

Thomas Widmer

Meta-Evaluation

Dr. Thomas Widmer

Meta-Evaluation

Kriterien zur Bewertung von Evaluationen

Verlag Paul Haupt Bern · Stuttgart · Wien

Thomas Widmer hat an der Universität Zürich Politikwissenschaft, Geschichte und Völkerrecht studiert (lic. Phil. I). Seit 1990 arbeitet er als Assistent und seit 1994 als wissenschaftlicher Mitarbeiter am Institut für Politikwissenschaft der Universität Zürich. Zudem ist er an der Philosophischen Fakultät I der Universität Zürich als Lehrbeauftragter tätig. 1995 hat er im Fach Politikwissenschaft promoviert (Dr. phil I). Seine Forschungsschwerpunkte bilden die Evaluationsforschung, die Policy-Analyse sowie die sozialwissenschaftliche Methodenlehre. Neben verschiedenen Artikeln in Fachzeitschriften hat er im Jahr 1991 eine Studie zur schweizerischen Luftreinhaltepolitik publiziert.

Elektronische Neuauflage
Zürich, 2006

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Widmer, Thomas :
Meta-Evaluation : Kriterien zur Bewertung von Evaluationen /
Thomas Widmer.–
Bern ; Stuttgart ; Wien : Haupt 1996
Zugl.: Zürich, Univ., Diss., 1995
ISBN 3-258-05454-1



INHALTSÜBERSICHT

<i>Detailliertes Inhaltsverzeichnis</i>	<i>VII</i>
<i>Vorwort</i>	<i>XV</i>
<i>1. Einleitung</i>	<i>1</i>
<i>2. Erkenntnisinteresse und Fragestellung</i>	<i>3</i>
2.1. Erkenntnisinteresse	3
2.2. Fragestellung	5
2.3. Hinweise zur Ausrichtung der Meta-Evaluation	6
<i>3. Kriterien der Meta-Evaluation</i>	<i>9</i>
3.1. Theoretischer Rahmen	9
3.2. Anforderungen an die Bewertungsinstrumente	9
3.3. Diskussion und Auswahl	13
3.4. Die eingesetzten Kriterien	14
3.5. Zwischenbemerkungen zu den Kriterien	39
3.6. Exkurs: Aktuelle Entwicklungen	41
<i>4. Empirisches Vorgehen</i>	<i>53</i>
4.1. Überblick	53
4.2. Auswahl der zehn Fallstudien	54
4.3. Zugrundeliegendes Datenmaterial	61
4.4. Auswertung des Datenmaterials	64
4.5. Vernehmlassung bei den Evaluatoren	65
4.6. Vergleichende Analyse	66
4.7. Hinweise zu speziellen personellen Konstellationen	66
<i>5. Die empirischen Fallstudien</i>	<i>69</i>
5.1. Fallstudie 'Lohngleichheitspostulat des Heimarbeitsgesetzes' (Studie Nr. 3)	71
5.2. Fallstudie 'Schneeräumung von Urner Passstrassen' (Studie Nr. 6)	121
5.3. Fallstudie 'Evaluation des BUWAL' (Studie Nr. 9)	175
<i>6. Vergleich zwischen den Fallstudien</i>	<i>241</i>
6.1. Vergleich entlang der Einzelkriterien	242
6.2. Vergleich entlang der Kriteriengruppen	258
6.3. Vergleich aufgrund des gesamten Kriterienrasters	265
6.4. Interdependenzen der Kriterien	270
6.5. Abschliessende Bemerkungen zum Fallstudienvergleich	288
<i>7. Folgerungen aus der empirischen Analyse</i>	<i>289</i>
7.1. Konsequenzen hinsichtlich der Kriterien	291
7.2. Folgerungen für die Ausgestaltung von Evaluationen	305
7.3. Zur Durchführung von Meta-Evaluationen	314
<i>8. Abkürzungsverzeichnis</i>	<i>321</i>
<i>9. Literaturverzeichnis</i>	<i>325</i>

DETAILLIERTES INHALTSVERZEICHNIS

<i>Vorwort</i>	<i>XV</i>
<i>1. Einleitung</i>	<i>1</i>
<i>2. Erkenntnisinteresse und Fragestellung</i>	<i>3</i>
2.1. Erkenntnisinteresse	3
2.2. Fragestellung	5
2.3. Hinweise zur Ausrichtung der Meta-Evaluation	6
<i>3. Kriterien der Meta-Evaluation</i>	<i>9</i>
3.1. Theoretischer Rahmen	9
3.2. Anforderungen an die Bewertungsinstrumente	9
3.3. Diskussion und Auswahl	13
3.4. Die eingesetzten Kriterien	14
3.4.1. Nützlichkeit	15
3.4.1.1. Identifikation der beteiligten Gruppen	15
3.4.1.2. Glaubwürdigkeit des Evaluators	16
3.4.1.3. Informationsrahmen und -auswahl	17
3.4.1.4. Bewertende Interpretation	17
3.4.1.5. Klarheit des Berichts	18
3.4.1.6. Verbreitung des Berichts	18
3.4.1.7. Rechtzeitigkeit des Berichts	19
3.4.1.8. Wirkung der Evaluation	19
3.4.2. Anwendbarkeit	20
3.4.2.1. Praktische Verfahren	20
3.4.2.2. Politische Überlebensfähigkeit	20
3.4.2.3. Kostenwirksamkeit	21
3.4.3. Korrektheit	22
3.4.3.1. Formale Verpflichtung	22
3.4.3.2. Deklaration von Interessenkonflikten	22
3.4.3.3. Vollständige und aufrichtige Offenlegung	23
3.4.3.4. Herstellung von Öffentlichkeit	23
3.4.3.5. Schutz individueller Menschenrechte	24
3.4.3.6. Menschliche Interaktionen	25
3.4.3.7. Ausgewogene Berichterstattung	25
3.4.3.8. Finanzielle Verantwortlichkeit	26
3.4.4. Genauigkeit	26
3.4.4.1. Identifikation des Gegenstandes	26
3.4.4.2. Kontextanalyse	27
3.4.4.3. Beschreibung von Zielen und Vorgehen	27
3.4.4.4. Verlässliche Informationsquellen	28
3.4.4.5. Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit	29

3.4.4.6.	Reliable Messung bzw. Verlässlichkeit	31
3.4.4.7.	Systematische Datenüberprüfung	32
3.4.4.8.	Analyse quantitativer Information	33
3.4.4.9.	Analyse qualitativer Information	33
3.4.4.10.	Gerechtfertigte Folgerungen	34
3.4.4.11.	Objektive Berichterstattung bzw. Bestätigbarkeit	35
3.4.5.	Die eingesetzten Kriterien im Überblick	36
3.5.	Zwischenbemerkungen zu den Kriterien	39
3.6.	Exkurs: Aktuelle Entwicklungen	41
3.6.1.	'The Program Evaluation Standards' 1994	41
3.6.1.1.	Allgemeines	41
3.6.1.3.	Feasibility Standards	44
3.6.1.4.	Propriety Standards	44
3.6.1.5.	Accuracy Standards	47
3.6.2.	'Evaluation Principles' der AEA	50
4.	<i>Empirisches Vorgehen</i>	53
4.1.	Überblick	53
4.2.	Auswahl der zehn Fallstudien	54
4.3.	Zugrundeliegendes Datenmaterial	61
4.4.	Auswertung des Datenmaterials	64
4.5.	Vernehmlassung bei den Evaluatoren	65
4.6.	Vergleichende Analyse	66
4.7.	Hinweise zu speziellen personellen Konstellationen	66
5.	<i>Die empirischen Fallstudien</i>	69
5.1.	Fallstudie 'Lohngleichheitspostulat des Heimarbeitsgesetzes' (Studie Nr. 3)	71
5.1.1.	Die Evaluation	71
5.1.1.1.	Das Thema	71
5.1.1.2.	Die Fragestellung	71
5.1.1.3.	Das Vorgehen	72
5.1.1.4.	Der Auftraggeber	72
5.1.1.5.	Die Forscher	72
5.1.2.	Beschreibung des Evaluationsprozesses	74
5.1.3.	Bewertung entlang der Kriterien	77
5.1.3.1.	Nützlichkeit	77
5.1.3.1.1.	Identifikation der beteiligten Gruppen	77
5.1.3.1.2.	Glaubwürdigkeit des Evaluators	78
5.1.3.1.3.	Informationsrahmen und -auswahl	79
5.1.3.1.4.	Bewertende Interpretationen	82
5.1.3.1.5.	Klarheit des Berichts	84
5.1.3.1.6.	Verbreitung des Berichts	85
5.1.3.1.8.	Wirkung der Evaluation	87

5.1.3.2.	Anwendbarkeit	88
	5.1.3.2.1. Praktische Verfahren	88
	5.1.3.2.2. Politische Überlebensfähigkeit	89
	5.1.3.2.3. Kostenwirksamkeit	90
5.1.3.3.	Korrektheit	91
	5.1.3.3.1. Formale Verpflichtung	91
	5.1.3.3.2. Deklaration von Interessenkonflikten	93
	5.1.3.3.3. Vollständige und aufrichtige Offenlegung	94
	5.1.3.3.4. Herstellung von Öffentlichkeit	94
	5.1.3.3.5. Schutz individueller Menschenrechte	94
	5.1.3.3.6. Menschliche Interaktionen	94
	5.1.3.3.7. Ausgewogene Berichterstattung	95
	5.1.3.3.8. Finanzielle Verantwortlichkeit	96
5.1.3.4.	Genauigkeit	96
	5.1.3.4.1. Identifikation des Gegenstandes	96
	5.1.3.4.2. Kontextanalyse	97
	5.1.3.4.3. Beschreibung von Zielen und Vorgehen	98
	5.1.3.4.4. Verlässliche Informationsquellen	99
	5.1.3.4.5. Glaubwürdigkeit und Übertragbarkeit	101
	5.1.3.4.6. Verlässlichkeit	103
	5.1.3.4.7. Systematische Datenüberprüfung	103
	5.1.3.4.8. Analyse quantitativer Informationen	103
	5.1.3.4.9. Analyse qualitativer Informationen	106
	5.1.3.4.10. Gerechtfertigte Folgerungen	108
	5.1.3.4.11. Bestätigbarkeit	110
5.1.4.	Gesamteindruck	111
5.1.5.	Anhang	112
	5.1.5.1. Unterlagen zur Evaluationsstudie	112
	5.1.5.2. Liste der kontaktierten Personen	113
	5.1.5.3. Literaturverzeichnis	114
5.1.6.	Stellungnahme der Evaluatoren	116
	5.1.6.1. Vorbemerkung	116
	5.1.6.2. Generelle Bemerkungen	116
	5.1.6.3. Zur Meta-Evaluation unserer Studie	116
	5.1.6.3.1. Kritik am Bearbeitungsteam	116
	5.1.6.3.2. Redaktionelle Kritik am Schlussbericht	117
	5.1.6.3.3. Methodische Kritikpunkte	117
	5.1.6.3.4. Inhaltliche Kritikpunkte	119
5.2.	Fallstudie 'Schneeräumung von Urner Passstrassen' (Studie Nr. 6)	121
5.2.1.	Die Evaluation	121
	5.2.1.1. Das Thema	121
	5.2.1.2. Die Fragestellung	121
	5.2.1.3. Das Vorgehen	122
	5.2.1.4. Der Auftraggeber	122
	5.2.1.5. Die Forscher	123
5.2.2.	Beschreibung des Evaluationsprozesses	124

5.2.3.	Bewertung entlang der Kriterien	126
5.2.3.1.	Nützlichkeit	126
5.2.3.1.1.	Identifikation der beteiligten Gruppen	126
5.2.3.1.2.	Glaubwürdigkeit des Evaluators	128
5.2.3.1.3.	Informationsrahmen und -auswahl	128
5.2.3.1.4.	Bewertende Interpretationen	130
5.2.3.1.5.	Klarheit des Berichts	131
5.2.3.1.6.	Verbreitung des Berichts	132
5.2.3.1.7.	Rechtzeitigkeit des Berichts	133
5.2.3.1.8.	Wirkung der Evaluation	134
5.2.3.2.	Anwendbarkeit	136
5.2.3.2.1.	Praktische Verfahren	136
5.2.3.2.2.	Politische Überlebensfähigkeit	137
5.2.3.2.3.	Kostenwirksamkeit	138
5.2.3.3.	Korrektheit	139
5.2.3.3.1.	Formale Verpflichtung	139
5.2.3.3.2.	Deklaration von Interessenkonflikten	142
5.2.3.3.3.	Vollständige und aufrichtige Offenlegung	144
5.2.3.3.4.	Herstellung von Öffentlichkeit	145
5.2.3.3.5.	Schutz individueller Menschenrechte	145
5.2.3.3.6.	Menschliche Interaktionen	145
5.2.3.3.7.	Ausgewogene Berichterstattung	146
5.2.3.3.8.	Finanzielle Verantwortlichkeit	147
5.2.3.4.	Genauigkeit	148
5.2.3.4.1.	Identifikation des Gegenstandes	148
5.2.3.4.2.	Kontextanalyse	149
5.2.3.4.3.	Beschreibung von Zielen und Vorgehen	150
5.2.3.4.4.	Verlässliche Informationsquellen	151
5.2.3.4.5.	Valide Messung	152
5.2.3.4.6.	Reliable Messung	155
5.2.3.4.7.	Systematische Datenüberprüfung	156
5.2.3.4.8.	Analyse quantitativer Informationen	157
5.2.3.4.9.	Analyse qualitativer Informationen	159
5.2.3.4.10.	Gerechtfertigte Folgerungen	159
5.2.3.4.11.	Objektive Berichterstattung	162
5.2.4.	Gesamteindruck	163
5.2.5.	Anhang	164
5.2.5.1.	Unterlagen zur Evaluationsstudie	164
5.2.5.2.	Liste der kontaktierten Personen	164
5.2.5.3.	Literaturverzeichnis	165
5.2.5.4.	Ergänzungen zur quantitativen Auswertung	167
5.2.6.	Stellungnahme der Evaluatoren	172
5.3.	Fallstudie 'Evaluation des BUWAL' (Studie Nr. 9)	175
5.3.1.	Die Evaluation	175
5.3.1.1.	Das Thema	175
5.3.1.2.	Die Fragestellung	175
5.3.1.3.	Das Vorgehen	176

5.3.1.4.	Der Auftraggeber	177
5.3.1.5.	Die Forscher	177
5.3.2.	Beschreibung des Evaluationsprozesses	181
5.3.3.	Bewertung entlang der Kriterien	185
5.3.3.1.	Nützlichkeit	186
5.3.3.1.1.	Identifikation der beteiligten Gruppen	186
5.3.3.1.2.	Glaubwürdigkeit des Evaluators	187
5.3.3.1.3.	Informationsrahmen und -auswahl	187
5.3.3.1.4.	Bewertende Interpretationen	189
5.3.3.1.5.	Klarheit des Berichts	190
5.3.3.1.6.	Verbreitung des Berichts	193
5.3.3.1.7.	Rechtzeitigkeit des Berichts	194
5.3.3.1.8.	Wirkung der Evaluation	195
5.3.3.2.	Anwendbarkeit	200
5.3.3.2.1.	Praktische Verfahren	200
5.3.3.2.2.	Politische Überlebensfähigkeit	201
5.3.3.2.3.	Kostenwirksamkeit	202
5.3.3.3.	Korrektheit	203
5.3.3.3.1.	Formale Verpflichtung	203
5.3.3.3.2.	Deklaration von Interessenkonflikten	205
5.3.3.3.3.	Vollständige und aufrichtige Offenlegung	206
5.3.3.3.4.	Herstellung von Öffentlichkeit	207
5.3.3.3.5.	Schutz individueller Menschenrechte	208
5.3.3.3.6.	Menschliche Interaktionen	208
5.3.3.3.7.	Ausgewogene Berichterstattung	209
5.3.3.3.8.	Finanzielle Verantwortlichkeit	210
5.3.3.4.	Genauigkeit	211
5.3.3.4.1.	Identifikation des Gegenstandes	211
5.3.3.4.2.	Kontextanalyse	212
5.3.3.4.3.	Beschreibung von Zielen und Vorgehen	213
5.3.3.4.4.	Verlässliche Informationsquellen	215
5.3.3.4.5.	Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit	217
5.3.3.4.6.	Reliable Messung bzw. Verlässlichkeit	217
5.3.3.4.7.	Systematische Datenüberprüfung	217
5.3.3.4.8.	Analyse quantitativer Informationen	218
5.3.3.4.9.	Analyse qualitativer Informationen	222
5.3.3.4.10.	Gerechtfertigte Folgerungen	222
5.3.3.4.11.	Objektive Berichterstattung bzw. Bestätigbark.	223
5.3.4.	Gesamteindruck	224
5.3.5.	Anhang	225
5.3.5.1.	Unterlagen zur Evaluationsstudie	225
5.3.5.2.	Liste der kontaktierten Personen	227
5.3.5.3.	Literaturverzeichnis	228
5.3.6.	Stellungnahme der Evaluatoren	234

6.	<i>Vergleich zwischen den Fallstudien</i>	241
6.1.	Vergleich entlang der Einzelkriterien	242
6.1.1.	Nützlichkeit	244
6.1.1.1.	Identifikation der beteiligten Gruppen	244
6.1.1.2.	Glaubwürdigkeit des Evaluators	245
6.1.1.3.	Informationsrahmen und -auswahl	245
6.1.1.4.	Bewertende Interpretationen	246
6.1.1.5.	Klarheit des Berichts	246
6.1.1.6.	Verbreitung des Berichts	246
6.1.1.7.	Rechtzeitigkeit des Berichts	247
6.1.1.8.	Wirkung der Evaluation	247
6.1.2.	Anwendbarkeit	248
6.1.2.1.	Praktische Verfahren	248
6.1.2.2.	Politische Überlebensfähigkeit	248
6.1.2.3.	Kostenwirksamkeit	249
6.1.3.	Korrektheit	249
6.1.3.1.	Formale Verpflichtung	249
6.1.3.2.	Deklaration von Interessenkonflikten	250
6.1.3.3.	Vollständige und aufrichtige Offenlegung	250
6.1.3.4.	Herstellung von Öffentlichkeit	251
6.1.3.5.	Schutz individueller Menschenrechte	251
6.1.3.6.	Menschliche Interaktionen	252
6.1.3.7.	Ausgewogene Berichterstattung	252
6.1.3.8.	Finanzielle Verantwortlichkeit	252
6.1.4.	Genauigkeit	253
6.1.4.1.	Identifikation des Gegenstandes	253
6.1.4.2.	Kontextanalyse	253
6.1.4.3.	Beschreibung von Zielen und Vorgehen	254
6.1.4.4.	Verlässliche Informationsquellen	254
6.1.4.5.	Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit	255
6.1.4.6.	Reliable Messung bzw. Verlässlichkeit	255
6.1.4.7.	Systematische Datenüberprüfung	255
6.1.4.8.	Analyse quantitativer Informationen	256
6.1.4.9.	Analyse qualitativer Informationen	256
6.1.4.10.	Gerechtfertigte Folgerungen	257
6.1.4.11.	Objektive Berichterstattung bzw. Bestätigbarkeit	257
6.2.	Vergleich entlang der Kriteriengruppen	258
6.2.1.	Nützlichkeit	258
6.2.2.	Anwendbarkeit	259
6.2.3.	Korrektheit	261
6.2.4.	Genauigkeit	262
6.3.	Vergleich aufgrund des gesamten Kriterienrasters	265
6.3.1.	Kriterien mit positiver Bewertung	265
6.3.2.	Kriterien mit negativer Bewertung	267
6.3.3.	Kriterien mit eingeschränkter Bewertbarkeit	267
6.4.	Interdependenzen der Kriterien	270

6.4.1.	Beziehungen der Einzelkriterien	273
6.4.2.	Netzwerkanalyse zu den Interdependenzen der Kriterien	282
6.4.3.	Beziehungen der Kriteriengruppen	286
6.5.	Abschliessende Bemerkungen zum Fallstudienvergleich	288
7.	<i>Folgerungen aus der empirischen Analyse</i>	289
7.1.	Konsequenzen hinsichtlich der Kriterien	291
7.1.1.	Untersuchungsergebnisse zum gesamten Kriterienraster	291
7.1.2.	Folgerungen und Vorschläge zu Einzelkriterien	292
7.1.3.	Vorschlag für einen revidierten Kriterienraster	300
7.2.	Folgerungen für die Ausgestaltung von Evaluationen	305
7.3.	Zur Durchführung von Meta-Evaluationen	314
8.	<i>Abkürzungsverzeichnis</i>	321
9.	<i>Literaturverzeichnis</i>	325

VORWORT

Dieses Buch befasst sich mit der Evaluationsforschung, einem Zweig der angewandten Sozialforschung, dem in den letzten Jahren in der Schweiz zunehmende Bedeutung zugekommen ist. Spezifisch geht es dabei um die Frage der Qualität von Evaluationen. Es interessiert dabei erstens die Frage, wie die Qualität von Evaluationsstudien eingeschätzt werden kann, zweitens wie gut spezifische Evaluationsstudien sind und drittens, wie die Qualität von Evaluationen erhalten bzw. verbessert werden kann.

Die Arbeiten an diesem Projekt sind im Rahmen des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) durch den Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF) finanziell unterstützt worden (SNF-Nr. 4027-39858).

Verschiedene Personen haben massgeblich zum Gelingen der vorliegenden Arbeit beigetragen. Prof. Dr. Ulrich Klöti, Leiter des Instituts für Politikwissenschaft an der Universität Zürich, hat die Arbeit betreut und durch seine Anmerkungen immer wieder zu kritischer Selbstreflexion veranlasst. Die Mitglieder der Expertengruppe und insbesondere der Programmleiter des NFP 27, Dr. Werner Bussmann, haben durch die finanzielle Unterstützung und die wertvollen Hinweise zur Studie beigetragen. Die von der Meta-Evaluation betroffenen Evaluatorinnen und Evaluatoren haben das vorliegende Projekt zwar nicht in allen Fällen mit offenen Armen begrüsst, waren aber trotz der teilweise vorhandenen Skepsis zu ausführlichen Gesprächen bereit, haben eine grosse Zahl von Dokumenten zur Verfügung gestellt und Entwürfe dieser Studie in mündlicher und schriftlicher Form kommentiert. Ohne deren Mitwirken wäre die Durchführung dieser Studie nicht möglich gewesen. Eine grosse Zahl von weiteren Personen, die sich in irgendeiner Form mit den hier untersuchten Evaluationsstudien befasst hatten, gaben bereitwillig Auskunft und überliessen uns umfangreiche schriftliche Materialien. Prof. James R. Sanders (Western Michigan University, Kalamazoo MI), Prof. Nick L. Smith (Syracuse University, Syracuse NY) und Prof. William M.K. Trochim (Cornell University, Ithaca NY) haben wichtige Unterlagen zur Verfügung gestellt und meine Ideen kritisch kommentiert. Meine Kolleginnen und Kollegen, Dr. Wolfgang Beywl (Universität zu Köln), Dr. Theo Haldemann (Universität St. Gallen), Dr. Armin Kühne (Zürich), Christine Rothmayr (Universität Zürich), Dr. Erwin Rüegg (Universität Potsdam), Walter Schenkel (Erasmus Universität, Rotterdam), Dr. Gerald Schneider (Universität Bern) und Uwe Serdült (Universität Zürich), haben verschiedene Arbeitspapiere und Entwürfe kritisch gelesen und hilfreiche Anmerkungen beigesteuert. Isa Reinhard hat den Text sprachlich überar-

beitet. Ich möchte mich bei all diesen Personen von ganzem Herzen für ihre Mithilfe bedanken.

Der grösste Dank gilt meiner Partnerin Isa Reinhard für das Verständnis, die Geduld und den Zuspruch. Ihre Unterstützung verhalf zur Überwindung mancherlei Hindernisse. Ihr und Nadja ist diese Arbeit gewidmet.

1. EINLEITUNG

Die vorliegende Studie befasst sich mit der Praxis einer angewandten sozialwissenschaftlichen Forschungsrichtung: der Evaluationsforschung. Der Schwerpunkt der Arbeit liegt in den Fragen, wie sich die Qualität von Evaluationen bewerten lässt und welche Folgerungen sich aus einer empirischen Bewertung ziehen lassen.

Die Untersuchung ist folgendermassen strukturiert: Nach der Spezifizierung des Erkenntnisinteresses und der Fragestellung (Kapitel 2) folgt die Bestimmung der zur Meta-Evaluation eingesetzten Bewertungskriterien (Kapitel 3). Im vierten Kapitel wird die empirische Vorgehensweise beschrieben und begründet, die in den nachfolgenden Fallstudien zu insgesamt zehn Evaluationsstudien umgesetzt wird. Die zehn ausgewählten Evaluationen, die in den Fallstudien entlang der Kriterien bewertet werden, stammen aus unterschiedlichen Politikbereichen. Für die vorliegende Publikation wurden drei der zehn Fallstudien ausgewählt (Kapitel 5). Die übrigen sieben Fallstudien sind an anderer Stelle publiziert worden (siehe Widmer 1996a-g). Auf der Basis dieser Fallstudien wird ein Vergleich entlang der einzelnen Bewertungskriterien vorgenommen (Kapitel 6). Daraus wiederum ergeben sich Folgerungen mit zwei verschiedenen Ausrichtungen: Erstens werden Schlüsse gezogen, die sich auf die Praxis der Evaluationsforschung richten und Empfehlungen zuhanden der Evaluationspraxis formulieren. Zweitens wird der Versuch unternommen, die eingesetzten Kriterien kritisch zu hinterfragen und Vorschläge zu deren Verbesserung zu präsentieren (Kapitel 7). Zum Schluss folgt der wissenschaftliche Apparat, also das Abkürzungs- (Kapitel 8) sowie das Literaturverzeichnis (Kapitel 9).

In diesem Text wird aus Gründen der einfachen Lesbarkeit bei gemischtgeschlechtlichen Personengruppen auf die Nennung der weiblichen Form verzichtet.

2. ERKENNTNISINTERESSE UND FRAGESTELLUNG

2.1. Erkenntnisinteresse

Der meta-evaluative Ansatz verfolgt die Absicht, die Praxis der Evaluationsforschung zu verstehen und zu verbessern. Etwas ausführlicher können exemplarisch folgende mögliche Zielsetzungen angeführt werden: die Einschätzung der Qualität, der Wirkung und der Verwendung von Evaluationsstudien zuhanden der Beteiligten oder interessierter Dritter; die Untersuchung der Charakteristik eines evaluativen Prozesses; die Vermeidung möglichen Missbrauchs; die Begutachtung evaluativer Arbeit; die Aufdeckung und Kontrolle von Verzerrungen in Evaluationen; die Einschätzung der Nützlichkeit neu entwickelter Evaluationsansätze. Diese Aufzählung ist nicht etwa als erschöpfend zu betrachten, sondern soll lediglich die Breite der denkbaren Zielsetzungen veranschaulichen. Es wird im Rahmen der vorliegenden Studie auch in keiner Weise möglich sein, alle diese Zielsetzungen zu berücksichtigen.

Der *Begriff Meta-Evaluation* bedarf einer etwas ausführlicheren Umschreibung, da er mit zwei unterschiedlichen Inhalten verwendet wird (vgl. dazu auch Cook/Gruder 1978, Posavac/Carey 1989: 281-285):

- Zum einen versteht man darunter Studien, die sich damit beschäftigen, inhaltliche Befunde aus verschiedenen Evaluationsstudien in irgendeiner Form zu kombinieren, zu integrieren oder zu kumulieren. Dieses Vorgehen könnte man auch mit der Bezeichnung 'kumulative Sekundäranalyse' oder 'substanzorientierte Meta-Evaluation' umschreiben. Es bestehen sowohl qualitative wie auch quantitative Ansätze. In der englischsprachigen Literatur werden für diese Vorgehensweise häufig die Begriffe 'meta-analysis' (siehe etwa Glass/McGaw/Smith 1981, Strube/Hartmann 1984, Wolf 1986, Hunter/Schmidt 1989, Hill/Andrews/Hoge 1991, Rossi/Freeman 1989: 263-264, Rosenthal 1991, Cook et al. 1992, Cordray 1992, Lipsey 1992, Vall 1993 oder Glaser/Shadish 1995) oder 'evaluation synthesis' (siehe Light 1984, Light/Pillemer 1984, Scriven 1991: 342-345, GAO 1992c, vgl. dazu auch Cooper/Hedges 1994) verwendet. Die Bezeichnung solcher Studien als Meta-Evaluationen ist in den letzten Jahren immer seltener geworden (ERS 1982, Hansen 1994: 559-560). Sie zielen darauf ab, mit einer Reduktion an Aufwand in der Datenerhebung sowie einer gleichzeitigen Erweiterung des hinzugezogenen - und bereits erhobenen - empirischen Datenmaterials substantielle Aussagen zu evaluativen Fragestellungen zu ermöglichen. Das Ziel dieses Ansatzes besteht somit in einer forschungsökonomisch sparsamen Vorgehensweise bei gleichzeitiger Erhöhung der empirischen Evidenz. Hinzu kommt der Vorteil gegenüber

alternativen Ansätzen, dass Studien dieser Art durchaus 'quick' sein können ohne unbedingt auch 'dirty' sein zu müssen. Eine unabdingbare Voraussetzung für diese Vorgehensweise besteht darin, dass eine ausreichend grosse Zahl von Evaluationsstudien zur interessierenden Fragestellung vorliegen muss. Auf diesen Ansatz wird im Rahmen dieser Untersuchung nicht mehr eingetreten. Ausführlichere Hinweise finden sich in der oben angegebenen Literatur.

- Im heutigen Sprachgebrauch wird unter Meta-Evaluation zumeist - und auch in der vorliegenden Untersuchung - die *Bewertung von Evaluationsstudien* verstanden (siehe etwa Scriven 1969, Stufflebeam 1978, Smith 1981b, Browne/Wildavsky 1984: 197, Stake 1986, Schwandt/Halpern 1988, Scriven 1991: 228-231, Shadish 1992, Gallegos 1994, Joint Committee 1994: 185-190, Georghiou 1995, Horvath 1995, Shula/Cousins 1995). Analog zu den Begriffsbildungen Meta-Theorie, Meta-Kommunikation usw. befassen sich Meta-Evaluationen mit der Evaluation von Evaluationen ("evaluating evaluations", Orata 1940). Dabei können verschiedene Aspekte von Interesse sein: die wissenschaftliche Qualität, die Praxisrelevanz, die Nutzung, die Wirkungen usw. einer oder mehrerer Evaluationen. Meta-Evaluationen können sowohl summativ wie auch formativ angelegt sein (siehe etwa Scriven 1991: 229-230, Joint Committee 1994: 185-186). Die vorliegende Untersuchung folgt diesem Ansatz.

Das *Ziel dieser Studie* besteht darin, ein deduktiv hergeleitetes und empirisch erprobtes Kriterienset für die Bewertung von Evaluationen zu entwickeln. Dies soll zur Bewertung sozialwissenschaftlich angelegter Evaluationen in der Schweiz geeignet sein. Dafür ist die Anpassung und Umsetzung von Kriterien zur Evaluation von Evaluationen im schweizerischen Kontext nötig. Diese Kriterien sollen als Instrumentarium zur Bewertung von Evaluationen dienen und damit einen Beitrag zur Qualitätskontrolle und Qualitätssicherung in der Evaluationsforschung leisten. Dies setzt jedoch voraus, dass diese Kriterien einige Bedingungen zu erfüllen vermögen. Diese werden wir im Kapitel 3 ausführlicher diskutieren. Es geht dabei in keiner Weise darum, die einzelnen, an der Evaluation beteiligten Personen zu bewerten. Dazu wäre ein anderes Vorgehen zu wählen (vgl. beispielsweise Joint Committee 1988). Unser Erkenntnisinteresse richtet sich auf die gesamte Evaluation, auf ihre Prozesse und Produkte. Die Meta-Evaluation soll weiter Hinweise auf spezifische Stärken und Schwächen der heutigen Evaluationspraxis in der Schweiz ermöglichen. Daraus sollen sich Empfehlungen für die Evaluationspraxis herleiten lassen.

Persönlich besteht ein dreifaches Interesse an der Durchführung dieser Meta-Evaluation. Zum einen interessieren die methodischen Aspekte der Untersuchung. Eine besondere Herausforderung besteht diesbezüglich darin, einen umfassenden Ansatz zu verfolgen,

der die unterschiedlichsten Ansätze und Ausprägungen der heutigen Evaluationsmethodik einzufangen versteht, ohne eine spezifische Richtung zu präferieren oder zu diskriminieren. Zweitens besteht ein Interesse daran, Erkenntnisse über die Praxis eines angewandten Forschungszweigs zu erhalten, der sich in den letzten Jahren in der Schweiz wachsender Bedeutung erfreuen konnte. Die Evaluationsforschung ist drittens deshalb von besonderem Interesse, da sie sich an der Nahtstelle von Politik und Wissenschaft (oder von Theorie und Praxis) bewegt. Damit setzt sich die Evaluation unterschiedlichen (und oft divergierenden) Ansprüchen, Erwartungen, Anforderungen und Verhaltensmustern, aber auch Macht- wie Wahrheitsvorstellungen aus.

2.2. Fragestellung

Aus dem dargelegten Erkenntnisinteresse ergibt sich eine dreiteilige Fragestellung. In einem *ersten Schritt* sind auf *theoretischer Ebene die Kriterien* zu bestimmen, die zur Evaluationsbewertung eingesetzt werden. In diesem Teil gehen wir folgender Fragestellung nach:

- Wie lässt sich eine Evaluation bewerten? Welche Kriterien dienen zur Einschätzung der Qualität einer Evaluation?

Diese Fragestellung wird in Kapitel 3 behandelt.

In einem *zweiten, empirischen Schritt* wird es darum gehen, ausgewählte Evaluationsstudien zu beschreiben und zu bewerten. Die dazugehörigen Fragen lassen sich folgendermassen formulieren:

- Wie gestaltet sich eine Evaluation? (Deskription)
- Wie ordnen sich empirische Evaluationsstudien hinsichtlich der Bewertungskriterien ein? Erfüllen die Evaluationen die Kriterien? (Evaluation)

In Kapitel 4 finden sich die zur Bearbeitung dieser Fragestellung erforderlichen Überlegungen methodischer Art zur empirischen Umsetzung. Das Kapitel 5 befasst sich mit der Beschreibung und Evaluation empirischer Fallstudien gemäss der oben angeführten Fragestellung.

In einem *dritten Schritt* ist aufgrund der empirischen Analyse eine *vergleichende Synthese* vorzunehmen, und zwar hinsichtlich der Bewertungskriterien und im Hinblick auf Empfehlungen zur Ausgestaltung von Evaluationen. Es werden in diesem Syntheseteil folgende Fragen behandelt:

- Inwiefern eignen sich die hergeleiteten Kriterien zur Bewertung von schweizerischen Evaluationen? Welche Änderungen drängen sich vor dem Hintergrund der empirischen Erfahrungen auf?
- Welche Empfehlungen für die weitere Evaluationspraxis lassen sich aufgrund der Bewertung bisher durchgeführter Evaluationen formulieren?

Diese Fragen werden in den Kapiteln 6 und 7 der vorliegenden Studie behandelt.

2.3. Hinweise zur Ausrichtung der Meta-Evaluation

Unsere Studie zielt darauf ab, Evaluationen zu bewerten. Es wird dagegen nicht möglich sein, aufgrund unserer Analyse auf die Qualitäten der Evaluatoren zu schliessen. Erstens wäre dazu sicher eine andere Vorgehensweise zu wählen. Wir beschäftigen uns mit Prozessen und Produkten eines Interaktionsprozesses. Daran sind mehrere Personen (Evaluator(en), Auftraggeber, Mitglieder von begleitenden Gremien, Vertreter verschiedenster Interessen und Organisationen usw.) beteiligt. Es ist deshalb nicht zulässig, aufgrund der Bewertung einer Evaluation auf personelle Qualitäten einzelner Personen zu schliessen. Zweitens richtet sich die Meta-Evaluation nur auf je eine Evaluation eines Evaluators (oder eines Evaluationsteams). Erkenntnisse aus der Untersuchung eines Projektes lassen sich nicht auf andere Evaluationen übertragen. Ebenso wenig lässt sich daraus auf die individuellen Qualitäten der beteiligten Personen schliessen. Zu beachten ist, dass jede Evaluation in einem spezifischen Kontext erarbeitet wurde. So sind beispielsweise die Handlungsspielräume der Evaluatoren sehr oft stark und in unterschiedlicher Weise begrenzt. Fragestellungen einer Evaluation etwa, oder ihr Untersuchungsbereich werden sehr häufig vom Auftraggeber definiert. Die Gegenstandsbereiche von Evaluationsstudien, ebenso wie zum Beispiel die Datenlage, weisen stark unterschiedliche Charakteristiken auf. Weiter führen auch die finanziellen und zeitlichen Ressourcen, die zur Durchführung der Evaluation zur Verfügung stehen, zu deutlich divergierenden Rahmenbedingungen. Zudem sind die untersuchten Evaluationen im Abstand von rund fünf Jahren entstanden. Da sich das Know-how in der Schweizer Evaluationsforschung in dieser Phase stark entwickelt hat, ist auch dies zu berücksichtigen. Scarvia B. Anderson und Samuel Ball führen dazu folgende Illustration an: "Certainly, given that the real world offers different opportunities, it is very difficult to compare the competencies of two evaluators in terms of products alone. As Humphrey Bogart said of the Academy Awards, the only way they could be fairly awarded was to have everyone in a given category play the same part in front of the same judges." (Anderson/Ball 1978: 190)

Aus diesen Überlegungen heraus ergibt sich, dass unsere Analyse keine Antworten auf Fragen, wie die folgenden, geben kann: Wer erarbeitet die beste/schlechteste Evaluation? Oder: Wer ist der beste/schlechteste Evaluator? Vielmehr besteht das Ziel darin, zu untersuchen, unter welchen Rahmenbedingungen welche spezifische Ausgestaltung der Evaluation - an der sich immer verschiedene Akteure beteiligen - erfolgreich sein kann.

3. KRITERIEN DER META-EVALUATION

3.1. Theoretischer Rahmen

Evaluationen befassen sich mit den unterschiedlichsten Programmen, Objekten, Verfahren, Personen, Instrumenten usw. Sie stellen damit Konstruktionen zweiter Ordnung dar, die sich mit Konstruktionen erster Ordnung (der sozialen Realität) auseinandersetzen. Meta-Evaluationen haben Evaluationen zum Untersuchungsgegenstand. Es handelt sich somit um Konstruktionen dritter Ordnung.

Diese Studie verfolgt auf der Metaebene den in der Evaluationsforschung weitverbreiteten, *programmtheoretischen Ansatz* (vgl. Bickman 1987, Chen 1990, Chen/Rossi 1992). Dieser besagt, dass die Evaluation (Konstruktion zweiter Ordnung) unter anderem die im zu evaluierenden Programm (Konstruktion erster Ordnung) (oft implizit) enthaltenen theoretischen Überlegungen expliziert und als Theorie übernimmt (Chen/Rossi 1980: 113-119, Chen/Rossi 1983: 289-290, Bickman 1987: 5-6, Chen 1990: 66-67, Palumbo/Gregware 1992: 145-146). Es handelt sich also um eine 'prozessgenerierte' Theorie. Umgesetzt auf die Meta-Evaluation (Konstruktion dritter Ordnung) bedeutet dies, dass die Evaluationsmethodologie (Konstruktion zweiter Ordnung) als Programmtheorie aufzufassen ist (Shadish 1992). Die Evaluationsforschung fällt bezüglich der methodologischen Ausrichtung durch eine Vielzahl sehr heterogener Orientierungen auf. Dies hängt unter anderem damit zusammen, dass sich die Evaluation als angewandte Forschung aus verschiedensten wissenschaftlichen Disziplinen heraus entwickelt hat, also schon von daher einen stark multidisziplinären Charakter aufweist. Dies führt dazu, dass es problematisch wird, über alle Paradigmen und Disziplinen hinweg geeignete Kriterien zur Bewertung von Evaluationen zu finden.

3.2. Anforderungen an die Bewertungsinstrumente

Die Bewertungskriterien müssen verschiedene Erfordernisse erfüllen, um in der Meta-Evaluation zweckmässig eingesetzt werden zu können:

- *Methodologische Offenheit*: Die Kriterien müssen so ausgestaltet sein, dass sie in der Lage sind, Evaluationen unterschiedlichster methodologischer Ausrichtungen zu bewerten. Kriterien, die sich auf spezifische erkenntnistheoretische Positionen abstützen, sind nicht geeignet. Der Kriterienraster muss gegenüber den verwendeten Evaluationsansätzen möglichst offen sein.

- *Umfassende Bewertung*: Weiterhin sollte die Bewertung einer Evaluation als Ganzes ermöglicht werden. Kriterien, die lediglich Teilaspekte (Nutzung, Wirkung, Methodik, Validität etc.) abdecken, reichen dazu nicht aus.
- *Weite Verbreitung, Anerkennung und Akzeptanz*: Die Kriterien sollten möglichst weit verbreitet sein und über eine breite Anerkennung durch Vertreter unterschiedlicher Evaluationsansätze verfügen. Sie sollten zudem im weiten Evaluationsfeld auf eine hohe Akzeptanz stossen. Zur Beurteilung stützen wir uns auf die Anerkennung durch relevante Fachorganisationen sowie auf die einschlägige Literatur.
- *Ausreichender Detaillierungsgrad*: Der zu verwendende Kriterienraster sollte die verschiedensten der zu bewertenden Aspekte einer Evaluation einzufangen vermögen. Dazu ist eine grössere Zahl von Kriterien und eine ausreichend differenzierte Formulierung der einzelnen Kriterien erforderlich.

Es besteht im Bereich der Evaluation eine grosse Vielfalt an Instrumenten zur Bewertung von politischen Programmen. Diese sind nun mit entsprechenden Anpassungen im Prinzip auch für den Bereich der Meta-Evaluation anwendbar. Diese *Kriterienraster* sind oft in recht knapper Weise in Form von sogenannten Checklisten zusammengestellt, die dem Evaluationspraktiker die Verwendung der Kriterien erleichtern sollen (siehe etwa Black 1993: 180, Scriven 1991: 204-211, Millman 1981). Daneben existieren aber auch einige Ansätze, die speziell zu meta-evaluativen Zwecken entwickelt wurden (Gowin 1981, Guba/Lincoln 1989: 233-251, Chen 1990: 58-65, Scriven 1991: 230, Shadish 1992). Diesen Ansätzen fehlt es jedoch an einer breiteren Anerkennung und/oder sie behandeln nur eine eingeschränkte Problemstellung.

Ein weiteres Hilfsmittel zur Beurteilung und Bewertung von Evaluationsstudien wird durch verschiedene Evaluations-Standards bereitgestellt. In den Vereinigten Staaten sind seit den siebziger Jahren verschiedene Kataloge von Standards entwickelt worden, welche die Anforderungen an Evaluationen definieren sollen.

Unter Evaluations-Standards werden breit anerkannte Prinzipien zur Einschätzung der Qualität von Evaluationen verstanden (Ridings/Stufflebeam 1981: 7, vgl. Newman/Brown 1996: 23). Damit scheinen diese Standards als Kriterien zur Bewertung von Evaluationsstudien geradezu prädestiniert zu sein, obwohl sie von ihrer ursprünglichen Zielsetzung her dazu geschaffen wurden, Evaluationspraktikern einen Orientierungsrahmen für ihre Arbeit zu geben und damit nicht primär eine (ex-post) bewertende, sondern eine präventive, formative und damit handlungsleitende Funktion wahrzunehmen. Zwei der am weitesten verbreiteten Evaluations-Standards seien im folgenden erläutert (vgl. auch

Comptroller General of the United States 1988 und Canadian Evaluation Society (CES) 1992).

Standards der 'Evaluation Research Society' (ERS-Standards)

Die ERS-Standards (ERS Standards Committee 1982) sind durch ein Komitee der 'Evaluation Research Society' erarbeitet worden. Diese Gesellschaft stellt eine berufständische Organisation von vor allem US-amerikanischen Evaluationsforschern dar und umfasst zwischen 2'000 und 3'000 Mitglieder. Seit 1985 bildet diese Organisation zusammen mit dem mitgliedermässig kleineren 'Evaluation Network' (ENet) die 'American Evaluation Association' (AEA; Berk/Rossi 1990: 110-111). Die ERS-Standards setzen sich aus sechs Sektionen mit insgesamt 55 Einzelstandards zusammen. Die Sektionen befassen sich mit den folgenden Aspekten:

- Formulierung und Aushandlung des Projektes (formulation and negotiation)
- Struktur und Design der Evaluation (structure and design)
- Datenerhebung und -aufbereitung (data collection and preparation)
- Datenanalyse und -interpretation (data analysis and interpretation)
- Kommunikation und Bekanntgabe der Ergebnisse (communication and disclosure)
- Verwendung der Ergebnisse (use of results)

Die Einzelstandards sind somit sequentiell, entlang der einzelnen Phasen eines traditionellen Evaluationsprozesses gruppiert. Die grundlegenden Annahmen, auf denen die einzelnen Kriterien, die Sektionen und die Standards als Ganzes aufbauen, bestehen in einem klar traditionell orientierten Wissenschaftsverständnis. Alternative wissenschaftliche Paradigmen (bspw. naturalistische und responsive Ansätze), in der heutigen Evaluationsforschung bereits recht weit verbreitet, sind dabei nicht berücksichtigt. Der Evaluator hat beispielsweise in seinem Evaluationsbericht klar zwischen objektiven Ergebnissen, Meinungen und Spekulationen zu unterscheiden (Standard No. 39). Wie Lee J. Cronbach (1982: 53) anmerkt, erscheint der Begriff 'values' in den Standards kein einziges Mal. Die Standards präferieren ganz deutlich quantitative und experimentelle Ansätze gegenüber allen anderen, bauen auf einer linear verlaufenden Struktur der Evaluationsaktivitäten auf und sehen eine starke Verpflichtung des Evaluators gegenüber dem Auftraggeber (Standard No. 7). Die ERS-Standards sind zwar in der Literatur der frühen achtziger Jahre relativ breit diskutiert worden, das Interesse hat seither aber deutlich abgenommen. Durch die im Jahre 1994 erschienen 'Guiding Principles' hat sich deren Bedeutung weiter reduziert (siehe unten, Kapitel 3.5.2.).

Standards des Joint Committee

Die "Standards for Evaluation of Educational Programs, Projects, and Materials" (Joint Committee 1981a), die ursprünglich ausschliesslich für die erziehungswissenschaftliche Evaluationsforschung entwickelt wurden, haben über ihren Fachbereich hinaus weite Beachtung gefunden. Sie bestehen aus insgesamt dreissig Einzelstandards, die in vier Gruppen zusammengefasst sind:

- Die *Nützlichkeitsstandards* (utility standards) dienen dazu, dass die Evaluation den Informationsbedürfnissen der beteiligten Gruppen (Entscheidungsträger, Programmverantwortliche, Programmpersonal, Programmteilnehmer *und* Programmbetroffene) nachkommt. Die Gruppe umfasst acht Einzelstandards.
- Die *Anwendbarkeitsstandards* (feasibility standards) sollen für praktikable und politisch tragbare Evaluationen sorgen. Hierzu führt das Joint Committee drei Einzelstandards an.
- Die *Korrekteitsstandards* (propriety standards) zielen primär auf die rechtliche und ethische Verantwortlichkeit in der Evaluationsforschung. Diese Gruppe besteht aus acht Normen.
- Die *Genauigkeitsstandards* (accuracy standards) beziehen sich primär auf die Einhaltung bestimmter Regeln der sozialwissenschaftlichen Arbeitsweise. Es werden insgesamt elf verschiedene Anforderungen aufgeführt.

In der Gruppe der Genauigkeitsstandards sind Standards enthalten, die für die Evaluation interpretativer Evaluationsstudien nur bedingt geeignet sind. Damit sind namentlich die Kriterien der Validität (Nr. 5), der Reliabilität (Nr. 6) und der Objektivität (Nr. 11) angesprochen (vgl. dazu Kirk/Miller 1986). Egon G. Guba und Yvonna S. Lincoln haben jedoch zu diesen sogenannte 'parallele Kriterien' entwickelt, die der alternativen Methode besser gerecht werden sollen. Der Validität, ein Begriff der vor allem durch die Unterscheidung von interner und externer Validität durch Donald T. Campbell (Cook/Campbell 1979: 37-94 und Campbell 1957) geprägt wurde, setzen Guba und Lincoln die Kriterien der Glaubwürdigkeit (credibility) für die interne, und der Übertragbarkeit (transferability) für die externe Validität gegenüber. Als Parallelkriterium zur Reliabilität bezeichnen die Autoren die Verlässlichkeit (dependability), zur Objektivität die Bestätigbarkeit (confirmability) im Sinne einer intersubjektiven Nachvollziehbarkeit (Guba/Lincoln 1989: 236-243).

3.3. Diskussion und Auswahl

Aufgrund der in Kapitel 3.2. angeführten Auswahlkriterien (methodologische Offenheit; umfassende Bewertung; weite Verbreitung, Anerkennung und Akzeptanz; ausreichender Detaillierungsgrad) gilt es nun einen der vorgestellten Ansätze auszuwählen.

Die oben diskutierten ERS-Standards (wie auch weitere Ansätze, vgl. etwa Raizen/Rossi 1981) beschränken sich auf eine kritisch-rationale bzw. neopositivistische Position, erfüllen also das Postulat der methodologischen Offenheit nicht. Aufgrund der Struktur der ERS-Standards ist auch eine Anpassung an alternative Ansätze nicht zu bewerkstelligen (vgl. dazu auch die Ergebnisse in Stockdill 1986).

Andere Ansätze zur Qualitätsbestimmung sind gerade eben für das traditionelle Paradigma nicht anwendbar (siehe Guba/Lincoln 1989: 233-251) und fallen deshalb ebenfalls ausser Betracht.

Der von D. Bob Gowin (1981) präsentierte Ansatz erreicht weder einen ausreichenden Detaillierungsgrad noch erfreut er sich weiter Verbreitung, Anerkennung oder Akzeptanz. Dasselbe gilt auch für die 'checklist'-Ansätze von Jason Millman (1981), Michael Scriven (1991: 204-211 und 230) oder Thomas R. Black (1993: 180).

Der Ansatz von William R. Shadish (1992) kommt deshalb für unsere Meta-Evaluation nicht in Frage, weil er sich auf spezifische Aspekte konzentriert, und damit das Kriterium einer umfassenden Bewertung nicht erfüllt. Zusätzlich ist dieser Ansatz auch nur wenig verbreitet und zu wenig detailliert spezifiziert.

Die Joint Committee-Standards erfüllen das Kriterium einer umfassenden Bewertung, da sie sich mit den verschiedenen Aspekten, die im Zusammenhang mit einer Evaluation relevant sind, befassen. Sie beschränken sich nicht auf wissenschafts- oder praxisorientierte Gesichtspunkte. In Anbetracht der heutigen Entwicklung hin zu einer immer vielfältigeren Evaluationsmethodologie scheinen sich die Standards des Joint Committees weiterhin ihre Bedeutung und eine allgemein hohe Akzeptanz erhalten zu können (Newman/Brown 1996 und 1992, Horvath 1995, Newman 1995: 108, Gallegos 1994, Mertens 1994: 24 und 1990, Patton 1994: 198-199, Worthen 1994: 11, Morris/Cohn 1993, Smith/Chircop/Mukherjee 1993, Nowakowski 1990, Schwandt 1990, Guba/Lincoln 1989: 229-233, Beywl 1988: 113-123 und Stockdill 1986), auch wenn (noch) nicht von einer allgemeinen Anerkennung im breiten Evaluationsfeld gesprochen werden kann (siehe etwa Love 1994: 32-33, vgl. auch unten, Kapitel 3.5.). Damit ist auch das für die Auswahl relevante Kriterium der weiten Verbreitung, Anerkennung und Akzeptanz erfüllt. Die Joint Committee-Standards umfassen insgesamt dreissig Einzelstandards, die

sich vier Gruppen zuordnen lassen. Der Detaillierungsgrad ist damit ausreichend. Das einzige Auswahlkriterium, das die Joint Committee-Standards nicht ganz zu erfüllen vermögen, ist jenes der methodologischen Offenheit. Da aber durch Egon G. Guba und Yvonna S. Lincoln für alternativ orientierte Evaluationsansätze Parallelkriterien bereitgestellt wurden (Guba/Lincoln 1989: 236-243), mit denen sich die Joint Committee-Standards ergänzen lassen, kann auf diese Weise auch das Erfordernis der methodologischen Offenheit eingehalten werden.

Mit diesen Ergänzungen eignen sich die Standards des Joint Committee am besten als Analyseinstrument für die hier zur Debatte stehende Problemstellung. Deshalb wird in diesem Projekt primär mit Kriterien gearbeitet, die von diesen Standards abgeleitet sind.

Seit Juli 1994 ist eine überarbeitete Fassung dieser Standards verfügbar (Joint Committee 1994). Aus zeitlichen Gründen konnte diese revidierte Fassung für die empirischen Teile der vorliegenden Arbeit nicht mehr berücksichtigt werden (vgl. aber Kapitel 7.1.). Wir werden weiter unten darauf zurückkommen (siehe Kapitel 3.5.).

Sämtliche uns bekannten Standards haben ihren Ursprung im nordamerikanischen Kontext (vgl. auch CES 1992, Comptroller General 1988). Inwiefern diese auf die schweizerischen Verhältnisse übertragbar sind, wird sich in der empirischen Umsetzung der Standards zeigen (vgl. zu dieser Frage vorerst Lewy 1984, Marklund 1984, Nevo 1984 und Smith/Chircop/Mukherjee 1993; siehe auch Hendricks/Conner 1995).

3.4. Die eingesetzten Kriterien

Wie bereits erwähnt, sind die Standards des 'Joint Committee on Standards for Educational Evaluation' (Joint Committee 1981a) in vier Gruppen gegliedert, die sich jeweils aus einer unterschiedlichen Anzahl von Einzelkriterien zusammensetzen. Die Joint Committee-Standards bestehen aus Einzelstandards, die graduell unterschiedliche Ausrichtungen aufweisen. Neben den eher wissenschaftlichen Fragen sind auch Aspekte abgedeckt, die als typisch für anwendungs- resp. praxisorientierte Forschung bezeichnet werden können. Als dritte Perspektive kommt die spezifisch evaluationsorientierte Ausrichtung hinzu.

Aufgrund der US-amerikanischen Standards haben wir jene Kriterien, die *wir* zur Bewertung einsetzen, abgeleitet. Je nach erkenntnistheoretischer Position sind drei der Kriterien unterschiedlich ausgestaltet. Für naturalistische, konstruktivistische, responsive oder interpretative Ansätze (im folgenden kurz auch 'alternative Ansätze' genannt) wurden die Kriterien der Validität (siehe Kapitel 3.4.4.5.), der Reliabilität (siehe Kapitel 3.4.4.6.) und der Objektivität (siehe Kapitel 3.4.4.11.) durch die von Guba/Lincoln vor-

geschlagenen Parallelkriterien ersetzt (Guba/Lincoln 1989: 236-243). Für die im konventionellen, realistischen, kausalitätsorientierten Paradigma ('traditionelle Ansätze') zu verortenden Evaluationen werden für die Bewertung die aufgrund der unveränderten Standards abgeleiteten Kriterien eingesetzt.

Bei der Formulierung der Kriterien sind wir den Joint Committee-Standards möglichst eng gefolgt. Änderungen erfolgten primär dort, wo sich das Original auf spezifisch erziehungswissenschaftliche Aspekte bezieht. Weiterhin erfolgte dort eine Anpassung, wo die Joint Committee-Standards aufgrund ihrer US-amerikanischen Herkunft für den schweizerischen Kontext offensichtlich nicht geeignet sind. Auf weitere Veränderungen wurde vorderhand verzichtet. Solche werden sich aber möglicherweise aufgrund der empirischen Analysen aufdrängen (und dann auch in legitimer Weise begründen lassen).

Wir werden im folgenden die einzelnen Kriterien vorstellen und kurz erläutern.

3.4.1. Nützlichkeit

3.4.1.1. Identifikation der beteiligten Gruppen

Die Beteiligten oder Betroffenen einer Evaluation sollten identifiziert werden, damit deren Interessen berücksichtigt werden können.

Der Kreis der beteiligten und betroffenen Personen umfasst folgende Akteure (nach Rossi/Freeman 1989: 423; vgl auch Weiss 1983, House 1993: 117-121, Kelly/Maynard-Moody 1994, Stevens/Dial 1994b: 8-9):

- *Entscheidungsträger und Programmgestalter*: Personen, die für Entscheide über die Einführung, Fortsetzung, Umgestaltung, Erweiterung, Reduktion oder Absetzung eines Programms verantwortlich sind.
- *Programmfinanzierende*: Personen oder Organisationen, die das Programm finanzieren.
- *Evaluationsfinanzierende*: Personen oder Organisationen, welche die Evaluation finanzieren.
- *Zielgruppe des Programms*: Personen oder Organisationen, die am Programm teilnehmen, angebotene Dienstleistungen nutzen, von regulativen Vorschriften betroffen sind usw.
- *Programmverantwortliche*: Personen oder Gruppen, welche die Durchführung des Programms leiten.

- *Evaluatoren*: Personen, Gruppen oder Institutionen, die für die Ausgestaltung und die Durchführung der Evaluation verantwortlich sind.
- *Programm-Mitbewerber*: Personen, Gruppen oder Institutionen, die um Ressourcen konkurrieren.
- *Evaluationsgemeinde*: Evaluatoren, welche die Evaluation zur Kenntnis nehmen und sich ein Urteil bilden wollen.

Diese Aufzählung soll lediglich die Bandbreite der möglichen 'stakeholders' veranschaulichen und erhebt keinerlei Anspruch auf Vollständigkeit. Weiterhin ist zu beachten, dass jeweils zwischen diesen Gruppen personelle Überschneidungen bestehen können.

Zudem haben verschiedene andere Autoren Versuche unternommen, die 'stakeholders' näher zu definieren. Egon G. Guba und Yvonna S. Lincoln unterscheiden grob in drei Gruppen: die Agenten (Personen, die an der Produktion, der Verwendung und der Implementation des Evaluandums beteiligt sind), die Begünstigten (Personen, die in irgendeiner Form vom Evaluandum profitieren) und die Opfer eines Programms (Personen, die in irgendeiner Form in negativer Weise vom Evaluandum betroffen sind) (nach Guba/Lincoln 1989: 40-41). Carol H. Weiss führt in ihrer 'stakeholder'-Typologie vier Gruppen an, die sich durch ihre Informationsbedürfnisse unterscheiden. Es sind dies die Programmgestalter, die Programmverantwortlichen, die Praktiker sowie die Klienten (oder Adressaten) eines Programms (Weiss 1983: 84-85). Lee J. Cronbach benennt fünf Gruppen, die eine sogenannte 'policy-shaping community' bilden: Politikgestalter, Programmverantwortliche, Programmvollziehende, Programmbetroffene und 'Illuminatoren', also Personen, die Informationen über das Programm interpretieren und verbreiten (Journalisten, Beobachter, Wissenschaftler etc.) (Cronbach et al. 1980).

Die Evaluation sollte in impliziter oder expliziter Weise diese Gruppen identifizieren, um deren Interessen berücksichtigen zu können. (Joint Committee 1981a: 21-23)

3.4.1.2. Glaubwürdigkeit des Evaluators

Die Personen, welche die Evaluation durchführen, sollten sowohl vertrauenswürdig wie auch kompetent sein, damit die Ergebnisse der Evaluation ein Höchstmass an Glaubwürdigkeit und Akzeptanz erreichen.

Evaluatoren sind vertrauenswürdig, sofern sie über eine entsprechende Ausbildung, technische Kompetenz, inhaltliche Kenntnisse, Erfahrung, Integrität, Fähigkeiten in Öffentlichkeitsarbeit und möglicherweise über weitere, seitens der Adressaten der Evaluation

erwarteten, Kenntnisse und Charakteristiken verfügen. Es wird nur in den seltensten Fällen möglich sein, dass eine einzelne Person alle diese Erfordernisse erfüllen kann. Durch eine geeignete personelle Zusammensetzung eines Evaluatorenteams lässt sich jedoch erreichen, dass das Kollektiv diesen Ansprüchen gerecht wird. Die Bedeutung der einzelnen Aspekte hängt vom Kontext, von der Art des Evaluandums sowie von der Ausgestaltung der Evaluation ab. Durch eine hohe Glaubwürdigkeit des Evaluators kann auch jene der Evaluation gesteigert werden. (Joint Committee 1981a: 24-26)

3.4.1.3. Informationsrahmen und -auswahl

Die beschafften Informationen sollten von einem Umfang und einer Auswahl sein, welche die Behandlung sachdienlicher Fragen zum Evaluandum ermöglichen und gleichzeitig die Interessen der Beteiligten und Betroffenen berücksichtigen.

Es können in einer Evaluation - angesichts der beschränkten Ressourcen - nie alle wünschbaren Informationen beschafft werden, so dass es sich aufdrängt, selektiv vorzugehen. Dieses Kriterium weist darauf hin, dass es notwendig ist, sich auf jene Aspekte zu konzentrieren, die zur Behandlung der Evaluationsfragen wesentlich sind. Dabei sollte darauf geachtet werden, dass die wichtigen Informationsbedürfnisse der 'stakeholders' berücksichtigt werden. Der Auftraggeber kann oft dazu beitragen, die erforderlichen Informationen zu beschaffen. (Joint Committee 1981a: 27-31)

3.4.1.4. Bewertende Interpretation

Die für die Interpretation der Ergebnisse verwendeten Perspektiven, Verfahren und Begründungen sollten sorgfältig beschrieben werden, damit die Grundlagen für die Werturteile klar werden.

Die bewertende Interpretation der beschafften Informationen ist der zentrale Aspekt jeder Evaluation. Dieses Kriterium richtet sich darauf, dass der verfolgte Ansatz zur Bewertung der beschafften Informationen sorgfältig ausgewählt, präzise beschrieben und überzeugend begründet wird. Dabei ist es wichtig, auch alternative Perspektiven und Verfahren in Betracht zu ziehen und bei der Wahl zu berücksichtigen, dass diesbezügliche Entscheidungen willkürlich erscheinen können und deshalb einen Diskussionsgegenstand darstellen. Aus diesen Gründen sollte der Evaluator darauf achten, dass der Evaluationsprozess gegenüber weiteren, bisher nicht beachteten Wertorientierungen offen bleibt und diese bei Bedarf in die Evaluation integriert werden können. (Joint Committee 1981a: 32-36)

3.4.1.5. Klarheit des Berichts

Der Evaluationsbericht sollte das evaluierte Objekt und den Kontext ebenso beschreiben, wie die Absichten, die Verfahren und die Ergebnisse der Evaluation, damit die Leser leicht verstehen, was gemacht worden ist, warum dies getan wurde, welche Informationen gewonnen, welche Schlüsse gezogen und welche Empfehlungen gemacht wurden.

Dieses Kriterium gilt für alle Arten von Berichterstattung, unabhängig davon, welche Medien (Text, Film, Projektionsfolien, Tonbänder oder mündliche Berichte) dazu eingesetzt werden. Die Klarheit des Berichts ist eine Bedingung zur Erreichung der Zielgruppen, zur Glaubwürdigkeit des Berichts und schliesslich auch zu dessen Verwendung. Die Klarheit des Berichts bezieht sich auf alle seine Teile, also sämtliche Texte, Illustrationen usw. Sie umfasst ebenso konzise und logische Argumentation, wie auch genau definierte technische Ausdrücke und anschauliche Graphiken sowie übersichtliche Tabellen. Es ist bei der Gestaltung von Evaluationsberichten jedoch darauf zu achten, dass durch die Erhöhung der Klarheit des Berichts die Aussagen des Berichts nicht verfälscht oder zu stark vereinfacht werden. (Joint Committee 1981a: 37-39)

3.4.1.6. Verbreitung des Berichts

Die Resultate der Evaluation sollten so verbreitet werden, dass alle Beteiligten und Betroffenen Zugang dazu haben und die Resultate auch verwenden können.

Bei der Verbreitung des Berichts ist darauf zu achten, dass alle Personen, die in irgendeiner Form an der Evaluation beteiligt oder von dieser betroffen sind, Zugang zum Bericht erhalten. Daraus folgt, dass der Bericht in vielen Fällen öffentlich zugänglich sein muss, da der Kreis der Beteiligten oder Betroffenen oft sehr gross ist. Die an einer Evaluation beteiligten Akteure (primär der Auftraggeber und der Evaluator, aber auch weitere Personen) tragen gemeinsam die Verantwortung, dass dies gewährleistet ist. Der Bericht sollte zudem in einer Form abgefasst sein, die den Bedürfnissen der Adressaten nachkommt. Dies führt oft dazu, dass die Evaluation - angepasst an den Adressatenkreis - in geeigneter Weise zu kommunizieren ist, also dass beispielsweise bei längeren Berichten Zusammenfassungen erstellt werden sollten oder dass ein Anhang für methodisch Interessierte angefügt wird. (Joint Committee 1981a: 40-43)

3.4.1.7. Rechtzeitigkeit des Berichts

Der Bericht sollte rechtzeitig veröffentlicht werden, so dass die Informationen von den Beteiligten und Betroffenen optimal verwendet werden können.

Die Evaluationsberichterstattung sollte zeitlich auf die Bedürfnisse der Adressaten abgestimmt sein. Ist dies nicht der Fall, läuft der Evaluator Gefahr, dass sein Bericht deutlich an Wert verliert. Dies würde eine Verschwendung von Ressourcen darstellen. Um eine zeitgerechte Berichterstattung zu gewährleisten, ist eine sorgfältige Zeitplanung erforderlich. Die Terminplanung einer Evaluation sollte auf die kritischen Terminbedürfnisse ausgerichtet sein, so dass mit Vorteil von den Schlussterminen aus rückwärts geplant wird. Der Evaluator sollte es vermeiden, ohne detaillierte Planung Abgabetermine zu vereinbaren oder zu versprechen. In der Planung ist ausreichend Zeit für Unvorhergesehenes einzurechnen. (Joint Committee 1981a: 44-46)

3.4.1.8. Wirkung der Evaluation

Evaluationen sollten so geplant und durchgeführt werden, dass die Beteiligten und Betroffenen ermuntert werden zu folgen.

Die Wirkung einer Evaluation hängt - neben einer Vielzahl anderer Faktoren - auch von deren Ausgestaltung und Durchführung ab. Dieses Kriterium ist darauf ausgerichtet, inwiefern bei der Planung und Durchführung einer Evaluation dieser Aspekt miteinbezogen wird. Zur Realisierung dieses Anspruchs können verschiedene Massnahmen beitragen, wie Einbezug von Vertretern der wichtigsten 'stakeholders' in die Planung und Durchführung der Evaluation, Berücksichtigung der Interessen bei der Formulierung von Empfehlungen usw. Damit kann der Einfluss einer Evaluation auf die Entscheidungen und Aktivitäten der Adressaten gesteigert werden. Der Evaluator, aber auch der Auftraggeber können massgeblich zur Wirkung der Evaluation beitragen (Joint Committee 1981a: 47-49)

3.4.2. Anwendbarkeit

3.4.2.1. Praktische Verfahren

Die Evaluationsverfahren sollten praktisch sein, so dass Störungen minimal gehalten und die benötigten Informationen beschafft werden können.

Die Wahl der Verfahren zur Durchführung einer Evaluation sollte unter Berücksichtigung der praktischen Umsetzbarkeit erfolgen. Wird versucht, theoretisch wertvolle aber unpraktische Verfahren umzusetzen, werden damit Ressourcen beansprucht, die zu keinen brauchbaren Resultaten führen. Unter Verfahren sind beispielsweise folgende Tätigkeiten zu verstehen: das Vorgehen bei der vertraglichen Vereinbarung mit dem Auftraggeber, die Wahl der Datenquellen und Instrumente, die Erhebung, Aufnahme, Speicherung und Aufbereitung von Daten, deren Analyse, sowie die Form, wie über die Resultate berichtet wird. Diese Verfahren sollten nicht nur in Bezug auf die zu behandelnde Fragestellung, sondern auch hinsichtlich der zur Verfügung stehenden Mittel (beziehungsweise deren Beschränkungen) und des Kontexts der Evaluation kritisch geprüft werden. (Joint Committee 1981a: 52-55)

3.4.2.2. Politische Überlebensfähigkeit

Evaluationen sollten mit Voraussicht auf die unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt werden, so dass deren Kooperation erreicht werden kann und so dass mögliche Versuche irgendeiner dieser Gruppen, die Evaluationsoperationen zu beschneiden oder die Resultate zu verzerren oder zu missbrauchen, vermieden werden können.

Dieses Kriterium erfordert, dass der Evaluator die von einer Evaluation betroffenen oder an dieser beteiligten Personen identifiziert und deren Interessen kennt. Dies umfasst eine Analyse des organisatorischen Kontextes sowie der Machtstrukturen. Werden die relevanten Interessen und deren Träger bei der Ausgestaltung und Durchführung einer Evaluation berücksichtigt, erlaubt dies auch, Versuche, die Evaluation zu tangieren oder ihre Resultate zu verzerren, zu vermeiden oder zumindest zu antizipieren. Damit wird erreicht, dass der Evaluator nicht unvorbereitet mit unerwarteten Reaktionen der Interessengruppen konfrontiert wird. Es lässt sich so auch die Möglichkeit reduzieren, dass Evaluationen verzerrt oder missbraucht werden. Nehmen die Rahmenbedingungen eine Form an, welche die Durchführung der Evaluation massiv tangieren, sollte überprüft

werden, ob nicht besser auf eine Evaluation verzichtet oder eine laufende Evaluation abgebrochen werden sollte. Der Auftraggeber sollte es vermeiden, eine Evaluation zu fordern, die lediglich auf seine eigenen Interessen abstellt. (Joint Committee 1981a: 56-59)

3.4.2.3. Kostenwirksamkeit

Die Evaluation sollte Informationen mit einem Wert produzieren, der die gemachten Aufwendungen rechtfertigt.

Eine Evaluation ist dann kostenwirksam, wenn ihr Nutzen gleich gross oder grösser als ihre Kosten ist.

Die *Kosten* bezeichnen den Wert aller Ressourcen, die durch die Evaluation benötigt werden, unabhängig davon ob sie direkt im Budget der Evaluation erscheinen. Sie umfassen also auch den Zeitbedarf für die Begleitung der Evaluation oder die Kosten, die durch andere Institutionen getragen werden. Die Kosten stellen also die Gesamtheit des (sozialen und monetären) Wertes aller zur Durchführung der Evaluation erforderlichen Ressourcen dar (Vollkosten).

Der *Nutzen* bezeichnet die Summe aller Werte, die durch die Evaluation hervorgebracht wurden. Der Nutzen umfasst also etwa durch die Diagnose der Effektivität beziehungsweise der Ineffektivität politischer Programme, aufgezeigte Einsparungsmöglichkeiten oder Wirkungsoptimierungen, ein verbessertes Verständnis von relevanten Prozessen oder Kenntnisse über die Akzeptanz eines evaluierten Programms. Dabei ist zu beachten, dass die Einschätzung des Nutzens je nach Perspektive unterschiedlich ausfallen kann.

Eine Evaluation ist nun so zu gestalten, dass ihr Kosten-Nutzen-Verhältnis möglichst optimal ausfällt. Sind beispielsweise von unterschiedlichen Untersuchungsanlagen identische Nutzen zu erwarten, ist jene, die mit dem kleinsten Aufwand verbunden ist, zu wählen. Sind verschiedene Alternativen mit den gleich hohen Kosten verbunden, ist jene mit dem höchsten erwarteten Nutzen umzusetzen. Werden verschiedene Designs in Erwägung gezogen, die über ein ebenbürtiges Kosten-Nutzen-Verhältnis verfügen, so ist der Umfang der Evaluation mit dem Auftraggeber auszuhandeln. Bleiben die Kosten in jedem Fall höher, als der zu erwartende Nutzen, ist auf eine Evaluation zu verzichten. (Joint Committee 1981a: 60-62; vgl. dazu das Konzept der 'cost-free evaluation' von Scriven 1983: 42-44 und 1991: 109-110)

3.4.3. Korrektheit

3.4.3.1. Formale Verpflichtung

Den Verpflichtungen der formalen Parteien einer Evaluation sollte von diesen in schriftlicher Form zugestimmt werden, damit diese Parteien zur Einhaltung aller Bedingungen verpflichtet sind oder damit sie diese formell aufs neue verhandeln müssen.

Zu Beginn einer Evaluation ist das Verhältnis zwischen Auftraggeber und Evaluator zu-
meist von gegenseitigem Respekt und Vertrauen geprägt. Dies ist ein optimales Umfeld
zur Festlegung aller wichtigen Rechte und Pflichten beider Parteien in einer schriftlichen
Vereinbarung. Dies kann in Form eines Vertrages, aber auch als Auftragsbestätigung, ge-
genseitig verpflichtenden Offerten, Projektskizzen oder in anderer geeigneter Weise er-
folgen. So sind beide Parteien rechtlich und moralisch an die gemeinsame Abmachung
gebunden. Werden im Verlauf der Zeit Änderungen nötig, besteht die Möglichkeit, die
Bedingungen neu auszuhandeln. Es ist aber keine Partei dazu verpflichtet, einseitige Ent-
scheide der anderen Partei zu akzeptieren. Eine formelle, schriftliche Vereinbarung redu-
ziert die Möglichkeit, dass Missverständnisse unter den Vertragsparteien auftreten. Soll-
ten solche trotzdem bestehen, erleichtert eine schriftliche Festlegung deren Bereinigung.
Die formalen Verpflichtungen sollten unter anderem Festlegungen in folgenden Berei-
chen enthalten: Finanzen, Zeit, Personen, Inhalt sowie Methodik und Vorgehen. Es ist
Wert darauf zu legen, dass die Vereinbarung die Rechte und Pflichten möglichst präzise
festhält und die unterschiedlichen Aspekte abdeckt. (Joint Committee 1981a: 65-69)

3.4.3.2. Deklaration von Interessenkonflikten

Interessenkonflikte - oft unausweichlich - sind offen und aufrichtig zu behandeln, damit sie den Prozess und die Resultate der Evaluation nicht kompromittieren.

Es besteht eine Vielzahl von Konstellationen, in denen sich Evaluatoren mit Interessen-
konflikten konfrontiert sehen. Auch der Evaluator hat Eigeninteressen, welche die Re-
sultate einer Evaluation beeinträchtigen können. Es sei hier nur daran erinnert, dass der
Evaluator mehr oder weniger darauf angewiesen ist, weiterhin Aufträge zu erhalten, dass
er über spezifische philosophische, theoretische oder politische Sichtweisen verfügt und
dass persönliche oder organisatorische Verbindungen bestehen. Konflikte, die das Eva-
luandum selbst betreffen, können dazu führen, dass sich der Evaluator einem Interessen-

konflikt ausgesetzt sieht. Sie können zur Folge haben, dass die Evaluation in ihrem Prozess, ihren Ergebnissen und Interpretationen verzerrt wird. Aufgrund der angedeuteten Vielfalt von möglichen Interessenkonflikten ist es oft nicht möglich, diese zu vermeiden. Vielmehr geht es darum, Wege zu entwickeln, um mit diesen umzugehen, ohne dass die Evaluation darunter leidet. (Joint Committee 1981a: 70-73)

3.4.3.3. Vollständige und aufrichtige Offenlegung

Mündliche und schriftliche Evaluationsberichte sollten in der Darlegung der sachdienlichen Ergebnisse und der Grenzen der Evaluation offen, direkt und ehrlich sein.

Die Berichterstattung, unabhängig davon, in welcher Form diese erfolgt, sollte direkt, offen, vollständig und ehrlich erfolgen. Sie sollte alle relevanten Ergebnisse umfassen. Auslassungen, die nur aufgrund der persönlichen Meinung des Evaluators erfolgen, sollten vermieden werden. Sie sollte klar auf die Zusammenhänge zwischen den Zielen der Evaluation, den erhobenen Informationen und den Resultaten hinweisen. Sie sollte sowohl die Perspektive der Befürworter, wie auch jene der Kritiker des evaluierten Programms miteinschließen. Mögliche Störfaktoren, welche die Aussagekraft der Evaluation beeinträchtigen, sind offen darzulegen und deren Auswirkungen sind klar zu benennen. Es ist zu vermeiden, Evaluationsberichte zu publizieren, die lediglich die Eigeninteressen des Evaluators, des Auftraggebers oder des Programmpersonals reflektieren. Andererseits ist aber auch darauf zu achten, dass der Geltungsanspruch der Evaluation nicht zu stark eingegrenzt wird, so dass sie die ihr zustehende Glaubwürdigkeit verlieren könnte. (Joint Committee 1981a: 74-76)

3.4.3.4. Herstellung von Öffentlichkeit

Die formellen Parteien einer Evaluation sollten das Recht der Öffentlichkeit auf Information respektieren und sicherstellen, wobei die Begrenzungen durch andere verwandte Prinzipien und Regelungen, wie die öffentliche Sicherheit und das Recht auf Privatsphäre, berücksichtigt werden müssen.

Unter den Begriff der Öffentlichkeit fallen in diesem Zusammenhang alle Personen oder Organisationen, die als 'stakeholder' einer Evaluation bezeichnet werden können, also alle Personen, die direkt oder indirekt an dieser beteiligt sind oder von dieser betroffen sind oder sein könnten (vgl. dazu ausführlicher oben, Kapitel 3.4.1.1.). Das Recht auf Öffent-

lichkeit wird dadurch begrenzt, dass die öffentliche Sicherheit, wie auch das Recht des Individuums auf Privatsphäre gewahrt bleibt. Erhalten die 'stakeholders' keinen Zugang zur Evaluation, besteht für diese keine Möglichkeit, Schwächen der Planung und Durchführung der Evaluation zu erkennen oder die Ergebnisse der Evaluation in konstruktiver Weise zu verwenden. Es entsteht daraus die Gefahr, dass ausgeschlossene Personen zu Opfern einer Evaluation werden, ohne dass sie sich dagegen wehren könnten. Evaluationen sind in einer Weise zu planen und durchzuführen, dass sie auch einer kritischen Begutachtung standhalten können. Ist dies der Fall, besteht seitens des Evaluators keinerlei Veranlassung, die Evaluation geheim zu halten. Üblicherweise steht die Regelung der Publikation einer Evaluation zumindest teilweise unter Kontrolle des Auftraggebers. Die Freiheit des Evaluators ist hier eingeschränkt. Deshalb tragen der Evaluator und der Auftraggeber gemeinsam die Verantwortung, dass dieses Kriterium erfüllt wird. (Joint Committee 1981a: 77-80)

3.4.3.5. Schutz individueller Menschenrechte

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

Menschliche Individuen verfügen über persönliche Rechte, die sich auf Gesetze abstützen oder durch ethische Praxis, 'common sense' und Höflichkeit anerkannt sind. Der Evaluator hat zu vermeiden, dass Individuen durch die Evaluation darin beschnitten werden. Ebenso hat der Evaluator darauf zu achten, dass die Evaluation dem Wohlergehen der Menschen nicht abträglich ist. Bei der Planung, wie auch bei der Durchführung einer Evaluation ist immer darauf Rücksicht zu nehmen, dass die Rechte und das Wohlergehen der Menschen nicht tangiert werden. So sind diejenigen Personen, die in eine Evaluation involviert werden, über diesen Umstand zu informieren. Die Auswirkungen der Evaluation sind gründlich darzulegen. Eingegangene Vereinbarungen, beispielsweise die Zusicherungen zur Wahrung der Anonymität von Gesprächspartnern, sind einzuhalten. Der Auftraggeber sollte es unterlassen, den Evaluator zu Entscheiden zu drängen, welche die angeführten Leitlinien tangieren könnten. Führen Evaluationen zu wohlbegründeten Folgerungen, welche das Wohlergehen einzelner Personen gefährden, ist vorsichtig abzuwägen, inwiefern die Verbreitung derartiger Folgerungen gerechtfertigt ist. (Joint Committee 1981a: 81-85)

3.4.3.6. Menschliche Interaktionen

Evaluatoren sollten in den Interaktionen mit anderen Personen im Zusammenhang mit der Evaluation die Würde und den Wert der Menschen respektieren.

Evaluatoren sollten es vermeiden, die Würde und die Selbstachtung der Personen, mit denen sie im Verlauf der Evaluation in Kontakt treten, in irgendeiner Form zu verletzen. Weiter sollte durch geeignetes Verhalten vermieden werden, dass sich gegenüber der Evaluation feindliche Gefühle entwickeln. Dies ist nicht nur ein Gebot der Menschenwürde, sondern hat durchaus auch praktische Gründe. Personen, die sich in ihrer Würde oder in ihrer Selbstachtung angegriffen fühlen, verlieren nicht nur an kreativem Potential, sondern verhalten sich auch oft so, dass der Evaluator in seinen Aktivitäten eingeschränkt wird. Deshalb ist es notwendig, die kulturellen und sozialen Werte der involvierten Personen zu kennen und zu verstehen. Es ist zudem zweckmässig darauf zu achten, welche Bedeutung die Evaluation für die einzelnen Personen hat. (Joint Committee 1981a: 86-89)

3.4.3.7. Ausgewogene Berichterstattung

Evaluationen sollten in der Präsentation der Stärken und Schwächen des Evaluandums vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder benannt werden können.

Ausgewogenheit in der Präsentation der Stärken und Schwächen eines Evaluandums bedeutet nicht, darauf zu achten, dass eine gleiche Anzahl von Stärken und Schwächen genannt wird, sondern zielt darauf ab, dass eine komplette und faire Einschätzung erfolgt, indem die Stärken und Schwächen aufgezeigt werden. Auch wenn das primäre Ziel einer Evaluation eine Schwachstellenanalyse bildet, entbindet dies nicht von der Pflicht, auch positive Aspekte ausfindig zu machen und diese zu benennen. Einen Grund dafür bildet der Umstand, dass es manchmal möglich ist, bestehende Schwächen mit vorhandenen Stärken zu korrigieren. Weiter ist zu bedenken, dass die Korrektur von Schwächen des Evaluandums dazu führen kann, dass Stärken geschwächt werden. Es ist deshalb nützlich, den Bericht vor Abschluss der Evaluation durch externe Personen mit möglicherweise anderen Einstellungen hinsichtlich der Darlegung der positiven und negativen Punkte durchsehen zu lassen. Wenn es aus irgendwelchen Gründen (wie zum Beispiel aus zeitlichen oder finanziellen Restriktionen) nicht möglich ist, gewisse Daten zu erheben, sind diese Lücken deutlich kenntlich zu machen. Der Auftraggeber sollte Interventionen un-

terlassen, die eine ausgewogene Berichterstattung in Frage stellen könnten. (Joint Committee 1981a: 90-92)

3.4.3.8. Finanzielle Verantwortlichkeit

Die Zuweisung und Ausgabe von Ressourcen durch den Evaluator sollte unter Berücksichtigung fundierter und verantwortlicher Verfahren und auch anderweitig klug und ethisch verantwortlich erfolgen.

Die finanzielle Verantwortlichkeit des Evaluators bezieht sich auf alle Zuweisungen und Ausgaben im Rahmen der Evaluation. Er ist dafür verantwortlich, dass die im Evaluationskontrakt enthaltenen Vereinbarungen eingehalten werden. Die ihm zur Verfügung gestellten Mittel, wie auch die Zuweisungen und Ausgaben sind sorgfältig zu erfassen und zu verwalten. Damit lässt sich die Gefahr, dass Mittel fahrlässig oder unvernünftig eingesetzt werden, deutlich reduzieren. Der Evaluator ist gehalten, die Vorschriften, Vereinbarungen und die Regeln der Buchhaltung peinlich genau einzuhalten, da er sich sonst der Gefahr aussetzt, dass finanzielle Aspekte als Argumente gegen die Evaluation verwendet werden könnten. Die Aufführung einer finanziellen Übersicht im öffentlichen Schlussbericht erhöht die Vertrauenswürdigkeit einer Evaluation. Die Effizienz einer Evaluation sollte jedoch nur soweit gesteigert werden, dass nicht die Qualität der Evaluation tangiert wird. (Joint Committee 1981a: 93-95)

3.4.4. Genauigkeit

3.4.4.1. Identifikation des Gegenstandes

Der zu evaluierende Gegenstand sollte ausreichend untersucht werden, so dass die in der Evaluation berücksichtigte(n) Form(en) des Gegenstandes klar identifiziert werden können.

Das Objekt einer Evaluation, sei es nun eine Massnahme, ein Programm oder eine Organisation sind genau zu untersuchen. Es ist darauf zu achten, dass sich das Evaluandum in unterschiedlichen zeitlichen und räumlichen Kontexten verschieden ausgestalten kann. Die Beschreibung des Evaluandums sollte klar machen, was präzise untersucht wurde. Dies ermöglicht den Adressaten auch einen Vergleich mit anderen Evaluationsgegenständen. Weiterhin hilft eine genaue Untersuchung des Evaluandums, Zusammenhänge zwischen diesem und seinen Wirkungen festzustellen und bisher unbeachtete Nebenwirkun-

gen aufzufinden. Dabei ist besonders darauf zu achten, inwiefern sich die Realisation des Evaluandums von der Planung unterscheidet, also inwiefern sich zwischen der vorgesehenen Form des Evaluandums und dessen Ausgestaltung in der Implementation Diskrepanzen zeigen. (Joint Committee 1981a: 99-103)

3.4.4.2. Kontextanalyse

Der Kontext des Evaluandums sollte ausreichend detailliert untersucht werden, so dass mögliche Beeinflussungen des Evaluandums identifiziert werden können.

Unter dem Kontext des Evaluandums verstehen wir die Kombination aller Rahmenbedingungen, welche den zu evaluierenden Gegenstand umgeben. Dazu zählen beispielsweise die geographische Lokalisierung, die institutionelle Einbindung, das soziale und politische Klima, die personellen Charakteristiken der massgeblich am Programm Beteiligten, die Struktur der Politikarena, benachbarte (und konkurrenzierende) staatliche (aber auch private) Programme oder die wirtschaftlichen Rahmenbedingungen. Diese und weitere Kontextfaktoren sind ausreichend genau zu untersuchen, um die Evaluation in Relation zu diesen planen, durchführen und vermitteln zu können. Dieses Kontextwissen ist notwendig, um die Evaluation realistisch in bezug auf die bestehenden Möglichkeiten und abgestimmt auf den Kontext ausgestalten zu können. Weiter ist zu beachten, dass Kontextfaktoren oft einen massgeblichen Einfluss (in positiver oder negativer Weise) auf die Auswirkungen eines Programms haben können. Zudem erlaubt eine fundierte Kontextanalyse den Adressaten, die Evaluation zu interpretieren und diesen Kontext mit anderen Kontexten vergleichen zu können, um die Übertragbarkeit der Resultate einzuschätzen. Es sollte vermieden werden, den Kontext zu eng zu definieren. Andererseits ist aber auch davor zu warnen, die Kontextanalyse zu ausführlich anzulegen, so dass die Analyse des eigentlichen Gegenstandes nicht ausreichend gründlich erfolgen kann. (Joint Committee 1981a: 104-106)

3.4.4.3. Beschreibung von Zielen und Vorgehen

Die Ziele und das Vorgehen der Evaluation sollten ausreichend genau beobachtet und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.

Die mit einer Evaluation verfolgten Absichten, wie auch die Vorgehensweise bei der Evaluation sind genau zu beobachten und zu beschreiben. Bei der Beobachtung und Beschreibung der Evaluationsziele ist speziell auf abweichende Vorstellungen der Beteilig-

ten und Betroffenen zu achten. Sind solche vorhanden, ist dies bei der Ausgestaltung der Evaluation zu berücksichtigen und bei der Berichterstattung entsprechend zu vermerken. Die Beobachtung und Beschreibung des Vorgehens umfasst nicht nur die detaillierte Beschreibung der Erhebung, der Organisation, der Analyse und der Berichterstattung, sondern ebenso die ursprünglich vorgesehene Vorgehensweise, die erfolgten Änderungen am Untersuchungsplan, wie auch die effektiv eingesetzten Methoden und Verfahren. Die Beobachtung und Beschreibung des Vorgehens ist aus verschiedenen Gründen notwendig: Sie dienen dem Evaluator selbst zur Einschätzung der eigenen Vorgehensweise und bieten ihm Orientierungshilfe. Sie eröffnen den Adressaten die Möglichkeit, die Evaluation nachvollziehen und beurteilen zu können. Nur so können die Ergebnisse einer Evaluation auch richtig eingeordnet und bewertet werden. Nur so erhält eine Evaluation ausreichende Transparenz. Dies gilt nicht nur für die 'stakeholders' eines Evaluandums, sondern auch für andere Evaluatoren, die an den spezifischen Vorgehensweisen interessiert sind. Weiter bildet diese Beschreibung eine wichtige Grundlage zur Evaluierung einer Evaluation. Ein Verzicht auf die Deklaration der Absichten und der Vorgehensweise kann eine Evaluation zu Unrecht gegen berechtigte Kritik immunisieren oder nichtgerechtfertigte Einwände aufkommen lassen. (Joint Committee 1981a: 107-111)

3.4.4.4. Verlässliche Informationsquellen

Die Informationsquellen sollten ausreichend detailliert beschrieben werden, so dass die Adäquatheit der Informationen eingeschätzt werden kann.

In einer Evaluation können verschiedenste Informationsquellen genutzt werden, wie etwa Einzelpersonen und Gruppen, Dokumente, Filme, Tonbandaufnahmen und so weiter. Die Nutzung unterschiedlicher Informationsquellen erlaubt einen Vergleich der enthaltenen Informationen. Evaluatoren sollten die Informationsquellen, die Auswahlkriterien, die Instrumente zur Aufarbeitung der Informationen, die Bestimmung von Stichproben für grössere Grundgesamtheiten usw. genau beschreiben. Diese Beschreibungen sollten es dem Adressaten erlauben, die Adäquatheit der Informationen zu beurteilen. Eine unvollständige oder fehlende Beschreibung der Quellen kann die Glaubwürdigkeit der Evaluation gefährden. Sie kann auch zur Fehlannahme durch die Adressaten führen, dass die Folgerungen und Empfehlungen gut informiert erfolgten, auch wenn dies nicht - oder nur eingeschränkt - der Fall war. Der Schlussbericht sollte nicht nur eine Beschreibung, sondern auch eine Qualifizierung der Informationen enthalten. (Joint Committee 1981a: 112-115)

3.4.4.5. Valide Messung (a) bzw. Glaubwürdigkeit (b) und Übertragbarkeit (c)

(a) Die Instrumente und Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt ist.

Die Validität einer Messung kann immer nur in Relation zu den mit ihr verfolgten Absichten beurteilt werden. Die Validierung von empirischen Erfassungen sollte verschiedene Komponenten enthalten:

- eine detaillierte inhaltliche Beschreibung der zu erfassenden Konstrukte,
- eine Analyse, was ein bestimmtes Instrument oder Verfahren zur Erfassung beitragen kann,
- eine detaillierte Beschreibung der spezifischen Umsetzung eines Instrumentes in der Evaluation,
- eine ausreichende Begründung zur Verwendung des spezifischen Instrumentes und
- eine umfassende Einschätzung der Validität des Einsatzes des Instrumentes und der darauf basierenden Interpretationen.

Dabei ist zu beachten, dass die Validität massgeblich davon abhängt, wie die Erhebung im konkreten Fall umgesetzt wurde. Die Validität ist der zentrale Aspekt jeder empirischen Erfassung. Die Reliabilität nicht-valider Instrumente ist irrelevant. Sind empirische Erfassungen als wenig valide einzuschätzen, oder fehlten die entsprechenden Mittel, um die Validität der Messung genauer zu überprüfen, ist dies in Berichten deutlich hervorzuheben. (Joint Committee 1981a: 116-119)

(b) Die Rekonstruktionen durch die Evaluation sollten eine korrekte Abbildung der Konstruktionen der Akteure darstellen.

Das alternative Paradigma geht von einer sozialen Konstruktion der Wirklichkeit aus. Dies steht im Gegensatz zum traditionellen Paradigma, in dem eine objektive Wahrheit zugrunde gelegt wird. Daher ist die konventionell verstandene, interne Validität im Rahmen des konstruktivistischen Paradigmas nicht adäquat. Die Glaubwürdigkeit tritt an die Stelle der internen Validität. Anstelle des Ziels im traditionellen Paradigma, die Realität möglichst richtig zu erfassen, ist hier von Interesse, ob die Rekonstruktionen durch die

Evaluation eine gültige Abbildung der Konstruktionen der Akteure darstellen. Ist dies der Fall, gilt die Rekonstruktion als glaubwürdig. Zur Erreichung dieser Zielsetzungen bestehen verschiedenste Hilfsmittel. So steigert sich die Glaubwürdigkeit, wenn der Evaluator über längere Zeit hinweg Gelegenheit erhält, die Konstruktionen der Akteure kennenzulernen, zu erfassen und zu verstehen. Dadurch erhöht sich die Tiefe der Erfassung und mögliche Desinformationen oder Verzerrungen können aufgedeckt werden. Diesem Ziel kann auch der Einsatz unterschiedlicher Instrumente zur Erhebung oder Auswertung von Daten dienen. Weiter ist es von Vorteil, zu Beginn der Evaluation, noch bevor andere Evaluationsaktivitäten unternommen werden, das Vorverständnis des Evaluators festzuhalten. So kann im Evaluationsprozess immer wieder überprüft werden, inwiefern sich die Rekonstruktionen vom Vorverständnis unterscheiden. Treten keine Änderungen ein, ist Vorsicht geboten, da das Vorverständnis des Evaluators die Evaluation stark beeinflussen könnte. Vorzugsweise sollte dieser Prozess durch eine aussenstehende Person begleitet werden. Die Rekonstruktionen des Evaluators können auch überprüft werden, indem zusätzliche Fallstudien erarbeitet werden, welche die in den Rekonstruktionen enthaltenen Fehlerhaftigkeiten - sofern vorhanden - erkennen lassen. Die wohl wichtigste Möglichkeit, die Glaubwürdigkeit einer Evaluation zu erhöhen, besteht darin, die Akteure dazu einzuladen, die Rekonstruktion ihrer Konstruktionen kritisch zu durchleuchten. Dadurch können sich auch weitere Erkenntnisse ergeben, die ansonsten unerkannt geblieben wären. (Guba/Lincoln 1989: 236-241)

(c) Eine präzise, detaillierte und vollständige Beschreibung (thick description) des untersuchten Kontexts sollte es erlauben zu entscheiden, ob die Aussagen der Evaluation auf andere Kontexte übertragbar sind.

Im Gegensatz zu Evaluationen, die dem traditionellen Paradigma folgen, geht das alternative Paradigma davon aus, dass es nicht die Aufgabe des Evaluators ist, die Übertragbarkeit seiner Evaluation auf andere Kontexte einzuschätzen. Seine Aufgabe ist es lediglich, die Informationen bereitzustellen, damit dies für Aussenstehende möglich ist. Durch die Erarbeitung einer 'thick description' werden die relevanten Informationen dafür bereitgestellt. Eine solchermassen präzise, detaillierte und möglichst umfassende Beschreibung sollte nicht nur das Evaluandum selbst, sondern auch dessen Kontext erfassen (Lincoln/Guba 1985: 362). Es sollte aufgrund der 'thick description' möglich sein, alle relevanten Faktoren zu erkennen, damit beurteilt werden kann, inwiefern sich ein anderes Programm in einem anderen Kontext von dem Evaluandum und seinem Kontext unterscheidet. (Guba/Lincoln 1989: 241-242)

3.4.4.6. Reliable Messung (a) bzw. Verlässlichkeit (b)

(a) Die Instrumente und Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Zuverlässigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt ist.

Reliable Verfahren zur Datenerhebung sind solche, welche konsistente Hinweise auf das zu untersuchende Merkmal erlauben. Jede Form der empirischen Erfassung ist bis zu einem gewissen Grad Fehlerquellen ausgesetzt, die den Erhebungsvorgang beeinflussen und möglicherweise zu Schwankungen in den Resultaten führen können. Die Reliabilität eines Erhebungsverfahrens kann nicht generell, sondern nur im konkreten Untersuchungskontext eingeschätzt werden. Ein Aspekt, der bei der Beurteilung der Reliabilität zu berücksichtigen ist, betrifft die Stabilität: Das heisst, dass sich die Messresultate nicht auf eine einmalige, spezifische Situation beziehen sollten. Ein weiterer Aspekt betrifft die Unabhängigkeit von Personen, die an der Erhebung beteiligt sind. Hinzu kommt die Frage, inwiefern die erhobenen Informationen eine ausreichend zuverlässige Erfassung des Evaluandums als Ganzes zulassen. Unter keinen Umständen sollte jedoch die Reliabilität einer Evaluation auf Kosten der Validität erhöht werden, weil erst die Einhaltung beider Erfordernisse eine qualitativ hochstehende Messung sicherstellt. (Joint Committee 1981a: 120-123)

(b) Der Evaluationsprozess sollte von aussen erkennbar sein, damit auch die getroffenen Entscheidungen und Interpretationen in der Evaluation untersucht und beurteilt werden können.

Im traditionellen Paradigma sind Veränderungen der Erhebungsinstrumente unter allen Umständen zu vermeiden, weil dadurch die Erfassung an Reliabilität verlieren würde. Derartige Veränderungen sind bei einer, an der alternativen Erkenntnistheorie orientierten Evaluation integraler Bestandteil des Konzepts, da sich die Untersuchungsanlage an sich verändernde Umstände und Konstellationen anpassen soll und aufgrund neuer Erkenntnisse immer wieder Änderungen auftreten müssen (emergentes Design). Umso wichtiger wird hier deshalb der Umstand der Verlässlichkeit. Darunter verstehen wir die Möglichkeit, die getroffenen Entscheidungen und vorgenommenen Interpretationen *von aussen* nachvollziehen zu können. Dies bedingt, dass Entscheide und Interpretationen ausführlich dargelegt und ausreichend begründet werden müssen. Nur so sind die Adressaten einer Evaluation in der Lage, den Evaluationsprozess zu erkennen, zu verstehen und einzu-

schätzen. Um eine ausreichende Dokumentation des Evaluationsprozesses bereitstellen zu können, sind die Evaluatoren dazu verpflichtet, alle relevanten Entscheidungen und Interpretationen und die dazugehörigen Begründungen in einem Journal (oder in einer anderen geeigneten Form) festzuhalten. Eine Beschreibung des Evaluationsprozesses sollte im öffentlichen Bericht enthalten sein. (Guba/Lincoln 1989: 242)

3.4.4.7. Systematische Datenüberprüfung

Die in einer Evaluation gesammelten, bearbeiteten und angegebenen Daten sollten überprüft und korrigiert werden, so dass die Resultate der Evaluation nicht fehlerhaft sind.

Eine systematische Datenüberprüfung umfasst alle Schritte, die unternommen werden können, um fehlerhafte Angaben zu vermeiden. Bei der Erhebung, der Erfassung, der Aufbereitung, der Analyse und der Interpretation bestehen eine Vielzahl möglicher Fehlerquellen. Diese reichen vom lapidaren Tippfehler bei der Datenerfassung bis zur Fehlinterpretation von erhobenen Daten. Evaluatoren sollten dafür besorgt sein, dass möglichst wenige Fehler dieser Art entstehen und wenn es zu solchen kommt, dass sie wenn immer möglich erkannt und korrigiert werden. Fehlerhaftigkeiten können nicht nur zu falschen Schlüssen führen, sondern sind - auch wenn es sich dabei um unwesentliche Fehler handelt - in der Lage, die ganze Evaluation zu diskreditieren und angreifbar zu machen. Evaluatoren sollten nicht von der Annahme ausgehen, dass hochqualifizierte und engagierte Personen keine Fehler machen. Eine sichere Prämisse ist die, sich an 'Murphy's Law' zu halten und damit zu rechnen, dass alles schief läuft. Üblicherweise häufen sich Fehlerquellen, wenn unter grossem Zeitdruck gearbeitet werden muss oder wenn notwendige Informationen nicht zur Verfügung stehen. Es sollte darauf geachtet werden, dass die primäre Form der Erfassung von Daten für Überprüfungs Zwecke weiterhin erhalten bleiben. Übertragungsfehler können reduziert werden, indem die Zahl der Übertragungen (bzw. der Arbeitsschritte) möglichst klein gehalten wird. Es ist unabdingbar, dafür zu sorgen, dass die erhobenen, erfassten, aufbereiteten, analysierten und interpretierten Daten jeweils einer systematischen Datenüberprüfung unterzogen werden. (Joint Committee 1981a: 124-126)

3.4.4.8. Analyse quantitativer Information

Quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, um haltbare Interpretationen sicherzustellen.

Quantitative Informationen bestehen aus Fakten und Behauptungen, die durch Zahlen repräsentiert werden. Mit quantitativer Analyse wird der Vorgang bezeichnet, bei dem derartige Informationen gesammelt, zusammengestellt, aufbereitet und validiert werden, um bestimmte Fragen zu beantworten. Viele Evaluationen verwenden quantitative Analysen in der einen oder anderen Form. Um haltbare Interpretationen sicherzustellen sind die entwickelten Regeln der quantitativen Methodik einzuhalten. Weiter ist darauf zu achten und kritisch zu reflektieren, ob die eingesetzten Verfahren in der Lage sind, die gesuchten Antworten auf die vorliegenden Fragen zu geben. Die Wahl des konkreten Verfahrens und die Art und Weise der spezifischen Umsetzung hat nicht nur hinsichtlich der zu bearbeitenden Fragestellung, sondern hat auch unter Berücksichtigung der Datenlage zu erfolgen. Das Verfahren und die Umsetzung muss sich an Fragestellung und Datenlage anpassen, um haltbare Interpretationen zu gewährleisten. Schwächen der Analysen, etwa hinsichtlich der Verletzung der verfahrensspezifischen Prämissen, sind offen darzulegen und in bezug auf ihre Implikationen kritisch zu diskutieren. Es ist von Vorteil bei der Durchführung quantitativer Analyse eine möglichst breite Datenbasis bereitzustellen. In jedem Fall ist es zu vermeiden, aufgrund von Überlegungen zur Datenerhebung und Datenanalyse die Fragestellung zu reduzieren oder diese in unzulässiger Weise zu vereinfachen. Weiter ist in diesem Zusammenhang zu beachten, dass nicht für jede Evaluation statistische Analysen erforderlich sind. Ebenso gilt es darauf zu achten, dass die Einhaltung (übertriebener) methodischer Striktheit nicht dazu führt, dass die Evaluation an Relevanz verliert. (Joint Committee 1981a: 127-129)

3.4.4.9. Analyse qualitativer Information

Qualitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, um haltbare Interpretationen sicherzustellen.

Qualitative Informationen bestehen aus Fakten und Interpretationen, die nicht in numerischer, sondern in sprachlicher Form vorliegen. Informationen dieser Art können aus einer Vielzahl verschiedener Quellen stammen: Strukturierten oder unstrukturierten Interviews, teilnehmenden oder nichtteilnehmenden Beobachtungen, Anhörungen, Dokumenten und Aufnahmen aller Art. Unter qualitativer Analyse verstehen wir Vorgänge, wie die Zu-

sammenstellung, Analyse und Interpretation qualitativer Informationen, die zur Bearbeitung der Fragestellung dienen. Normalerweise erfolgt dabei keine Quantifizierung der Informationen. Währenddem qualitative Informationen zwar auch in quantitativer Weise ausgewertet werden können, wird bei einer qualitativen Analyse versucht, mehr Tiefe und Perspektive zu erreichen, als dies allein aufgrund einer quantitativen Analyse möglich wäre. Im Gegensatz zu quantitativ angelegten Analysen ist es hier zumeist nicht möglich, im voraus die Beobachtungsprotokolle, die Kategorisierungen und Methoden zur Zusammenfassung festzulegen. Qualitative Analysen erfordern oft einen interaktiven wie auch iterativen Prozess, in dem der Evaluator oft mehrfach an relevante Akteure herantritt, um weitere Informationen zu beschaffen oder um Daten, Interpretationen oder Folgerungen einer Überprüfung zu unterziehen. Auch bei qualitativen Analysen ist es wichtig, die Wahl des Verfahrens unter Massgabe der zu bearbeitenden Fragestellungen, wie auch der Datenlage zu treffen. Mögliche Schwächen der Analyse sind offen darzulegen und hinsichtlich ihrer Konsequenzen zu diskutieren. Eine qualitative Analyse ist zu beenden, wenn die Quellen ausgeschöpft sind und wenn vermehrt Redundanzen zu beobachten sind. Bei qualitativen Analysen ist es von Vorteil, wenn die gleiche Information von mehreren Analysten interpretiert wird, um die Verlässlichkeit der Interpretationen zu erhöhen. (Joint Committee 1981a: 130-134)

3.4.4.10. Gerechtfertigte Folgerungen

Die in einer Evaluation erfolgten Folgerungen sollten explizit begründet sein, so dass die Beteiligten und Betroffenen diese einschätzen können.

Die Folgerungen einer Evaluation, die Bewertungen und Empfehlungen umfassen, müssen begründet erfolgen. Sie müssen aufgrund logischer Argumentation und geeigneter Informationen gezogen werden, damit sie gegenüber Beteiligten und Dritten vertretbar sind. Um ausreichend gerechtfertigt zu sein, müssen die Folgerungen zusammen mit einer Beschreibung der Evaluationsverfahren, der Informationen und der zugrundeliegenden Annahmen dargestellt werden. Ebenso sollte eine Diskussion alternativer Folgerungen enthalten sein, wobei begründet sein sollte, weshalb letztere verworfen wurden. Bei der Etablierung von Prämissen sollte darauf geachtet werden, dass sie nicht Annahmen enthalten, die von relevanten Akteuren nicht geteilt werden.

Dieses Kriterium ist aus zwei Gründen wichtig. Erstens könnten die Folgerungen fehlerhaft sein, so dass die Adressaten aufgrund der Folgerungen der Evaluation ungeeignete Massnahmen ergreifen könnten. Zweitens erhöht eine Ausrichtung auf dieses Kriterium die Überzeugungskraft der Folgerungen, da den Adressaten Informationen in ausreichen-

dem Masse zur Verfügung gestellt werden, um die Berechtigung der Folgerungen überprüfen zu können. Das soll nun aber nicht bedeuten, dass Evaluationen nicht auch Vermutungen, Schätzungen und Interpretationen enthalten sollen. Diese sind jedoch deutlich als solche zu deklarieren, damit die Adressaten nicht davon ausgehen, dass Aussagen dieser Art ausreichend belegt wären. (Joint Committee 1981a: 135-137)

3.4.4.11. Objektive Berichterstattung (a) bzw. Bestätigbarkeit (b)

(a) Die Evaluationsverfahren sollten über Vorkehrungen verfügen, um die Ergebnisse und Berichte einer Evaluation gegen Verzerrungen durch persönliche Gefühle und Vorlieben irgendeiner Partei in der Evaluation zu schützen.

Evaluationsberichte sind soweit objektiv, als sie auf unparteiisch zusammengestellten Informationen beruhen und nicht dazu neigen, einseitige Positionen zu vertreten. Die erste Bedingung ist schwer zu erfüllen. Persönliche Beurteilungen fließen in die Auswahl von Evaluationskriterien und Instrumenten ein. Auch die Wahl von Verfahren zur Erhebung und Auswertung des Datenmaterials oder die Formulierung von Empfehlungen enthalten subjektive Elemente. Viele der hier dargelegten Kriterien sind darauf ausgerichtet, die erste Bedingung zu erfüllen. Das hier zu diskutierende Kriterium bezieht sich aber auf die zweite Bedingung. Es soll Hilfestellung dazu bieten, wie Berichte zu entwickeln und zu präsentieren sind, damit sie möglichst wenig einseitig ausfallen und damit die Adressaten auf mögliche Verzerrungen aufmerksam gemacht werden. Evaluationsberichte können in unterschiedlicher Weise nicht-objektiv ausfallen. So kann es zu Einseitigkeiten kommen, da der Evaluator nicht alle relevanten Perspektiven in den Bericht einbezogen hat. Der Ursprung für eine einseitige Berichterstattung kann jedoch auch im Evaluationsauftrag enthalten sein, insbesondere dann, wenn Evaluationen in der Absicht vergeben werden, um beispielsweise zu täuschen, um zu vertuschen, um zu beschönigen oder um Entschiede zu bestätigen, die bereits auf anderer Grundlage getroffen worden sind. Einseitige Berichte können aber auch aus Unbekümmertheit oder Unwissenheit des Evaluators zustande kommen. Wieder andere Berichte sind verzerrt, weil sich der Evaluator dem Druck des Auftraggebers oder anderer Gruppen von Beteiligten oder Betroffenen nicht widersetzen konnte. Aber auch der kontinuierliche Austauschprozess mit 'stakeholders' kann dazu führen, dass der Bericht unausgeglichen ausfällt. Um derartige Verzerrungen zu vermeiden, ist es empfehlenswert, mit dem Auftraggeber bei der Aushandlung des Evaluationskontraktes die entsprechenden Schritte festzulegen, um Einseitigkeit in der Berichterstattung zu vermeiden. Weiter kann die Objektivität des Berichts erhöht werden, indem wenn immer möglich abweichende Positionen dargestellt werden. Insbesondere ist

darauf zu achten, dass nicht nur die Perspektiven der direkt Beteiligten, sondern auch jene der Betroffenen ausreichend Berücksichtigung finden. Die Unabhängigkeit des Evaluators kann durch externe Überprüfungen und durch Vermeidung einer zu starken Bindung der Evaluatoren an eine Gruppe der 'stakeholders' erhöht werden. Weiter ist es von Vorteil, den Bericht durch kritische 'stakeholders' begutachten zu lassen. Der Evaluator sollte die Kompetenz, Berichte zu publizieren nicht an andere abgeben. Er sollte weiter darauf achten, dass er bei mündlichen Präsentationen miteinbezogen wird. (Joint Committee 1981a: 138-141)

(b) Die Konstruktionen der Evaluation sollten auf ihre Ursprünge und Wurzeln im empirischen Material auch von aussen rückführbar sein.

Das alternative Paradigma lehnt die Objektivität als Ziel einer Evaluation ab, da eine konstruktivistische Erkenntnistheorie keine objektive Wahrheit, sondern lediglich subjektive Konstruktionen kennt. Um erkennen zu können, ob die Konstruktionen der Evaluation, also ihre Beschreibungen und Interpretationen auf Konstruktionen der Akteure rückführbar sind, ist dieser Bezug im Evaluationsbericht in transparenter Weise herzustellen. Die Interpretationen, Folgerungen und Empfehlungen müssen auf die 'Rohdaten' zurückführbar sein. Um dies zu gewährleisten, muss die Evaluation sowohl die Rohdaten, wie auch die Prozesse, die zu deren Verdichtung, Verwesentlichung oder Interpretation eingesetzt wurden, ausführlich darstellen. Nur so kann der Evaluationsprozess von aussen erkennbar sein. Und nur so ist es möglich, in einer Evaluation Bestätigbarkeit zu erreichen. (Guba/Lincoln 1989: 242-243)

3.4.5. Die eingesetzten Kriterien im Überblick

1. Nützlichkeit

1.1. Identifikation der beteiligten Gruppen

Die Beteiligten oder Betroffenen einer Evaluation sollten identifiziert werden, damit deren Interessen berücksichtigt werden können.

1.2. Glaubwürdigkeit des Evaluators

Die Personen, welche die Evaluation durchführen, sollten sowohl vertrauenswürdig wie auch kompetent sein, damit die Ergebnisse der Evaluation ein Höchstmass an Glaubwürdigkeit und Akzeptanz erreichen.

1.3. Informationsrahmen und -auswahl

Die beschafften Informationen sollten von einem Umfang und einer Auswahl sein, welche die Behandlung sachdienlicher Fragen zum Evaluandum ermöglichen und gleichzeitig die Interessen der Beteiligten und Betroffenen berücksichtigen.

1.4. *Bewertende Interpretation*

Die für die Interpretation der Ergebnisse verwendeten Perspektiven, Verfahren und Begründungen sollten sorgfältig beschrieben werden, damit die Grundlagen für die Werturteile klar werden.

1.5. *Klarheit des Berichts*

Der Evaluationsbericht sollte das evaluierte Objekt und den Kontext ebenso beschreiben, wie die Absichten, die Verfahren und die Ergebnisse der Evaluation, damit die Leser leicht verstehen, was gemacht worden ist, warum dies getan wurde, welche Informationen gewonnen, welche Schlüsse gezogen und welche Empfehlungen gemacht wurden.

1.6. *Verbreitung des Berichts*

Die Resultate der Evaluation sollten so verbreitet werden, dass alle Beteiligten und Betroffenen Zugang dazu haben und die Resultate auch verwenden können.

1.7. *Rechtzeitigkeit des Berichts*

Der Bericht sollte rechtzeitig veröffentlicht werden, so dass die Informationen von den Beteiligten und Betroffenen optimal verwendet werden können.

1.8. *Wirkung der Evaluation*

Evaluationen sollten so geplant und durchgeführt werden, dass die Beteiligten und Betroffenen ermuntert werden zu folgen.

2. **Anwendbarkeit**

2.1. *Praktische Verfahren*

Die Evaluationsverfahren sollten praktisch sein, so dass Störungen minimal gehalten und die benötigten Informationen beschafft werden können.

2.2. *Politische Überlebensfähigkeit*

Evaluationen sollten mit Voraussicht auf die unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt werden, so dass deren Kooperation erreicht werden kann und so dass mögliche Versuche irgendeiner dieser Gruppen, die Evaluationsoperationen zu beschneiden oder die Resultate zu verzerren oder zu missbrauchen, vermieden werden können.

2.3. *Kostenwirksamkeit*

Die Evaluation sollte Informationen mit einem Wert produzieren, der die gemachten Aufwendungen rechtfertigt.

3. **Korrektheit**

3.1. *Formale Verpflichtung*

Den Verpflichtungen der formalen Parteien einer Evaluation sollte von diesen in schriftlicher Form zugestimmt werden, damit diese Parteien zur Einhaltung aller Bedingungen verpflichtet sind oder damit sie diese formell aufs neue verhandeln müssen.

3.2. *Deklaration von Interessenkonflikten*

Interessenkonflikte - oft unausweichlich - sind offen und aufrichtig zu behandeln, damit sie den Prozess und die Resultate der Evaluation nicht kompromittieren.

3.3. *Vollständige und aufrichtige Offenlegung*

Mündliche und schriftliche Evaluationsberichte sollten in der Darlegung der sachdienlichen Ergebnisse und der Grenzen der Evaluation offen, direkt und ehrlich sein.

3.4. *Herstellung von Öffentlichkeit*

Die formellen Parteien einer Evaluation sollten das Recht der Öffentlichkeit auf Information respektieren und sicherstellen, wobei die Begrenzungen durch andere verwandte

Prinzipien und Regelungen, wie die öffentliche Sicherheit und das Recht auf Privatsphäre, berücksichtigt werden müssen.

3.5. *Schutz individueller Menschenrechte*

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

3.6. *Menschliche Interaktionen*

Evaluatoren sollten in den Interaktionen mit anderen Personen im Zusammenhang mit der Evaluation die Würde und den Wert der Menschen respektieren.

3.7. *Ausgewogene Berichterstattung*

Evaluationen sollten in der Präsentation der Stärken und Schwächen des Evaluandums vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder benannt werden können.

3.8. *Finanzielle Verantwortlichkeit*

Die Zuweisung und Ausgabe von Ressourcen durch den Evaluator sollte unter Berücksichtigung fundierter und verantwortlicher Verfahren und auch anderweitig klug und ethisch verantwortlich erfolgen.

4. Genauigkeit

4.1. *Identifikation des Gegenstandes*

Der zu evaluierende Gegenstand sollte ausreichend untersucht werden, so dass die in der Evaluation berücksichtigte(n) Form(en) des Gegenstandes klar identifiziert werden können.

4.2. *Kontextanalyse*

Der Kontext des Evaluandums sollte ausreichend detailliert untersucht werden, so dass mögliche Beeinflussungen des Evaluandums identifiziert werden können.

4.3. *Beschreibung von Zielen und Vorgehen*

Die Ziele und das Vorgehen der Evaluation sollten ausreichend genau beobachtet und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.

4.4. *Verlässliche Informationsquellen*

Die Informationsquellen sollten ausreichend detailliert beschrieben werden, so dass die Adäquatheit der Informationen eingeschätzt werden kann.

4.5. *Valide Messung (a) bzw. Glaubwürdigkeit (b) und Übertragbarkeit (c)*

(a) Die Instrumente und Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt ist.

(b) Die Rekonstruktionen durch die Evaluation sollten eine korrekte Abbildung der Konstruktionen der Akteure darstellen.

(c) Eine präzise, detaillierte und vollständige Beschreibung (thick description) des untersuchten Kontextes sollte es erlauben, zu entscheiden, ob die Aussagen der Evaluation auf andere Kontexte übertragbar sind.

4.6. *Reliable Messung (a) bzw. Verlässlichkeit (b)*

(a) Die Instrumente und Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Zuverlässigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt ist.

(b) Der Evaluationsprozess sollte von aussen erkennbar sein, damit auch die getroffenen Entscheidungen und Interpretationen in der Evaluation untersucht und beurteilt werden können.

- 4.7. *Systematische Datenüberprüfung*
Die in einer Evaluation gesammelten, bearbeiteten und angegebenen Daten sollten überprüft und korrigiert werden, so dass die Resultate der Evaluation nicht fehlerhaft sind.
- 4.8. *Analyse quantitativer Information*
Quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, um haltbare Interpretationen sicherzustellen.
- 4.9. *Analyse qualitativer Information*
Qualitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, um haltbare Interpretationen sicherzustellen.
- 4.10. *Gerechtfertigte Folgerungen*
Die in einer Evaluation erfolgten Folgerungen sollten explizit begründet sein, so dass die Beteiligten und Betroffenen diese einschätzen können.
- 4.11. *Objektive Berichterstattung (a) bzw. Bestätigbarkeit (b)*
(a) Die Evaluationsverfahren sollten über Vorkehrungen verfügen, um die Ergebnisse und Berichte einer Evaluation gegen Verzerrungen durch persönliche Gefühle und Vorlieben irgendeiner Partei in der Evaluation zu schützen.
(b) Die Konstruktionen der Evaluation sollten auf ihre Ursprünge und Wurzeln im empirischen Material auch von aussen rückführbar sein.

3.5. Zwischenbemerkungen zu den Kriterien

Aufgrund der vorangehenden Beschreibung der Kriterien drängen sich bereits vor deren empirischen Anwendung einige Bemerkungen auf. Die Darstellung der Kriterien macht deutlich, dass der Kriterienraster die Erfordernisse, die üblicherweise an eine sozialwissenschaftliche Typologie gestellt werden, in verschiedener Hinsicht nicht erfüllt:

- *Eindeutigkeit*: Eine eindeutige Zuordnung empirischer Sachverhalte zu einer theoretischen Kategorie ist nicht immer möglich. Es wird nicht zu vermeiden sein, dass ein und dieselbe Beobachtung bezüglich verschiedener Kriterien relevant wird. Daraus wird bei der empirischen Umsetzung der Kriterien zur Bewertung der Evaluationen das Problem entstehen, dass verschiedentlich Wiederholungen auftreten werden.

Beispiel: Wird in einem Evaluationsbericht darauf verzichtet, die Vorgehensweise zu beschreiben, wirkt sich dies sowohl beim Kriterium 'Klarheit des Berichts' (siehe Kapitel 3.4.1.5.) als auch beim Kriterium 'Beschreibung von Zielen und Vorgehen' (siehe Kapitel 3.4.4.3.) aus.

- *Einheitlichkeit*: Aufgrund der Beschreibungen wird deutlich, dass sich die verschiedenen Kriterien auf deutlich verschiedenen analytischen Ebenen bewegen. Das heisst, dass verschiedene Kriterien als Vorbedingungen für andere Kriterien bezeichnet werden können. Daraus kann das Problem entstehen, dass wegen der Nicht-Erfüllung einer Vorbedingung auch nachgelagerte Kriterien nicht erfüllt werden können.

Beispiel: Werden in einer Evaluation Informationen beschafft, welche die sachdienliche Behandlung von Fragen zum Evaluandum nicht erlauben, hat dies Auswirkungen bei der Bewertung des Kriteriums 'Informationsrahmen und -auswahl' (siehe Kapitel 3.4.1.3.). Ohne sachdienliche Informationen wird es kaum möglich sein, den zu evaluierenden Gegenstand ausreichend zu untersuchen, womit auch das Kriterium 'Identifikation des Gegenstandes' (siehe Kapitel 3.4.4.1.) nicht erfüllt werden kann.

- *Konsistenz:* Die Konsistenz der Typologie ist nicht gegeben, da verschiedene Kriterien konfligierende Erfordernisse formulieren. Die Kriterien enthalten verschiedentlich Zielkonflikte. Die Einhaltung eines Kriteriums kann zum Verstoss bei einem anderen Kriterium führen. Deshalb wird sich das Problem ergeben, dass eine Erfüllung aller Kriterien nicht möglich sein wird und dass sich verschiedene - miteinander verknüpfte (siehe unten) - Optimierungsprobleme einstellen.

Beispiel: Im Kriterium 'Ausgewogene Berichterstattung' (siehe Kapitel 3.4.3.7.) wird eine vollständige und faire Präsentation der Stärken und Schwächen eines Evaluandums gefordert. Dies kann die Kooperation einer relevanten Interessengruppe gefährden und damit zu Problemen hinsichtlich des Kriteriums 'Politische Überlebensfähigkeit' (siehe Kapitel 3.4.2.2.) führen.

- *Unabhängigkeit:* Die einzelnen Kriterien sind nicht unabhängig voneinander, sondern hochgradig interdependent. Verschiedenste Kriterien beeinflussen einander in positiver oder negativer Weise. Die Kriterien sind stark miteinander verflochten. Dies erschwert einerseits die Zuordnung von Beobachtungen zu einzelnen Kriterien, andererseits machen die vielfachen Verknüpfungen häufige Querverweise notwendig.

Beispiel: Ein typisches Beispiel für diese Problematik stellen die beiden Kriterien 'Valide Messung' (siehe Kapitel 3.4.4.5.a.) und 'Reliable Messung' (siehe Kapitel 3.4.4.6.a.) dar. Diese beiden Erfordernisse bedingen sich gegenseitig. Eine nicht-reliable Messung kann nicht valide sein. Ist eine Messung nicht valide, dann wird die Qualität der Messung nicht besser, wenn die Reliabilität erhöht wird.

Im Rahmen der empirischen Umsetzung werden sich vielleicht weitere Probleme ergeben. Die bereits jetzt auf theoretischer Ebene festgestellten Schwierigkeiten werden sich konkretisieren und vertiefen lassen. Wir werden diese Thematik nach den empirischen Fallstudien wieder aufnehmen (siehe Kapitel 6. und 7.1.).

3.6. Exkurs: Aktuelle Entwicklungen

3.6.1. 'The Program Evaluation Standards' 1994

3.6.1.1. Allgemeines

Zur Ableitung der Kriterien haben wir uns im vorangehenden Kapitel auf die 'Standards for Evaluations of Educational Programs, Projects, and Materials' (Joint Committee 1981a) abgestützt. Diese Standards sind im Jahre 1981 erschienen. Seither haben sie eine verhältnismässig weite Verbreitung gefunden, wobei jedoch nicht von einer allgemeinen Anerkennung dieser Standards gesprochen werden kann (siehe etwa Love 1994: 32-33; vgl. oben Kapitel 3.3.). Im Sommer 1994 hat nun das 'Joint Committee on Standards for Educational Evaluation' eine revidierte Version der 1981er Standards publiziert (Joint Committee 1994, vgl. dazu auch Stufflebeam 1994: 334-336). Diese Version wird nunmehr von insgesamt fünfzehn Organisationen unterstützt, unter denen sich auch solche finden, die keine erziehungswissenschaftliche Ausrichtung aufweisen. Zu nennen ist an dieser Stelle vor allem die 'American Evaluation Association' (AEA). Ein weiteres Anzeichen für eine Ausdehnung des Wirkungskreises zeigt sich daran, dass die 1994er-Version mit 'The Program Evaluation Standards' bezeichnet ist und sich lediglich im Untertitel ('How to Assess Evaluations of Educational Programs') ein Hinweis auf den Erziehungsbereich findet. Die neue Version der Standards ist sodann auch vom 'American National Standards Institute' (ANSI) als 'American National Standard' anerkannt worden (vgl. dazu Sanders 1994). Die aktuelle Literatur bezieht sich nach wie vor sehr häufig auf die Standards (Caracelli/Riggin 1994; Tomlinson/Bland/Moon/Callahan 1994, insbes.: 156; Stufflebeam 1994; Worthen 1994, insbes.: 11; Mertens 1994: 24; Love 1994: 32-33; Stevens/Dial 1994b: 5-6; LeCompte 1994). Die AEA hat vorderhand darauf verzichtet, eigene Standards aufzustellen bzw. die Standards einer ihrer Vorläuferorganisationen, der 'Evaluation Research Society' (ERS) (ERS Standards Committee 1982) weiterzuentwickeln (vgl. dazu aber Kapitel 3.5.2.). Es zeichnet sich immer deutlicher ab, dass die Standards des 'Joint Committee' *die* Standards der amerikanischen Evaluationsforschung darstellen (Patton 1994: 194).

Weil die neue Version der Standards erst im Juli 1994 erhältlich waren, mussten wir uns für die empirischen Teile der vorliegenden Meta-Evaluation auf die ältere Version abstützen. Dies ist einerseits sicherlich bedauerlich, andererseits zeigt sich, dass die vorgenommenen Änderungen nicht derart zentral und revolutionär sind. Wir werden im folgenden die wesentlichsten Änderungen kurz erläutern.

Die Grundstruktur, also die Aufteilung in die vier Gruppen von 'Utility Standards', 'Feasibility Standards', 'Propriety Standards' und 'Accuracy Standards' ist beibehalten worden. Neu enthalten die Standards selbst Literaturhinweise. In der älteren Version wurde bewusst darauf verzichtet (Wildemuth 1981: ii), wobei eine begleitende Bibliographie erstellt worden war (Wildemuth 1981). Der Terminus 'audiences' ist durch die Begriffe 'stakeholders' respektive 'intended users' ersetzt worden. Diese begriffliche Veränderung zeigt einen Aspekt der Entwicklung der Evaluationsforschung in den 80er Jahren. Die ursprüngliche Vorstellung passiv konsumierender Adressaten wich dem Bild aktiv mitwirkender 'stakeholders'. Zur sprachlichen Vereinfachung wurden in den neuen Standards die Begriffe 'program, project, and material' durch 'program' ersetzt, wobei dies jedoch keine Einschränkung des Gegenstandsbereichs implizieren soll, sondern lediglich der sprachlichen Straffung dient (siehe Joint Committee 1994: 208).

Bezüglich der einzelnen Standards sind diverse Änderungen vorgenommen worden, die wir im Anschluss daran aufzeigen werden. Wir haben uns dabei darauf konzentriert, die Bezeichnungen (die Titel) und Beschreibungen (der Satz) der beiden Versionen der Standards zu vergleichen. Auf Änderungen in den weiteren Erläuterungen ('Overview', 'Guidelines', 'Pitfalls', 'Caveats', 'Common Errors' und 'Illustrative Cases') werden wir im folgenden nicht detaillierter eintreten können.

3.6.1.2. Utility Standards

Neben den bereits genannten Revisionen in der Terminologie (wie 'audience' zu 'stakeholders') ergeben sich folgende Änderungen:

1981:

Information Scope and Selection (A3):

Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

1994:

Information Scope and Selection (U3):

Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

Wie aus dieser Gegenüberstellung ersichtlich wird, liegt vor allem eine sprachliche Klärung bzw. Vereinfachung vor. Eine inhaltliche Veränderung ist kaum festzumachen.

Die beiden Standards, die sich mit der Verbreitung und der Rechtzeitigkeit des Berichts befassen, sind in einem Standard zusammengefasst worden:

1981:

Report Dissemination (A6):

Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

Report Timeliness (A7):

Release of reports should be timely, so that audiences can best use the reported information.

1994:

Report Timeliness and Dissemination (U6):

Significant interim findings and evaluations reports should be disseminated to intended users, so that they can be used in a timely fashion.

Neben der Zusammenfassung fällt an der revidierten Form auf, dass sich dieses Kriterium explizit auch auf wichtige Zwischenergebnisse bezieht. Weiter entfällt die Betonung der Einschätzung durch die Adressaten ebenso wie die genauere Beschreibung der Adressaten. Die Differenz zwischen den beiden Versionen ist aber auch hier nicht bedeutend. Weil diese beiden Standards zusammengefasst wurden, umfasst die Gruppe der 'Utility Standards' neu nur noch sieben (anstatt früher acht) Standards.

Der ehemalige Standard 'Valuational Interpretation' (A4) erhält die neue Bezeichnung 'Values Identification' (U4). Der Text des Standards bleibt jedoch identisch.

Die Beschreibung zum Standard 'Report Clarity' ist sprachlich gestrafft worden, ohne dass inhaltliche Divergenzen zu erkennen wären. Die beiden Versionen lauten:

1981:

Report Clarity (A5):

The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

1994:

Report Clarity (U5):

Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

Der neue Standard zur Wirkung der Evaluation ist einerseits noch stärker auf die Nutzung ausgerichtet und bezieht nun auch die Berichterstattung mit ein:

1981:

Evaluation Impact (A8):

Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

1994:

Evaluation Impact (U7):

Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Auch hier sind die vorgenommenen Änderungen verhältnismässig klein.

3.6.1.3. Feasibility Standards

Die Standards 'Practical Procedures' (B1 resp. F1) und 'Political Viability' (B2 resp. F2) blieben nahezu unverändert. Beim dritten Standard in dieser Gruppe hat sich die Gewichtung leicht verschoben. Die beiden Versionen lauten:

1981:

Cost Effectiveness (B3):

The evaluation should produce information of sufficient value to justify the resources expended.

1994:

Cost Effectiveness (F3):

The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

Während die frühere Version das Hauptaugenmerk auf die Produktion wertvoller Information legt, zielt die neuere Version vor allem auf die Rechtfertigung der eingesetzten Ressourcen. Fraglich ist, ob der Standard in der Version von 1994 klarer formuliert ist als sein Vorläufer (vgl. dazu auch Patton 1994: 196).

3.6.1.4. Propriety Standards

Die Gruppe der 'Propriety Standards' umfasst auch in der revidierten Version acht Standards. In dieser Gruppe ist ein Standard weggelassen und durch einen anderen ersetzt worden. Diese beiden Standards lauten:

1981:

Public's Right to Know (C4):

The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

1994:

Service Orientation (P1):

Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

Michael Quinn Patton sieht vor allem zwei Gründe, weshalb der Standard 'Public's Right to Know' in der 1994er-Version nicht mehr berücksichtigt wurde (Patton 1994: 196-197). Erstens sei verschiedentlich festgestellt worden, dass der Standard für nicht-westliche Kontexte weniger geeignet erscheint. Zweitens habe sich gezeigt, dass er bei formativ angelegten Evaluationen nicht zweckmässig ist, da die Nützlichkeit in diesen Fällen oft reduziert wird, wenn die Evaluationsergebnisse in die breite Öffentlichkeit getragen werden.

Der neue Standard 'Service Orientation' ist in seinem Wortlaut etwas unklar. Durch die weiteren Erläuterungen wird jedoch ausreichend geklärt, was die Absicht ist, die hinter diesem Standard steht. Es heisst dort: "In short, evaluations should serve program participants, community, and society." Evaluationen sollen also keinen Selbstzweck darstellen, sondern zur Verbesserung von Programmen beitragen, damit die Gesellschaft davon profitiert (Verantwortung gegenüber der Öffentlichkeit). Damit wird auch klar, dass die Ausrichtung dieses neuen Standards gar nicht so weit entfernt von jener des gestrichenen Standards 'Public's Right to Know' liegt, auch wenn die Dienstleistungsorientierung als neuer Aspekt hinzukommt.

Der Standard 'Formal Obligation' (C1) wurde neu mit 'Formal Agreements' (P2) bezeichnet. An der Formulierung sind jedoch keinerlei Veränderungen vorgenommen worden.

Die neue Version des Standards 'Conflict of Interest' (C2 resp. P7) ist lediglich von der Zwischenbemerkung entlastet, dass Interessenkonflikte häufig unvermeidlich seien. Ansonsten sind Bezeichnung wie Wortlaut identisch.

Umfangreicher sind hingegen die Änderungen beim früheren Standard 'Full and Frank Disclosure' (C3):

1981:

Full and Frank Disclosure (C3):

Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.

1994:

Disclosure of Findings (P6):

The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

Diese beiden Standards unterscheiden sich in verschiedener Hinsicht. Während die ältere Version die Art der Vermittlung besonders betont ("open, direct, and honest"), legt die neuere Version das Schwergewicht eher auf die Umschreibung des Kreises jener Personen, die Zugang zu den Resultaten erhalten sollten. Diese Veränderung zeigt sich auch in der Bezeichnung der beiden Standards. Weiterhin weist die 1981er-Version speziell auf die Form der Vermittlung hin, während die 1994er-Ausgabe speziell betont, dass die Einhaltung dieses Standards Sache der formellen Parteien einer Evaluation ist, und damit ausdrücklich nicht nur die Evaluatoren, sondern auch die Auftraggeber in die Pflicht nimmt. Bei beiden Versionen wird explizit darauf hingewiesen, dass sich dieser Standard nicht nur auf die Resultate, sondern auch auf die Grenzen der Evaluation bezieht.

Beim Standard 'Rights of Human Subject' (C5 resp. P3) beschränkt sich die Revision auf eine sprachliche Veränderung ohne inhaltliche Implikationen.

Die Änderung beim Standard 'Human Interactions' präzisiert mit einem Nachsatz die Absicht dieser Norm:

1981:

Human Interactions (C6):

Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation.

1994:

Human Interactions (P4):

Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

Ein weiterer Standard, der im Verlauf des Revisionsprozesses verschiedenen Veränderungen unterworfen war, lautet wie folgt:

1981:

Balanced Reporting (C7):

The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas adressed.

1994:

Complete and Fair Assessment (P5):

The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas adressed.

Es zeigen sich im wesentlichen zwei Änderungen: Die erste betrifft die Bezeichnung des Standards, bei der auf 'ausgewogen' zugunsten von 'vollständig und fair' verzichtet wird. Diese Änderung ist kritisiert worden, da es unrealistisch erscheine, vollständig zu sein, und dass es wichtiger sei, eine ausgewogene und faire Berichterstattung zu erreichen (Patton 1994: 197). Die zweite Änderung betrifft den Austausch des Ausdrucks "presentation" durch "examination and recording", was unseres Erachtens deutlich machen soll, dass sich dieser Standard nicht nur auf die Berichterstattung bezieht, sondern die ganze Evaluation betrifft. Dies kommt auch in der Bezeichnung des Standards zum Ausdruck, wird doch nicht mehr von "reporting", sondern von "assessment" gesprochen.

Bei der 'Fiscal Responsibility' besteht die Differenz wiederum nur aus einem angefügten Nachsatz, der die Absicht des Standards verdeutlichen soll:

1981:

Fiscal Responsibility (C8):

The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.

1994:

Fiscal Responsibility (P8):

The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

3.6.1.5. Accuracy Standards

Der erste Standard in dieser Gruppe wurde neu formuliert und trägt auch eine neue Bezeichnung:

1981:

Object Identification (D1):

The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.

1994:

Program Documentation (A1):

The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

Unseres Erachtens liegt hier vor allem eine deutliche sprachliche Verbesserung vor. Die Bezeichnung des Standards, wie die Umschreibung selbst auch, legt neu grösseres Gewicht auf die Dokumentation, als auf die Identifikation des Gegenstandes.

Die Standards 'Context Analysis' (D2 resp. A2), 'Defensible Information Sources' (D4 resp. A4), 'Justified Conclusions' (D10 resp. A10) bleiben bis auf kleinere sprachliche Revisionen identisch und bei 'Described Purposes and Procedures' (D3 resp. A3) wurden keinerlei Änderungen vorgenommen.

Die Standards zur Validität und zur Reliabilität sind vor allem neu bezeichnet worden. Sie präsentieren sich folgendermassen:

1981:

Valid Measurement (D5):

The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

1994:

Valid Information (A5):

The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

Reliable Measurement (D6):

The informationgathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.

Reliable Information (A6):

The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained ist sufficiently reliable for the intended use.

Neben der geänderten Bezeichnung sind vor allem sprachliche Anpassungen vorgenommen worden. Auffällig ist, dass sich das Validitätserfordernis nach wie vor auf die Inter-

pretation bezieht, während die Reliabilität hinsichtlich der Information beurteilt werden soll. In der Bezeichnung wird diese Differenz jedoch nicht kenntlich gemacht.

Der Standard 'Systematic Data Control' (D7) heisst neu 'Systematic Information' (A7) und ist sprachlich leicht verändert worden, wobei keine wesentliche inhaltliche Verschiebung festzustellen ist.

Die Standards 'Analysis of Quantitative Information' (D8 resp. A8) und 'Analysis of Qualitative Information' (D9 resp. A9) wurden praktisch unverändert übernommen, wobei sich die Zielsetzung verändert darstellt. Anstatt "supportable interpretations" (1981) sicherzustellen, soll die Analyse zu einer wirksamen Beantwortung der evaluativen Fragestellung führen. Es wird damit neu ein stärkerer Bezug zur evaluativen Funktion der Analyse hergestellt, als dies in der früheren Version der Fall war. Eher qualitativ orientierte Evaluatoren stossen sich daran, dass nach wie vor zuerst die quantitative Analyse behandelt wird und fordern, dass die Standards in alphabetischer Reihenfolge stehen sollten (Patton 1994: 198).

Die Gruppe der 'Accuracy Standards' schliesst in der 1994er-Version mit einem vollständig neuen Eintrag:

1994:

Metaevaluation (A12):

The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Wir werden in unserer Untersuchung diesem Standard folgen. Da wir eine reine ex-post-Analyse vornehmen, werden wir ihn jedoch nicht vollständig einhalten können. Alle durch uns evaluierten Evaluationen waren bei Projektbeginn bereits abgeschlossen (vgl. unten, Kapitel 4.2.).

3.6.2. 'Evaluation Principles' der AEA

Im Juni 1994 hat die 'American Evaluation Association' (AEA) die 'Guiding Principles for Evaluators' herausgegeben (AEA 1994, vgl. dazu Shadish/Newman/Scheirer/Wye 1995). Diese umfassen die folgenden fünf Leitsätze:

1. *Systematic Inquiry*: Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
2. *Competence*: Evaluators provide competent performance to stakeholders.
3. *Integrity/Honesty*: Evaluators ensure the honesty and integrity of the entire evaluation process.
4. *Respect for People*: Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
5. *Responsibilities for General and Public Welfare*: Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Diese fünf Prinzipien werden durch eine Reihe weiterer Aussagen illustriert, wobei diese Erläuterungen weder alle möglichen Anwendungen der 'principles' darstellen sollen, noch als Regeln zu verstehen sind, die als Basis zur Sanktionierung dienen könnten. Die 'principles' sind von einer vier Personen umfassenden 'Task Force' der AEA entwickelt worden. Sie sollen alle zuvor unternommenen Versuche der AEA, bzw. ihrer Vorläuferorganisationen 'Evaluation Research Society' (ERS) und 'Evaluation Network' (ENet), zur Formulierung von Normen, ersetzen. Dies betrifft somit insbesondere die oben diskutierten ERS-Standards aus dem Jahre 1982 (vgl. oben Kapitel 3.2.). Die 'Program Evaluation Standards' des 'Joint Committee' hingegen werden davon explizit ausgenommen. Die beiden Normierungen sollen, gemäss den Angaben im Vorwort der 'Guiding Principles' parallel nebeneinander bestehen. Betrachtet man die inhaltliche Ausrichtung der 'Guiding Principles' und vergleicht diese mit den oben diskutierten Standards, fallen primär zwei Aspekte auf (siehe dazu Sanders 1995). Erstens ergeben sich zwischen den beiden Texten inhaltliche Übereinstimmungen. Alle in den 'Guiding Principles' enthaltenen Ansprüche finden sich in der einen oder anderen Form in den 'Program Evaluation Standards' wieder. Zweitens besteht zwischen den beiden aber ein deutlicher Unterschied. Die 'Guiding Principles' sind viel genereller angelegt, viel weniger spezifiziert und auch weniger elaboriert, als die Standards. Die 'Principles' sind sicherlich weniger detailliert ausgestaltet, sie haben aber gegenüber den Standards auch einen gewichtigen Vorteil. Durch ihre Knapp-

heit und durch ihren übergreifenden Charakter stellen sie eine willkommene Orientierungshilfe dar. Bei den Standards besteht doch, rein schon durch ihren Umfang, die Gefahr, dass man sich darin verliert.

Stellt sich zum Schluss dieses Exkurses die Frage, inwiefern die 'Guiding Principles' (wären sie nicht erst im Juni 1994 erschienen) geeignet gewesen wären, als Vorgaben für die von uns in der Meta-Evaluation eingesetzten Kriterien zu dienen. Unseres Erachtens hätte dies zwar einerseits sicherlich zu einer Straffung der Analyse beigetragen. Andererseits verfügen die 'Principles' nicht über einen ausreichend hohen Detaillierungsgrad (vgl. die in Kapitel 3.3. angeführten Auswahlkriterien) und lassen im Vergleich zu den Joint Committee-Standards (noch) mehr Interpretationsspielraum zu. Dies hätte die Meta-Evaluation erschwert, da der Umgang mit dem Ermessensspielraum bereits bei der gewählten Vorgehensweise eines der Hauptprobleme darstellt. Zudem sind die 'Guiding Principles' mehr als Verhaltensrichtlinien für Evaluatoren zu verstehen und stellen weniger spezifische Kriterien für die Praxis dar, wie dies die Standards des Joint Committee tun (für einen Vergleich siehe Covert 1995, Newman 1995: 108 und Sanders 1995). Deshalb sind wir der Ansicht, dass die Berücksichtigung der 'Guiding Principles' wohl kaum zu einer Verbesserung der Meta-Evaluation hätte beitragen können. Sie sollten unseres Erachtens aber unbedingt als Verhaltensgrundsätze für die Evaluationspraxis zur Kenntnis genommen werden.

4. EMPIRISCHES VORGEHEN

4.1. Überblick

Der im vorangehenden Kapitel hergeleitete Kriterienraster wird im empirischen Teil auf zehn Evaluationsstudien angewendet, die in den letzten fünf Jahren in der Schweiz erarbeitet wurden. Die zehn Fallstudien bestehen jeweils aus einer Beschreibung des Evaluationsprozesses sowie aus einer Bewertung der Evaluation hinsichtlich der einzelnen Kriterien. Die Kriterien sind als Idealkriterien zu verstehen, das bedeutet, dass die Kriterien eine anzustrebende Zielvorstellung darstellen. Somit steht einer vollständigen Erfüllung aller Kriterien nicht nur die fehlende Konsistenz des Kriterienrasters entgegen (Kapitel 3.5.). Es kommt hinzu, dass die Kriterien einen Sollzustand verkörpern, der in Anbetracht der gegebenen Rahmenbedingungen für Evaluationen in der Praxis kaum vollständig erreichbar ist. Dieses Vorgehen wurde gewählt, da eine kritische Betrachtung der Evaluationen beabsichtigt wurde und um sicherzustellen, dass die Meta-Evaluation auch die Evaluation als Ganzes (mit ihrem Kontext) erfassen kann. Die Kriterien erhöhen zwar die Systematik der Bewertung, sie sind jedoch nicht in der Lage, den Meta-Evaluator vom Bewertungszwang zu entlasten. Die Bewertung bedarf einer subjektiven Einschätzung, um zu einer Evaluation zu gelangen. Aus diesem Grund haben wir uns bei der Darstellung der Fallstudien darum bemüht, die Argumentation ausführlich darzulegen, zu belegen und zu erläutern (siehe Kapitel 5 und Widmer 1996a-g). Weiterhin wurden die Evaluatoren zu einer Stellungnahme zu den Fallstudienentwürfen aufgefordert, um Fehler, Lücken und einseitige Gewichtungen möglichst auszuschliessen. Schriftliche Stellungnahmen sind im Anschluss an die Fallstudien im Bericht enthalten. Aufgrund der zehn Fallstudien wurde ein Vergleich der Evaluationen vorgenommen (siehe Kapitel 6). Anhand dieses Vergleichs, ergänzt mit Erkenntnissen aus den Einzelanalysen, wurden Folgerungen abgeleitet, die sich auf zwei unterschiedliche Aspekte beziehen: Einerseits befassen sich die Folgerungen mit der methodischen Frage, inwiefern sich der eingesetzte Kriterienraster bewährt hat und versucht Vorschläge für Modifikationen zu formulieren (vgl. Kapitel 7.2). Andererseits besteht das praxisorientierte Ziel darin, Schlüsse für die zukünftige Ausgestaltung von Evaluationen zu ziehen (vgl. Kapitel 7.2.).

Die empirische Meta-Evaluation ist *nicht* darauf ausgerichtet, die Qualität der *Evaluatoren* zu bewerten. Die Analyse richtet sich auf die Evaluation von *Evaluationen*. Diese wiederum entstehen durch (Inter-)Aktionen verschiedener Individuen, Gruppen und Organisationen im Kontext diverser Rahmenbedingungen. In diesem Umfeld leisten selbstverständlich auch die Evaluatoren einen Beitrag. Sie sind aber bei weitem nicht

immer für bestehende Schwächen (und Stärken) einer Evaluation alleine verantwortlich, sondern höchstens mitverantwortlich.

4.2. Auswahl der zehn Fallstudien

Die Grundgesamtheit, aus der die zehn Fallstudien ausgewählt werden, wird durch folgende Abgrenzungen definiert. Es werden lediglich Evaluationen berücksichtigt,

- die in der Schweiz entstanden sind,
- die von externen Evaluatoren durchgeführt wurden (zu internen Evaluationen vgl. Love 1991 und 1993, Mathison 1991, Torres 1991, Winberg 1991, Duffy 1994, Sonnichsen 1994, Lovell 1995),
- deren Resultate in irgendeiner Form publiziert wurden,
- die den Anspruch erheben, im weitesten Sinn einen sozialwissenschaftlichen Ansatz zu verwenden,
- die bereits abgeschlossen worden sind (aber wenn möglich nicht vor allzu langer Zeit),
- die nicht als akademische Arbeiten entstanden sind (also keine Diplom- oder Lizentiatsarbeiten bzw. Dissertationen) und
- die nicht in die Bereiche der Erziehungswissenschaften oder der Sozial- und Präventivmedizin fallen.

Um in diesem Rahmen eine möglichst breite Auswahl an Evaluationsstudien zur Verfügung zu haben, konsultierten wir folgende Quellen:

1. Verzeichnisse von Evaluationsstudien oder von sozialwissenschaftlichen Studien im allgemeinen (Zimmermann/Kolly/Mahon 1989, Forschungsstelle Wissenschaft und Politik 1993, SIDOS 1993, SGS div. Jahre),
2. nicht-publizierte Projektlisten, die in der Bundesverwaltung oder in deren Auftrag erstellt wurden,
3. eine eigene Dokumentation zur Berichterstattung über Evaluationsstudien in der Presse, die wir seit einigen Jahren geführt haben,
4. Bibliotheken, deren Bestände Evaluationsstudien enthalten (Institut für Politikwissenschaft und weitere Institute der Universität Zürich; ETH Zürich; Zentralbibliothek Zürich; Réseau des bibliothèques romandes et tessinoise RERO) und

5. Jahresberichte des Schweizerischen Nationalfonds, insbesondere bezüglich der hier relevanten Nationalen Forschungsprogramme (SNF, div. Jahre).

Diese Quellen garantieren zwar in keiner Weise eine vollständige Erfassung aller in Frage kommenden Evaluationsstudien. Wir sind jedoch der Ansicht, dass die Liste mit den über 110 Evaluationsstudien die wesentlichen Bereiche abdecken.

Zur Absicherung, dass keine besonders relevanten Studien unberücksichtigt bleiben, ist bei vier Personen, die in der schweizerischen Evaluationsforschung einen Expertenstatus einnehmen, eine kurze Umfrage durchgeführt worden. Den Experten wurde dabei die Frage gestellt, welche Evaluationsstudien sie im Rahmen einer Meta-Evaluation in Betracht ziehen würden. Es wurden dazu folgende Personen kurz befragt:

- Werner Bussmann, Programmleiter NFP 27, Bern
- Luzius Mader, Bundesamt für Justiz, ehem. Sekretariat AGEVAL, Bern
- Markus Spinatsch, Leiter der Parlamentarischen Verwaltungskontrollstelle, Bern und
- Willi Zimmermann, Forschungsstelle Wissenschaft und Politik, Bern.

Aufgrund dieser Befragung zeigte sich, dass das von uns erstellte Inventar fast alle relevanten Evaluationen im oben definierten Evaluationsfeld umfasst, musste es - aufgrund der Umfrage - nur gerade um eine Studie ergänzt werden.

Anhand des Zeitpunktes des Projektabschlusses (bzw. des Publikationstermins des Schlussberichts) wurde eine Vorauswahl getroffen. Es wurden lediglich Studien berücksichtigt, die in den Jahren ab 1990 fertiggestellt wurden. Diese Abgrenzung ist wünschbar, da sich unsere Untersuchung auf die heutige Evaluationspraxis beziehen soll. Sie ist aber auch zweckmässig, da bei älteren Studien die Informationsbeschaffung schwierig gewesen wäre. Dadurch verkleinerte sich die Zahl der zur Auswahl stehenden Evaluationsstudien auf 53.

Zu den verbleibenden Studien wurden anhand eines Erfassungsrasters zur Verfügung stehende Angaben zu den nachfolgenden Dimensionen erfasst:

Dimensionen zum Evaluandum:

- *Politische Ebene* (Bund, Kanton, Gemeinde)
- *Struktur der Policy-Arena*: Mit der Struktur der Policy-Arena sind primär die Konflikt- und Konsensusprozesse gemeint. Ein Politikbereich, in dem ein Evaluandum zu lokalisieren ist, kann sich diesbezüglich stark unterscheiden. Neben unterschiedlichen Konfliktivitätsgraden, also der Umstrittenheit einer Politik (vgl. Widmer 1987), kann sich

die Policy-Arena (Lowi 1964 und 1972) auch dadurch unterscheiden, ob sie über längerfristig konstante Koalitionen verfügt, oder ob diese nur sehr kurzlebig sind (vgl. dazu Sabatier/Jenkins-Smith 1993). Der dritte Aspekt betrifft die Polarität einer Policy-Arena, also die Frage durch wieviele Koalitionen (bzw. Konfliktlinien) sie geprägt ist.

- *Art des Evaluandums* (Einzelmaßnahme, Gesamtprogramm, Organisation) (vgl. dazu Kaufmann/Rosewitz 1983: 34-36)
- *Steuerungsprinzip(ien) des Evaluandums*: (Gebot/Verbot (regulativ), positive/negative Anreize ('incentives'), Angebot, Überzeugung ('persuasion') und Vorbild) (Windhoff-Héritier 1987: 27-34)

Dimensionen zur Evaluation:

- *Umfang der Evaluation* (finanzielle, personelle und zeitliche Ressourcen)
- *Erkenntnistheoretische Ausrichtung*: Darunter verstehen wir den erkenntnistheoretischen Ansatz der in der Evaluation verfolgt wird. Es lassen sich dabei zwei verschiedene Grundausrichtungen ausmachen. Einmal handelt es sich dabei um das von uns als 'traditionelles Paradigma' bezeichnete Verständnis. Darunter fallen alle realistischen, (post-)positivistischen, an Objektivität und Wertfreiheit orientierten Ansätze. Die zweite Grundausrichtung vereint konstruktivistische, partizipativ-orientierte, responsive und perspektivistisch angelegte Ansätze, welche die Objektivität und Wertfreiheit als wissenschaftliche Zielorientierung verneinen (vgl. dazu Guba/Lincoln 1994).
- *Vergleichsdimension* (Längsschnitt- oder Querschnittuntersuchungen bzw. Kombinationen)
- *Methodik*: Darunter verstehen wir alle Ansätze, Verfahren, Techniken und Instrumente, die dazu dienen, die empirische Umsetzung einer Evaluation mit wissenschaftlichen Mitteln zu gewährleisten. Dazu zählen Evaluationsansätze und Untersuchungsanlagen ebenso wie Verfahren zur Datenerhebung und Datenauswertung.

Dimensionen zu den Evaluatoren:

- *Kontext der Evaluatoren*: Hier sind primär zwei Formen zu unterscheiden, nämlich Evaluatoren, die als Selbständige oder als Angestellte eines privatwirtschaftlichen Unternehmens Evaluationsaufträge bearbeiten sowie jene, die an einer Universität oder Hochschule tätig sind. Neben diesen beiden Formen sind auch Mischkategorien möglich, also gemischt zusammengesetzte Evaluatorenteams oder Einzelpersonen, die sowohl privatwirtschaftlich wie universitär aktiv sind.

- Disziplinäre Herkunft der Evaluatoren

Dimensionen zu den Auftraggebern:

- *Politische Ebene* (Bund, Kanton, Gemeinde)
- *Art des Auftraggebers* (Exekutive, Legislative bzw. Forschungsförderung)

Eine weitere, anfänglich ins Auge gefasste Dimension, jene ob der Auftraggeber öffentlicher oder privater Natur ist, musste fallengelassen werden, da die Zahl der vorliegenden Evaluationen im Auftrag Privater zu klein war.

Die Expertengruppe des NFP 27 gab zudem vor, dass in der Meta-Evaluation auch einige Evaluationen, die im Rahmen dieses Forschungsprogramms entstanden sind, berücksichtigt werden sollen, und zwar sowohl Vollprojekte, wie auch die im Rahmen des genannten Programms mitfinanzierten sogenannten Kurzevaluationen. Wir haben uns dazu entschlossen, vier Forschungsprojekte sowie zwei sogenannte Kurzevaluationen aus diesem Programm zu berücksichtigen.

Alle Informationen, die uns zu den 46 Evaluationsstudien vorlagen, wurden anhand dieses Schemas systematisiert. Dimensionen, zu denen wir über keine Informationen verfügten, wurden offengelassen.

Auf dieser Basis wurde die Auswahl der zehn zu evaluierenden Evaluationsstudien vorgenommen. Es wurde versucht, im Rahmen der abgesteckten Grenzen eine möglichst weit gestreute Auswahl zu treffen. Dabei wurde in den genannten Dimensionen eine möglichst breite Streuung der Stichprobe angestrebt.

Die Gruppe der berücksichtigten Evaluationen setzt sich aus folgenden Studien zusammen (Autoren und Erscheinungsjahr in Klammern):

1. 'Umweltpolitik und technische Entwicklung' (Andreas Balthasar und Carlo Knöpfel, 1994)
2. 'Unternehmerische Innovationsprozesse' (Willy Bierter und Hans-Martin Binder, 1993)
3. 'Bedingt wirksam: Das Lohngleichheitspostulat des Heimarbeitsgesetzes' (Frohmut W. Gerheuser und André Schmid, 1993)
4. 'Wohneigentumsförderung durch den Bund' (Hans-Rudolf Schulz, Christoph Muggli und Jörg Hübschle, 1993)

5. 'Evaluation der Zürcher Gemeinschaftszentren (ZGZ)' (Andreas Näf et al., 1993)
6. 'Schneeräumung auf Uner Passstrassen' (Heini Sommer und Stefan Suter, 1993)
7. 'Evaluation Kehrichtsackgebühr' (Barbara Haering et al., 1990)
8. 'Verbrauchsabhängige Heizkostenabrechnung - Förderung erneuerbarer Energien' (Wolf Linder et al. 1991)
9. 'Evaluation des BUWAL' (Peter Knoepfel et al., 1991)
10. 'Evaluierung der bundeseigenen EWR-Informationskampagne' (Claude Longchamp, 1993)

Die ersten vier Evaluationen sind Forschungsprojekte, die im Rahmen des NFP 27 entstanden sind. Die Studien 5 und 6 sind im Auftrag eines Kantons bzw. einer Stadt erarbeitet worden, wobei sich das NFP 27 an der Finanzierung dieser Studien unter dem Titel 'Kurzevaluationen' mitbeteiligt hat. Die Evaluationen 7 und 8 wurden im Auftrag der Arbeitsgruppe Gesetzesevaluation (AGEVAL) des EJPD realisiert. Die zwei letzten Studien sind durch die Geschäftsprüfungskommission des Nationalrats bzw. durch das Integrationsbüro EDA/EVD veranlasst worden.

Bezüglich einiger der oben angeführten Dimensionen, die zur Auswahl der Evaluation beigezogen wurden, lassen sich diese Evaluationen grob wie in Tabelle 1 angeben einordnen.

Tabelle 1a: Übersicht über die zehn Fallstudien (Teil 1)

Projekt	Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5
<i>Titel</i>	Umweltpolitik und technische Entwicklung	Unternehmerische Innovationsprozesse	Lohnleichheit bei Heimarbeit	Wohneigentumsförderung des Bundes	Zürcher Gemeinschaftszentren
<i>Autoren</i>	Andreas Balthasar, Carlo Knöpfel	Willy Bierter, Hans-Martin Binder	Frohmut W. Gerheuser, André Schmid	Hans-Rudolf Schulz, Jörg Hübschle, Christoph Muggli	Martin Kaufmann, Andreas Näf, Markus Hämmerle
<i>Auftraggeber</i>	NFP 27	NFP 27	NFP 27	NFP 27	Sozialamt Stadt Zürich / NFP 27
<i>Ebene Evaluandum</i>	Bund, Kantone und Gemeinden	Bund und Kantone	Bund	Bund	Gemeinde
<i>Art Eval.</i>	Einzelmassn.	Programm	Einzelmassn.	Teilprogramm	Organisation
<i>Kontext Evaluator</i>	privat	privat	privat	privat	privat
<i>Disziplinäre Herkunft Evaluator</i>	Politikwiss. Ökonomie	Physik/Sozialw. Politologie Ökonomie	Soziologie Ökonomie	Ökonomie	Ökonomie Betriebsingenieur Recht
<i>Dauer Evaluation</i>	24 Monate	39 Monate	30 Monate	26 Monate	4 Monate
<i>Kosten (rd.)</i>	254'000.-- sFr.	219'000.-- sFr.	136'000.-- sFr.	230'000.-- sFr.	88'000.-- sFr.
<i>Erkenntnistheorie</i>	traditionell	alternativ	alternativ	traditionell	traditionell
<i>Ansatz</i>	summat./format.	format./summat.	summat./format.	summat./format.	format./summat.
<i>Dimension</i>	quer/längs	quer/längs	quer/längs	quer/längs	quer
<i>Methodik</i>	Fallstudie, Marktanalyse, Statistik, Dokumenten, Interviews, Sekundärstatistik, Umfrage	Fallstudie, Dokumentenanalyse, Leitfaden-Interv., Gruppendisk., Inhaltsanalyse	Fallstudien, Dokumenten, offene Interviews, Sekundärstatistik	Dokumentenanal. Sekundärstatistik, offene Interviews, stand. Befragung	Dokumentenanal. Befragung, Gruppendiskussion, Kosten-Nutzen-Analyse

Tabelle 1b: Übersicht über die zehn Fallstudien (Teil 2)

Projekt	Nr. 6	Nr. 7	Nr. 8	Nr. 9	Nr. 10
<i>Titel</i>	Schneeräumung auf Urner Passstrassen	Evaluation Kehrichtsackgebühr	Verbrauchsabh. Heizkostenabr. und erneuerbare Energien	Evaluation des BUWAL	Evaluierung EWR-Informationskampagne des Bundes
<i>Autoren</i>	Heini Sommer, Stefan Suter	Barbara Haering, Samuel Mauch, Adrian Aebersold, Corine Mauch	Wolf Linder, Sandro Looser, Mila Trombitas, Daniel Wittwer	Peter Knoepfel, Willi Zimmermann, Giorgio, Sailer, Enzo Matafora	Claude Longchamp
<i>Auftraggeber</i>	Baudirektion Kt. Uri / NFP 27	AGEVAL (EJPD)	AGEVAL (EJPD)	GPK Nationalrat (Sektion EDI)	Integrationsbüro EDA / EVD
<i>Ebene Evaluandum</i>	Kanton	Gemeinden	Kantone	Bund	Bund
<i>Art Eval.</i>	Einzelmassn.	Programm	Programme	Organisation	Programm
<i>Kontext Evaluator</i>	privat	privat	universitär	universitär/privat	privat
<i>Disziplinäre Herkunft Evaluator</i>	Ökonomie	Nat'wiss., Raumplanung, Bauing., Ingenieure	Politikwiss., Psychologie, Gesch., Ökonomie	Verwaltungsw., Politikwiss., Ökonomie	Politikwissenschaft
<i>Dauer Evaluation</i>	6 Monate	14 Monate	12 Monate	14 Monate	9 Monate
<i>Kosten (rd.)</i>	40'000.-- sFr.	109'000.-- sFr.	150'000.--sFr.	165'000.--sFr.	40'000.-- sFr.
<i>Erkenntnistheorie</i>	traditionell	traditionell	traditionell	-	traditionell
<i>Ansatz</i>	summat./format.	summat./format.	summat./format.	format./summat.	summat./format.
<i>Dimension</i>	längs/quer	längs/quer	längs/quer	quer/längs	quer/längs
<i>Methodik</i>	Sekundärstatistik, Befragung, Interviews	Sekundärstatistik, Interviews, Dokumentenanalyse, Forumsprozess	Dokumentenanalyse, Interviews, Sekundärstatistik, Fallstudien	stand. Befragung, Interviews, Dokumentenanalyse, Fallstudien, Rechtsgutachten	standardisierte Befragung

Anhand dieser Tabelle wird ersichtlich, dass es mit der erfolgten Auswahl gelungen ist, in jeder Dimension jede Kategorie zu belegen. Dies gilt jedoch lediglich bei getrennter Betrachtung der einzelnen Dimensionen. Aufgrund der Anzahl verschiedener Dimensionen und der beschränkten Fallzahl ist es selbstverständlich nicht einzulösen, jede mögliche Kombination der Ausprägungen in den einzelnen Dimensionen abzudecken. Aus diesem Grund können die Resultate unserer Untersuchung auch lediglich Hinweise sowie Thesen aber keine empirisch überprüften (erhärteten) Aussagen liefern ('plausible inference', nach Cronbach et al. 1980). Da nicht für jede kombinatorische Möglichkeit ein Fall zur Verfügung steht, wird auch der Vergleich zwischen den Evaluationen erschwert.

4.3. Zugrundeliegendes Datenmaterial

Die Evaluation der einzelnen Evaluationsstudien baut primär auf einer Dokumentenanalyse auf. Zu jeder der zehn Evaluationsstudien wurden alle erhältlichen schriftlichen Unterlagen beschafft (vgl. dazu Schwandt/Halpern 1988). Dazu gehören beispielsweise folgende Dokumente:

- Evaluationskontrakte, Verträge, Auftragsvereinbarungen, Geschäftsbedingungen
- Ausschreibungsunterlagen
- Offerten, Projektskizzen, Forschungsgesuche
- Erhebungsdokumente, wie Fragebogen, Leitfaden etc.
- Zwischenberichte, Arbeitspapiere, Sitzungsunterlagen, Sitzungsprotokolle
- Schlussberichte, Schlussberichtsentwürfe, Rohberichte, Konzepte, Dispositionen
- Anhänge, Tabellenbände
- Kurzberichte, Pressemitteilungen, Presserohstoffe, Zusammenfassungen, Medienberichterstattung
- Finanzielle Zwischen- und Schlussberichte, Rechnungen, Kostenaufstellungen
- Projekt-Journale, Log-Bücher
- Briefwechsel
- Aktennotizen

Zu jeder Fallstudie findet sich im empirischen Teil (Kapitel 5) eine detaillierte Liste mit den verwendeten Unterlagen. In Ergänzung zu diesen projektbezogenen Materialien haben wir die einschlägige Sekundärliteratur zur behandelten Thematik konsultiert. Weiter haben wir Publikationen der an den Evaluationen beteiligten Personen erhoben. Diese Literatur ist ebenfalls in den Anhängen zu den einzelnen Fallstudien vermerkt.

Der Zugang zu den relevanten Unterlagen gestaltete sich bei den verschiedenen Evaluationen unterschiedlich:

- *Projekte im Rahmen des NFP 27 (Evaluationen 1-4)*: Für Unterlagen zu allen Projekten, die im Rahmen des NFP 27 entstanden sind oder an denen sich das NFP 27 finanziell mitbeteiligte, stellte uns der Programmleiter, Werner Bussmann, seine ausführliche Dokumentation zur Verfügung. Viele Unterlagen sind uns auch von Auftraggebern oder Evaluatoren überlassen worden. Weiter sind verschiedene Dokumente durch zuständige Stellen zugänglich gemacht worden. Der Zugang zu den Dokumenten gestaltete sich hier in der Regel problemlos.
- *Kurzevaluationen mit Kostenbeteiligung des NFP 27 (Evaluationen 5 und 6)*: Für die beiden vom NFP 27 mitfinanzierten Kurzevaluationen konnten wir uns ebenfalls auf die Dokumentation der Programmleitung NFP 27 abstützen. Zudem konnten wir uns hier auf eine vertragliche Verpflichtung mit den Auftraggebern berufen, in der festgehalten wurde, dass die Programmleitung des NFP 27 oder eine von ihr beauftragte Person für die Auswertung ein Einsichtsrecht in projektrelevante Akten erhält. Während dies bei einem der Projekte absolut unproblematisch ablief, erhielten wir bei der zweiten Evaluation auch nach mehrfachem Insistieren keinen Einblick in bestimmte Unterlagen. Der zuständige Vertreter des Auftraggebers liess uns zudem mitteilen, dass er für ein Gespräch nicht zur Verfügung stehen könne (obwohl dies ebenfalls einen Vertragsbestandteil darstellte).
- *AGEVAL-Studien (Evaluationen 7 und 8)*: Bei den beiden im Auftrag der AGEVAL erarbeiteten Evaluationen konnten wir uns auf die Dokumentation der Arbeitsgruppe abstützen. Den Zugang zu diesen Unterlagen hat uns freundlicherweise W. Bussmann ermöglicht, der als Mitglied des Sekretariats der AGEVAL wirkte. Da W. Bussmann beim Projekt 'Evaluation Kehrrechtsackgebühr' zudem Mitglied des begleitenden Ausschusses war, war die Dokumentation zu diesem Projekt dichter, als bei der Studie zur Energiepolitik. Versuche, auch die Dokumentation des zweiten AGEVAL-Projektes zu verdichten, scheiterten, da diese Unterlagen nicht mehr auffindbar waren. Auch bei diesen beiden Projekten haben verschiedene Stellen ergänzende Materialien zur Verfügung gestellt.
- *Evaluation des BUWAL (Evaluation 9)*: Bei dieser Evaluation drängte es sich auf, beim Auftraggeber, der Geschäftsprüfungskommission (GPK) des Nationalrates ein Gesuch um Akteneinsicht und Auskunftserteilung zu stellen. Diesem wurde - unter Vorbehalt der Wahrung der Anonymität - zugestimmt, so dass wir Zugang zur Dokumentation des Sekretariats der GPK erhielten. Aufgrund des bewilligten Gesuchs war es uns möglich mit verschiedenen Personen Gespräche zu führen.
- *Evaluierung EWR-Informationskampagne (Evaluation 10)*: Diese Evaluation ist im Auftrag des Integrationsbüros EDA/EVD erstellt worden. Der Projektnehmer, Claude

Longchamp, hat uns verschiedene Unterlagen mit Einverständnis des Auftraggebers zur Verfügung gestellt. Weitere Dokumente haben wir direkt vom Integrationsbüro erhalten. Ergänzende Unterlagen wurden uns auf informellem Weg zugänglich gemacht. Andere Unterlagen konnten wir im Verlauf der Meta-Evaluation jedoch nicht erhalten. Weil uns nur eine unvollständige Dokumentation vorlag, haben wir uns dazu entschlossen, ein längeres Leitfaden-Interview mit C. Longchamp vollständig zu transkribieren. Das Transkript wurde dem Gesprächspartner vorgelegt und von ihm durchgesehen (siehe dazu unten).

Bei Projektbeginn war vorgesehen, zwei der zehn Fallstudien auszuwählen, um diese vertieft anhand von Interviews mit beteiligten oder betroffenen Akteuren zu untersuchen. Die übrigen acht Evaluationen sollten lediglich anhand von schriftlichen Unterlagen bewertet werden. Im Verlauf der Projektarbeiten hat sich jedoch gezeigt, dass es nicht möglich ist, alleine aufgrund der genannten Unterlagen eine verlässliche Bewertung der Evaluationsstudien vorzunehmen. Aus diesem Grund musste von der ursprünglich geplanten Vorgehensweise abgewichen werden. Wir entschlossen uns zu allen zehn Studien in Ergänzung zur Dokumentenanalyse mit ausgewählten Akteuren Gespräche zu führen. Natürlich konnte dies nicht so detailliert und ausführlich geschehen, wie dies für die zwei typischen Fälle ursprünglich geplant war. Je nach Vollständigkeit der projektbezogenen Unterlagen musste die Zahl der Gesprächspartner angepasst werden. In jedem Fall wurde zumindest mit einem Mitglied des Forschungsteams und mit einem Vertreter des Auftraggebers ein Gespräch geführt. Die weiteren Gesprächspartner wurden aufgrund spezifischer Informationsbedürfnisse oder aufgrund konkreter Hinweise ausgewählt. Die Auswahl erfolgte somit stark pragmatisch. Insgesamt wurden rund siebzig Personen kontaktiert. Eine Liste dieser Personen findet sich ebenfalls als Anhang bei den zehn Fallstudien (vgl. Kapitel 5 sowie Widmer 1996a-g).

Die Länge und die Art der Gespräche waren sehr unterschiedlich. Während in einigen Fällen längere (persönliche) Interviews im Umfang von eineinhalb bis über zwei Stunden erfolgten, beschränkten sich andere Gespräche auf Telefonate von zehn bis fünfzehn Minuten Dauer. Die längeren Gespräche sind mit Unterstützung eines Leitfadens geführt worden. In einem Fall wurde das Gespräch mit einem Tonband aufgenommen und anschliessend im Wortlaut (in Schriftsprache) transkribiert. Bei allen anderen Gesprächen wurde im Anschluss an das Gespräch anhand von Gesprächsnotizen ein Kurzprotokoll erstellt. Die Gespräche wurden unter Zusicherung der Anonymität durchgeführt, wobei die Gesprächspartner darauf aufmerksam gemacht wurden, dass eine Liste der kontaktierten Personen angeführt werden wird, dass aber keinerlei Rückschlüsse auf Aussagen einzelner Personen möglich sein werden. Im Regelfall wurde dieses Vorgehen akzeptiert,

wobei in drei Fällen zwar sehr wohl eine Gesprächsbereitschaft vorhanden war, aber keine Namensnennung gewünscht wurde. Diese Personen sind in den Listen nicht aufgeführt. Die angefragten Personen gaben in den meisten Fällen sehr gerne Auskunft und waren zudem sehr oft bereit, weitere Unterlagen zur Verfügung zu stellen oder selbst weitere Nachforschungen anzustellen. Nur gerade in einem Fall wollte der zuständige Projektleiter mit der Untersuchung nichts zu tun haben und verwies etwas abrupt auf einen Projektbearbeiter ("Ich kann ihnen zu diesem Projekt keine Auskunft erteilen, wenden sie sich an X, die Nummer finden sie im Telefonbuch von Y."). Der daraufhin kontaktierte Projektbearbeiter war sehr zuvorkommend und stellte alle gewünschten Informationen gerne zur Verfügung. Ebenfalls in einem Fall war der zuständige Vertreter des Auftraggebers nicht bereit, mit uns in Kontakt zu treten. Dort wurden wir mehrfach von Mitarbeitern getröstet. Weiter zeigte sich bei jenen Studien, die bereits vor mehr als drei Jahren abgeschlossen worden sind, dass das menschliche Erinnerungsvermögen beschränkt ist. In diesen Fällen waren gewisse Informationen nur noch sehr schwer oder gar nicht mehr zu erhalten. Die schriftlichen Materialien haben in diesen Fällen mehr Gewicht erhalten.

4.4. Auswertung des Datenmaterials

Aufgrund der zur Verfügung stehenden Materialien und den Informationen aus den Gesprächen wurde in einem ersten Schritt eine 'thick description' zum Evaluationsvorgang erstellt. Dabei wurde die Evaluation als sozialer Prozess (Braybrooke/Lindblom 1963) aufgefasst, der in einer Fallbeschreibung rekonstruiert wurde.

Die Auswertung des oben beschriebenen Materials erfolgte in einem zweiten Schritt anhand der bereits in Kapitel 3 vorgestellten Kriterien. Informationen aus den Unterlagen und Gesprächen wurden den dreissig Kriterien zugeordnet. Dabei kam es relativ häufig vor, dass ein bestimmter Aspekt für mehrere Kriterien relevant erschien (vgl. dazu auch Kapitel 3.5.). Die Informationen sind auf Karteikarten zu den einzelnen Kriterien festgehalten worden. Die Kartei wurde sodann als Arbeitsgrundlage zur Erstellung der Fallstudientexte verwendet. Zumeist wurde in einem ersten Schritt die Analyse alleine aufgrund der vorliegenden schriftlichen Unterlagen vorgenommen und erst in einer zweiten Phase die Gespräche geführt. Die Vorgehensweise bei der Zuordnung zu den einzelnen Kriterien entspricht jener der inhaltlichen Strukturierung in einer deduktiv angelegten Inhaltsanalyse, wie sie von Philipp Mayring vorgeschlagen wird (Mayring 1993, insbesondere 83; vgl. auch Lamnek 1989, II: 202-213 und Spöhring 1989: 200-207).

Aufgrund der Struktur des Kriterienrasters ergaben sich für dessen empirische Umsetzung verschiedene Probleme (vgl. Kapitel 3.5.):

- Die fehlende Eindeutigkeit führt zu Wiederholungen in den Fallstudien. Wir versuchten diese durch den häufigen Einsatz von Querverweisen zu reduzieren. Vielfach mussten jedoch trotzdem Wiederholungen angeführt werden, um unsere Argumentation transparent darstellen zu können.
- Die fehlende Einheitlichkeit hat zur Folge, dass bei gewissen Kriterien vorgelagerte Mängel, Lücken, Schwierigkeiten die Bewertung des nachgelagerten Kriteriums mitbeeinflussten.
- Die fehlende Unabhängigkeit der Kriterien versuchten wir mit Querverweisen aufzuzeigen. Dadurch bestand die Möglichkeit, die vielfältigen Interdependenzen unter den verschiedenen Kriterien empirisch zu repräsentieren. Wir versuchten dies zu nutzen, indem wir bei der kritischen Analyse der eingesetzten Kriterien speziell auf diese Interdependenzen achteten (siehe unten, Kapitel 6.4.).

4.5. Vernehmlassung bei den Evaluatoren

Im Verlauf der Arbeiten wurde immer deutlicher, dass es unabdingbar sein wird, die geplante Vorgehensweise um einen zusätzlichen Arbeitsschritt zu erweitern. Um sicherzustellen, dass die empirischen Fallstudien möglichst wenige Fehlerhaftigkeiten, Lücken und Fehleinschätzungen enthalten und um den an den untersuchten Studien beteiligten Evaluatoren die Gelegenheit zu geben, ihre eigene Perspektive zur Evaluation bzw. zur Meta-Evaluation darzulegen, entschlossen wir uns, die Fallstudien den betreffenden Evaluatoren zur Stellungnahme vorzulegen. Die Evaluatoren wurden darum gebeten, sich zur vorgelegten Fallstudie in mündlicher oder schriftlicher Form zu äussern. Zudem sahen wir vor, dass die Evaluatoren die Möglichkeit erhalten sollten, ihre Perspektive unfiltriert in einer schriftlichen Stellungnahme, die den betreffenden Fallstudien unverändert angefügt wird, darlegen zu können. Damit folgen wir auch einer Empfehlung von Michael Scriven (1991: 229-230) und der Praxis des US-amerikanischen General Accounting Office (GAO), das die Evaluationsberichte vor Abschluss einer Evaluation den evaluierten Institutionen zur Stellungnahme unterbreitet.

Dabei wurde ein zweistufiges Verfahren eingesetzt. Aufgrund einer ersten Reaktion der Evaluatoren erfolgte eine Überarbeitung der Fallstudientexte. Die so revidierten Texte wurde den Evaluatoren ein weiteres Mal vorgelegt. Daraufhin formulierten die Evaluatoren - sofern sie dies wünschten - eine schriftliche Stellungnahme.

Diesen Hinweisen ist noch eine Bemerkung anzufügen: Es ist für die Evaluatoren mit einem verhältnismässig grossen - unbezahlten - Aufwand verbunden, auf unseren Wunsch zur Stellungnahme einzutreten. Einerseits wegen des erwähnten Aufwandes andererseits auch deshalb, weil die Fallstudien teilweise erhebliche Kritik enthalten, sind die Fallstudien unterschiedlich aufgenommen worden.

4.6. Vergleichende Analyse

Aufgrund der Fallstudien zu den einzelnen Evaluationen erfolgen zwei weitere Untersuchungsschritte, bei denen einer fallvergleichenden Vorgehensweise gefolgt wurde. Der erste der beiden Schritte befasst sich mit der Frage, welche Aufschlüsse sich für die weitere Praxis der Evaluationsforschung in der Schweiz ergeben. In diesem Teil wird ein spezielles Augenmerk darauf gerichtet, welche Stärken und Schwächen die spezifischen Ausprägungen bei der Evaluationsausgestaltung aufweisen. Weiterhin wird darauf geachtet, ob sich entlang der in Kapitel 4.2. angeführten Dimensionen systematische Strukturen beobachten lassen. Beim zweiten, fallvergleichenden Analyseteil stand die Frage im Mittelpunkt, ob sich der eingesetzte Kriterienraster als Ganzes und in den einzelnen Teilen für die Bewertung von schweizerischen Evaluationen bewährte. Dabei stützen wir uns wiederum auf die Fallstudien zu den einzelnen Evaluationen und auf einen Vergleich unter diesen ab.

Diese vergleichende Analyse ist nicht dazu geeignet, empirisch erhärtete, bewährte Aussagen (im Sinne Poppers 1989: 198-226) zu generieren. Vielmehr handelt es sich um plausible Folgerungen ('plausible inference', Cronbach et al. 1980). Die in dieser Untersuchung eingesetzte Vorgehensweise wird auch als 'multiple case study design' (Greene/ David 1984) oder 'structured case study methodology' (Horvath 1995) bezeichnet.

4.7. Hinweise zu speziellen personellen Konstellationen

Die 'Evaluatoren-Gemeinde' in der Schweiz ist verhältnismässig klein und teilweise mehr Gemeinschaft als Gesellschaft. Für die praktische Durchführung einer Meta-Evaluation ergeben sich dadurch einerseits Vorteile (so etwa der zumeist leichte Zugang zu den Akteuren) aber auch einige problematische Aspekte. Deshalb erscheint es uns an dieser Stelle zweckmässig zu sein, wenn wir unsere Beziehungen zu den beteiligten Evaluatoren offenlegen.

- Der Projektbearbeiter hat bei verschiedenen Gelegenheiten unabhängig von dieser Studie folgende Personen kennengelernt: Andreas Balthasar, Hans-Martin Binder,

Frohmut W. Gerheuser, Peter Knoepfel, Wolf Linder, Claude Longchamp, Stefan Suter und Willi Zimmermann.

- Verschiedene Personen sind dem Projektleiter aus früheren Kontakten bekannt. Dies gilt in besonderer Weise für Hans-Martin Binder, der beim Projektleiter der Meta-Evaluation seine Lizentiatsarbeit verfasst hat und an der Forschungsstelle für Politische Wissenschaft an der Universität Zürich als Assistent tätig war. Durch vielfache Treffen, gleichzeitigen Einsitz in verschiedenen Gremien usw. sind dem Projektleiter seine Kollegen Wolf Linder und Peter Knoepfel bekannt. Im Zusammenhang mit den Nachbefragungen bei eidgenössischen Urnengängen (sog. VOX-Analysen) haben sich verschiedentlich Kontakte mit Claude Longchamp ergeben.
- Der Projektleiter ist Mitglied der Expertengruppe des NFP 27 und befasste sich deshalb auch schon in dieser Funktion mit den Evaluationen 1 bis 6.
- Wolf Linder nimmt bezüglich dieser Meta-Evaluation eine Doppelrolle ein. Er war Projektleiter einer der hier evaluierten Studien, ist aber gleichzeitig (und wie der Leiter der vorliegenden Untersuchung) Mitglied der Expertengruppe des NFP 27, welche diese Meta-Evaluation in Auftrag gegeben, begutachtet und genehmigt hat.

5. DIE EMPIRISCHEN FALLSTUDIEN

Die zehn ausgewählten Evaluationsstudien werden anhand der in Kapitel 3 präsentierten Kriterien in je einer Fallstudie bewertet.

Die Fallstudien sind folgendermassen aufgebaut: Im ersten Kapitel wird die untersuchte Evaluation kurz vorgestellt. Diese Beschreibung enthält Angaben zur Thematik, zur Fragestellung, zum gewählten Vorgehen und zum Auftraggeber sowie zu den Evaluatoren. Im zweiten Abschnitt findet sich eine chronologische Deskription des Evaluationsprozesses. Der Hauptteil der Fallstudie befasst sich sodann im dritten Kapitel mit der Bewertung der Evaluationen. Dabei folgen wir dem Aufbau des Kriterienrasters. Somit gehen wir zuerst auf die Kriterien zur Nützlichkeit, auf jene zur Anwendbarkeit, auf jene zur Korrektheit und zum Schluss auf jene zur Genauigkeit ein. Auf die Bewertung mittels der Einzelkriterien folgt ein kurzer Gesamteindruck zur entsprechenden Evaluation, in der die wesentlichsten Aspekte summarisch dargelegt werden und in einen weiteren Zusammenhang gestellt werden. Hier wird auch verstärkt auf die für den Evaluationsprozess relevanten Rahmenbedingungen eingegangen. Jede Fallstudie wird begleitet von einem Anhang, in dem sich ein Verzeichnis der verwendeten schriftlichen Unterlagen zur Evaluation, ein Verzeichnis der im Rahmen der Meta-Evaluation kontaktierten Personen sowie eine Zusammenstellung der zusätzlich konsultierten Dokumente finden. In einigen Fällen enthält der Anhang zudem weitere Informationen. Im Anschluss an die Fallstudie wird - sofern eine solche vorliegt - die Stellungnahme der betroffenen Evaluatoren angefügt. Diese Texte stammen von den Evaluatoren und wurden von unserer Seite nicht redigiert oder andersweitig inhaltlich verändert.

Die vorliegende Publikation enthält nur drei der insgesamt zehn Fallstudien. Um den Umfang des vorliegenden Textes nicht zu stark anwachsen zu lassen, mussten wir drei der zehn Fallstudien auswählen. Es sind dies die Fallstudien 'Lohngleichheitspostulat des Heimarbeitsgesetzes' (Studie Nr. 3), 'Schneeräumung von Urner Passstrassen' (Studie Nr. 6) und 'Evaluation des BUWAL' (Studie Nr. 9). Die Auswahl wurde so getroffen, dass hinsichtlich der in Kapitel 4.2. diskutierten Dimensionen möglichst verschiedenartige Evaluationsstudien berücksichtigt werden.

Die restlichen sieben Fallstudien sind in der Reihe "Schlussberichte" des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) publiziert worden. Es handelt sich dabei um die folgenden Studien:

- 'Umweltpolitik und technische Entwicklung' (Studie Nr. 1) (siehe Widmer 1996a)
- 'Unternehmerische Innovationsprozesse' (Studie Nr. 2) (siehe Widmer 1996b)

- 'Wohneigentumsförderung durch den Bund' (Studie Nr. 4) (siehe Widmer 1996c)
- 'Evaluation der Zürcher Gemeinschaftszentren' (Studie Nr. 5) (siehe Widmer 1996d)
- 'Evaluation Kehrichtsackgebühr' (Studie Nr. 7) (siehe Widmer 1996e)
- 'Heizkostenabrechnung und Förderung erneuerbarer Energien' (Studie Nr. 8) (siehe Widmer 1996f)
- 'Evaluierung der EWR-Informationskampagne' (Studie Nr. 10) (siehe Widmer 1996g)

Da diese Texte einen integralen Bestandteil dieser Untersuchung darstellen, empfehlen wir dem interessierten Leser, diese Fallstudien ebenfalls zu konsultieren.

5.1. FALLSTUDIE 'LOHNGLEICHHEITSPOSTULAT DES HEIMARBEITSGESETZES' (STUDIE NR. 3)

5.1.1. Die Evaluation

5.1.1.1. Das Thema

Die Evaluation (Gerheuser/Schmid 1993; Schlussbericht (SB)) untersucht die Wirksamkeit des Lohngleichheitspostulats im Heimarbeitsgesetz (HArG, SR 822.31). Das Lohngleichheitspostulat ist im Gesetz folgendermassen formuliert:

"Der Lohn für Heimarbeit richtet sich nach den im eigenen Betrieb für gleichwertige Arbeit geltenden Ansätzen. Fehlt ein vergleichbarer Betriebslohn, so ist der im betreffenden Wirtschaftszweig übliche regionale Lohnansatz für ähnliche Arbeiten anzuwenden. Den unterschiedlichen Arbeitsbedingungen zwischen Betrieb und Wohnort des Heimarbeitnehmers sowie den mit der Heimarbeit verbundenen Mehr- und Minderaufwendungen für Arbeitgeber und Arbeitnehmer ist angemessen Rechnung zu tragen." (HArG, Art. 4, Abs. 1)

Die Studie befasst sich mit einem Wirtschaftsbereich, der insgesamt gesehen seit dem Beginn des Jahrhunderts stark an Bedeutung verloren hat. Die Zahl der Heimarbeitnehmer - zum grössten Teil Frauen - hat stark abgenommen. Weiterhin bezieht sich das Heimarbeitsgesetz lediglich auf industrielle und gewerbliche Heimarbeit, also auf die herkömmlichen Heimarbeitsformen; die Heimarbeit im Dienstleistungssektor untersteht dem Gesetz nicht.

Empirisch untersucht wird die Wirksamkeit des Lohngleichheitspostulats anhand vier verschiedener Wirtschaftsbranchen in vier Regionen der Schweiz.

5.1.1.2. Die Fragestellung

Die Untersuchung verfolgt die Absicht, "zu überprüfen, ob die Heimarbeitnehmer den gleichen Lohn beziehen wie die Betriebsarbeitnehmer" (SB: 2). Diese inhaltliche Zielsetzung verfolgen die Autoren auf zwei Arten. Einerseits klären sie anhand eines Vergleichs der Lohnsituation bei der Betriebsarbeit und bei der Heimarbeit ab, "wieweit das Lohngleichheitspostulat eingehalten wird" (SB: 3). Andererseits wird versucht einen Einblick in die Wirkungsweisen der Regelung zu erhalten, um so die Wirksamkeit des Lohngleichheitspostulates beurteilen zu können.

Daraus sollen sich, so die praktisch-orientierte Zielsetzung, "konkrete Schlussfolgerungen [ergeben], wie das Lohngleichheitspostulat des Heimarbeitsgesetzes aufgrund unserer Erkenntnisse in Zukunft gehandhabt werden soll" (SB: 3, Ergänzung TW).

Aufgrund der methodischen Erfahrungen, welche die Evaluatoren im Projektverlauf erworben haben, sollen auch "einige generelle Schlussfolgerungen zur Evaluation gesetzlicher Regelungen" (SB: 3) angestrebt werden.

5.1.1.3. Das Vorgehen

Zur Evaluation der genannten gesetzlichen Vorschrift wählen die Forscher einen Fallstudien-Ansatz. Sie untersuchen die Situation in vier Wirtschaftsbranchen in verschiedenen Regionen, nämlich Heimarbeit

- für das Militär im Kanton Uri,
- in der Walliser Uhrenindustrie,
- in der Textil- und Bekleidungsindustrie der Ostschweiz und
- im Maschinen- und Apparatebau der Ostschweiz.

Zur Datenerhebung stützt sich die Evaluation vor allem auf die Befragung von Heimarbeitnehmern und -arbeitgebern sowie von Vollzugsbeamten auf eidgenössischer und kantonaler Ebene ab. Insgesamt wurden 54 (bzw. 55; vgl. SB: 135 bzw. 8) Gespräche geführt. Ergänzend wurden sekundärstatistische Analysen und eine Dokumentenanalyse relevanter Materialien (wie Parlamentsprotokolle, Botschaften des Bundesrates, Fachliteratur etc.) beigezogen.

5.1.1.4. Der Auftraggeber

Diese Studie ist im Rahmen des Nationalen Forschungsprogramms 'Die Wirksamkeit staatlicher Massnahmen' (NFP 27) erarbeitet worden.

5.1.1.5. Die Forscher

Das Forschungsteam besteht insgesamt aus vier Personen, nämlich aus André Schmid (Projektleiter), Frohmut Gerheuser, Anton Bellwald und Daniela Perrig-Gmünder. Der Grossteil der Projektarbeit ist von den beiden erstgenannten geleistet worden. Anton Bellwald (Dr. rer. pol., Volkswirtschaftler, Jahrgang 1933) hat sich vor allem zu Beginn des Projektes mit Fragen konzeptioneller Art beschäftigt. Daniela Perrig-Gmünder (lic. iur., Rechtsanwältin) befasste sich mit den juristischen Aspekten des Projekts.

Frohmüt Gerheuser (Jahrgang 1939) hat in Deutschland Soziologie studiert (dipl. Soziologe) und verfügt über eine umfangreiche praktische Erfahrung in angewandter Sozialforschung. Nach einer langjährigen Tätigkeit bei einem privaten Planungsbüro (vgl. etwa Gerheuser 1971, Gerheuser/Schumann 1973, Gerheuser/Castelazzi 1982, Gerheuser 1984, Henz/Gerheuser/Caspar 1984, Caspar/Gerheuser/Henz 1985, Gerheuser et al. 1988, Gerheuser/Sartoris 1989; für einen Überblick vgl. Gerheuser 1988) machte er sich im Jahre 1988 selbständig und eröffnete das 'Büro für Politikberatung und Sozialforschung' in Brugg (AG). Neben anderen Arbeiten entstanden seither drei grössere Studien. Im Auftrag der 'Arbeitsgruppe Gesetzesevaluation' (AGEVAL) befasste sich Gerheuser mit einer Wirkungsabklärung der beruflichen Vorsorge (Gerheuser 1991). Im Rahmen des Nationalen Forschungsprogramms 'Stadt und Verkehr' (NFP 25) erarbeitete er mit einem Forschungsteam eine Studie zur Gewerbeverdrängung in den grösseren schweizerischen Städten (Gerheuser/Perret-Gentil 1993) und vor kurzem ist die im Auftrag der Parlamentarischen Verwaltungskontrolstelle von Gerheuser verfasste Evaluation zur Regionalpolitik des Bundes publiziert worden (Parlamentarische Verwaltungskontrolstelle 1994, siehe auch Spinatsch/Gerheuser 1994).

André Schmid (Jahrgang 1957) ist Betriebswirt (lic. rer. pol., 1982), seit 1985 Sachbearbeiter und seit 1989 auch Geschäftsmitinhaber im Büro 'Planval - Ökonomie / Ökologie / Raumplanung AG' mit Sitz in Brig. Leiter dieses privaten Büros ist der bereits erwähnte Anton Bellwald. Das Büro Planval - damals noch unter dem Namen Bellwald, Jaeger & Partner - hatte im Rahmen des Nationalen Forschungsprogramms 'Regionalprobleme' (NFP 5) zwei Projekte bearbeitet (siehe Abt et al. 1981, Zurschmitten 1984, Bellwald et al. 1985 und Bellwald 1985). Neben der Tätigkeit bei Planval ist André Schmid als Referent an der Schweizerischen Tourismusfachschule in Sierre tätig. Aufgrund unserer Literaturrecherche konnten keine Publikationen von André Schmid eruiert werden.

Das Projekt wurde begleitet durch eine Projektkommission, die sich aus Fachleuten des Bundesamtes für Industrie, Gewerbe und Arbeit (BIGA), des Bundesamtes für Justiz (BJ), des Eidgenössischen Arbeitsinspektorates, eines kantonalen Amtes für Heimarbeit und der Zentralstelle für Heimarbeit zusammensetzte.

5.1.2. Beschreibung des Evaluationsprozesses

Ende Januar 1990 hat André Schmid eine Projektskizze mit dem Titel 'Herkömmliche und neue Formen der Heimarbeit - Lohnvergleich, Entwicklungstendenzen und arbeitsrechtliche Schlussfolgerungen' zuhanden der Expertengruppe des NFP 27 eingereicht. Dieser Themenbereich war im Ausführungsplan dieses NFP explizit vorgeschlagen worden (SNF 1989: 12). Das damalige Forschungsteam umfasste neben dem Projektleiter André Schmid noch Anton Bellwald und Daniela Perrig-Gmünder. In der Skizze wurde von einer Projektdauer von 9 Monaten und einem Aufwand von rund 135'000.- sFr. ausgegangen. Aufgrund der Skizzenevaluation durch die Expertengruppe wurde André Schmid aufgefordert, eine überarbeitete Forschungsskizze vorzulegen. Diese wurde Ende Juni 1990 fertiggestellt. Im Vergleich zur ersten Skizze ist das Projektteam durch Frohmüt Gerheuser ergänzt worden. Sie enthält zudem präzisere Angaben zur Methodik und zum geplanten Vorgehen. Der Projektumfang ist sowohl in zeitlicher (18 statt 9 Monate), wie auch in finanzieller Hinsicht (200'000.- statt 135'000.- sFr.) erweitert worden.

Aufgrund der revidierten Projektskizze entschliesst sich die Expertengruppe dazu, das Forschungsteam zur Ausarbeitung eines Forschungsgesuchs aufzufordern. Die Vorabklärungen bezüglich einer Mitfinanzierung durch die für den Vollzug des Heimarbeitsgesetzes zuständige Bundesstelle, das Bundesamt für Industrie, Gewerbe und Arbeit (BIGA), verlaufen negativ. Es bestehe zwar seitens des BIGA sehr wohl ein Interesse an der Studie, es sei jedoch nicht möglich, einen finanziellen Beitrag zu sprechen. In einem Schreiben an die Programmleitung weist André Schmid darauf hin, dass die Frist zur Gesuchseinreichung wegen äusseren Umständen nicht eingehalten werden könne. Die Expertengruppe erklärt sich damit einverstanden, auf die Frist vom 9. Oktober 1990 zu verzichten.

Am 27. November 1990 kann sodann das endgültige Forschungsgesuch eingereicht werden. Im Gegensatz zu den Projektskizzen wird nun im Gesuch der Schwerpunkt der Untersuchung auf die herkömmlichen Heimarbeitsformen gelegt. Neue Formen der Heimarbeit, vor allem im Dienstleistungssektor, werden nur am Rande behandelt. Das Projektgesuch geht von einer Projektdauer von 14 bis 18 Monaten und Kosten im Umfang von 120'000.- sFr. aus. Der Projektbeginn wird auf den 1. Januar 1991 in Aussicht gestellt. Rückwirkend auf diesen Termin wird das Projekt am 17. Januar 1991 vom Forschungsrat gutgeheissen. Das Forschungsteam befasst sich zu Beginn des Evaluationsprozesses vor allem mit Fragen der methodischen Vorgehensweise, mit der projekt-internen Arbeitsteilung und mit Vorbereitungsarbeiten für die Befragungen. Ende Juni 1991 findet die erste Sitzung der sogenannten Projektkommission, also der

Projektbegleitgruppe, statt. Anschliessend sind vom Projektteam die Befragungen der Vollzugsbeamten und anderer Experten in Angriff genommen worden. Ab Mitte September 1991 begannen auch die Befragungen der Heimarbeitnehmer und Heimarbeitgeber. Am 25. Februar 1992 legen die Evaluatoren den wissenschaftlichen und den finanziellen Zwischenbericht vor. Ende März 1992 findet eine Besprechung zwischen der Programmleitung des NFP 27 (Werner Bussmann) und dem Forschungsteam (André Schmid, Frohmüt Gerheuser) statt. Die Evaluatoren stellen den Stand des Projektes dar. Trotz einer leichten Verzögerung im Projektablauf, soll das Projekt rechtzeitig abgeschlossen werden können. Weiterhin wird über die methodischen Probleme der Evaluation gesprochen. Es wird vereinbart, dass per Ende Juni 1992 ein Schlussberichtsentswurf vorliegen sollte, der sodann in der Projektbegleitgruppe und in der Expertengruppe des NFP 27 zur Diskussion gestellt wird. Dieser Entwurf liegt Mitte August 1992 vor. Neben dem eigentlichen Bericht wird zu diesem Zeitpunkt auch ein Anhang zur Methodendiskussion und eine Zusammenfassung fertiggestellt.

Am 24. August 1992 wird vom Revisorat des SNF eine Verlängerung der Projektdauer um drei Monate bestätigt. Damit endet das Projekt nicht wie geplant auf Ende Juni, sondern erst auf Ende September 1992. Am 1. September 1992 trifft sich die Projektkommission zur Diskussion des Schlussberichtsentswurfs. Bei Programmleitung und Expertengruppe des NFP 27 stösst der Entwurf auf teilweise heftige Kritik. Es wird entschieden, dass der Bericht überarbeitet werden muss. Die Kritik bezieht sich gleichermassen auf konzeptionelle, formale und inhaltliche Aspekte. Lediglich der Anhang mit der Methodendiskussion findet eine einhellig positive Aufnahme. Anlässlich der Besprechung des Schlussberichtsentswurfs erklärt sich Frohmüt Gerheuser bereit, den Schlussbericht im Sinne der geäusserten Kritikpunkte zu überarbeiten. Für einen zweiten Entwurf wird ein Termin Mitte Dezember 1992 vorgesehen. Gerheuser legt dazu als Grundlage einen neuen Gliederungsentswurf vor, der auf ein positives Echo stösst. Aufgrund zeitlicher Indisponibilität ist er jedoch nicht in der Lage, den Entwurf termingerecht fertigzustellen. Neu wird ein Termin auf Ende März 1993 festgelegt. In der Folge wird die Projektdauer um weitere sechs Monate verlängert, womit das Projektende nun auf Ende März 1993 zu liegen kommt.

Ende März wird der Schlussberichtsentswurf von Gerheuser abgeschlossen, so dass er am 7. April 1993 im Rahmen der Expertengruppe diskutiert werden kann. Der überarbeitete Entwurf wird nun von der Expertengruppe positiv beurteilt und formell angenommen. Er soll in der Reihe 'Schlussberichte' des NFP 27 veröffentlicht werden. Anfangs August 1993 liegt der Schlussbericht in gedruckter Form vor. In der Folge wurde die Projektdauer ein weiteres Mal um drei Monate verlängert. Damit verschob sich das Projektende

auf Ende Juni 1993. Aufgrund der von Frohmüt Gerheuser zusätzlich geleisteten Arbeiten wird ein Zusatzkredit im Umfang von 16'264.- sFr. gewährt. Ende Juli 1993 erscheint die Zusammenfassung der Studie in der Reihe 'Kurzberichte' des NFP 27. Am 30. September 1993 erfolgt die finanzielle Schlussberichterstattung durch die Projektleitung.

5.1.3. Bewertung entlang der Kriterien

Die Auswahl des passenden Kriteriensets war bei dieser Studie nicht einfach zu treffen. Es fehlen explizite Hinweise auf eine paradigmatische Ausrichtung. Es ist auch nicht ganz auszuschliessen, dass das Forschungsteam in sich diesbezüglich unterschiedliche Auffassungen vertreten könnte. Aufgrund indirekter Schlüsse kommt jener Kriterienraster zur Anwendung, der sich an einem alternativen Paradigma orientiert. Beispielsweise lässt sich aufgrund von Vorarbeiten der Mitglieder der Forschergruppe darauf schliessen. Es finden sich so etwa Hinweise in der Arbeit Gerheusers zur Evaluation des BVG (Gerheuser 1991). Dort wird die Validität folgendermassen umschrieben: "Als valid sehe ich die Resultate einer solchen Untersuchung dann an, wenn die Erkenntnisse über die untersuchten Sachverhalte umfassend und in ihrem Kontext plausibel und verallgemeinerbar sind." (Gerheuser 1991: 121) Hier von Interesse ist dabei vor allem die Kontextgebundenheit der Erkenntnisse und die holistische Perspektive, die Hinweise auf ein naturalistisches Evaluationsverständnis darstellen (vgl. dazu Lincoln/Guba 1985). Weiter finden sich auch in der Evaluation selbst indirekte Hinweise. So beispielsweise folgende Aussage: "Erst über Einsichten in die gesamthaften Wirkungszusammenhänge lässt sich auf den Einfluss und die 'Wirksamkeit' des Gesetzes schliessen." (SB: 130-131) Die holistische Sichtweise spricht auch hier eher für das alternative Paradigma.

5.1.3.1. Nützlichkeit

5.1.3.1.1. Identifikation der beteiligten Gruppen

Die Studie unterscheidet drei verschiedene Akteurgruppen, nämlich die Heimarbeitgeber, die Heimarbeitnehmer und die Vollzugsorgane. Deren Interessenlagen und Handlungsmöglichkeiten werden detailliert dargestellt und mit den zu evaluierenden Massnahmen in Bezug gesetzt (Kapitel 5.3., 5.4. und 5.5., SB: 76-102). Diese Analyse erfolgt aufgrund der empirischen Untersuchungen, besonders anhand der durchgeführten Befragungen mit Heimarbeitgebern (14 bzw. 13 Gesprächspartner; siehe SB: 8 bzw. 135), Heimarbeitnehmern (34; SB: 8, 135) und Vollzugsbeamten (7, SB: 8, 135). Soweit dies anhand der vorliegenden Informationen beurteilt werden kann, gewährleistet die Datenerhebung eine ausreichende Basis für die Analyse. Bei den Gesprächspartnern, die im Vollzug des Heimarbeitsgesetzes tätig sind, fehlen jedoch Vertreter der Judikative. Die Gerichte werden zwar von den Evaluatoren explizit als Vollzugsträger genannt (SB: 72 und 73), in der empirischen Analyse ist dieser Bereich aber nicht abgedeckt worden. Es ist zu vermuten, dass dies deshalb nicht geschehen ist, da es nur selten zu gerichtlichen Streitfällen kommt (vgl. SB: 91-92, vgl. dazu auch Kapitel 5.1.3.4.5.). Die Befragung der Heimarbeitgeber

war im Forschungsgesuch nicht vorgesehen. Im Bereich der Heimarbeiter sind mehr Gespräche durchgeführt worden, als dies geplant war. Damit hat die Evaluation in diesem Bereich eine deutliche Ausweitung der empirischen Basis erfahren.

Die Projektbegleitgruppe setzt sich, wie bereits erwähnt (Kapitel 5.1.5.), aus Vertretern des Vollzugs zusammen. Namentlich handelt sich um Vertreter verschiedener Bundesstellen (BIGA, BJ, Eidg. Arbeitsinspektorat), eines kantonalen Amtes für Heimarbeit und um einen Vertreter der Zentralstelle für Heimarbeit. Unseres Wissens hat sich die Begleitgruppe lediglich zweimal getroffen, nämlich zu Beginn des Projekts und zur Besprechung des Schlussberichtsentswurfs (ESB). Nicht vertreten sind darin die Heimarbeiter und -geber, die Gerichte sowie die Fraueninteressen. Letzteres erscheint uns deshalb von Bedeutung, da die grosse Mehrheit der Heimarbeiter weiblich sind und sich dadurch potentiell zwei Diskriminierungsfaktoren - Heimarbeit versus Betriebsarbeit und das Geschlecht - überlagern. Die Evaluatoren haben im Forschungsgesuch angekündigt, auch Vertreter der Arbeitgeber- und Arbeitnehmerverbände in die Projektbegleitung zu integrieren (FG: 4). Begründungen für diese Abweichung, wir denken hier etwa an die fehlende Bereitschaft der Interessenorganisationen zur Mitwirkung, liefert der Schlussbericht nicht. Trotz diesen punktuellen Schwachstellen, erscheint die Interessenberücksichtigung in ausreichendem Ausmass erfüllt worden zu sein. Den angeführten Lücken steht eine wesentliche Ausweitung in anderen Bereichen gegenüber. Wünschenswert erscheint uns, im Schlussbericht die Gründe, warum auf den Einbezug bestimmter Akteure verzichtet wurde, zu explizieren. Trotzdem kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt.

5.1.3.1.2. Glaubwürdigkeit des Evaluators

Das Forschungsteam ist sehr disparat zusammengesetzt. Dies betrifft nicht nur die disziplinäre Ausrichtung der beiden Forscher, welche den Hauptteil der Projektarbeit geleistet haben, sondern auch die regionale Herkunft und den Erfahrungsschatz der beiden.

Frohmut Gerheuser ist deutscher Staatsangehöriger, hat Soziologie studiert und ist seit sehr vielen Jahren im Bereich der angewandten Sozialwissenschaft und der Politikberatung tätig. Er gehört hier sicherlich zur Gruppe der erfahreneren Personen in der Schweiz. Er hat sich im Verlauf seiner Tätigkeit in verschiedenen Politikbereichen fundierte Kenntnisse erworben, die ihm auch erlauben, in für ihn neuen Politikfeldern schnell Fuss zu fassen. Aus der Projektarbeit, die sehr oft auch interdisziplinär angelegt war, ist er auch in der Lage, sich auf unterschiedlichste Kontexte einzustellen. Weiterhin verfügt Gerheuser über fundierte Erfahrungen mit evaluativen Fragestellungen (Gerheuser 1991).

Er hat überdies in verschiedenen Nationalen Forschungsprogrammen (NFP) Projekte bearbeitet, so dass ihm auch dieser Kontext sehr wohl bekannt ist. Aufgrund dieser Überlegungen kann ihm sicherlich eine hohe Glaubwürdigkeit attestiert werden.

Die Bewertung der Glaubwürdigkeit des Projektleiters, *André Schmid*, fällt uns relativ schwer, weil wir nur wenige Kenntnisse von seiner Tätigkeit haben. Gemäss diesen Informationen verfügt Schmid über einen deutlich kleineren Erfahrungshorizont als Gerheuser. Im Vergleich zu diesem ist er auch noch ziemlich jung. Seine Tätigkeiten beim Büro Planval scheinen sich vor allem auf die Beratung konzentriert zu haben. Wie bereits erwähnt, sind wir auf keine Publikationen aus seiner Feder gestossen. Wie weit er sich in den sozialwissenschaftlichen Vorgehensweisen auskennt, können wir hier nicht beurteilen. Andererseits hat Schmid durch seine Herkunft - er stammt aus dem Kanton Wallis und ist seit einiger Zeit dort tätig - sicherlich einen guten Zugang zu den entsprechenden Stellen in diesem Kanton. Er kennt die Perspektive der Bergregionen besonders gut, was bei mindestens zwei der vier Fallstudien von Vorteil ist.

Etwas ungewöhnlich ist der Umstand, dass bei diesem Forschungsteam nicht der Ältere und Erfahrenere der beiden Hauptbeteiligten die Projektleitung übernommen hat. Dies mag damit in Zusammenhang stehen, dass ja, wie oben bereits dargestellt worden ist (vgl. Kapitel 5.1.2.), Gerheuser erst später zur Mitarbeit an diesem Projekt beigezogen worden ist und Schmid der Initiator des Projekts war. Unseres Wissens haben die beiden Forscher vor diesem Projekt noch nie zusammengearbeitet. Die Arbeitsteilung ist so erfolgt, dass Schmid die Fallstudien in der West- und Zentralschweiz bearbeitete, während Gerheuser die Untersuchungen in der Ostschweiz durchführte. Die Zusammenarbeit ist gegen Ende des Projekts nicht besonders harmonisch verlaufen.

Aus den obenerwähnten Gründen fällt eine gesamthafte Beurteilung der Glaubwürdigkeit des Evaluationsteams schwer. Die Verschiedenheit der beiden Forscher kann sich sowohl als sehr ergiebig wie auch als äusserst schwierig erweisen. Für eine fundiertere Analyse bedarf es hier aber der Durchführung von weitergehenden Befragungen. Nur aufgrund der Dokumentenanalyse (siehe Kapitel 5.1.5.1. und 5.1.5.3.) und der von uns geführten kurzen Gespräche (siehe Kapitel 5.1.5.2.) ist eine Einschätzung nicht möglich.

5.1.3.1.3. Informationsrahmen und -auswahl

Die Evaluation bedient sich verschiedener Informationsquellen. Dies sind:

- Befragungen von Vollzugsbeamten, Heimarbeitgebern und Heimarbeitnehmern,
- statistische Unterlagen und Fachliteratur sowie

- Botschaften des Bundesrates und Parlamentsprotokolle.

Die Autoren messen den Quellen unterschiedliche Bedeutung zu: "Im Zentrum standen die Expertengespräche und Befragungen." (SB: 6) Etwas unklar bleibt dabei die Unterscheidung in Expertengespräche und Befragungen. Der Schlussbericht gibt dazu leider keine weiteren Hinweise. Die Befragungen wurden verteilt über die vier Branchen und Regionen geführt. Es handelt sich dabei um Leitfadengespräche mit einer Dauer von ein bis zwei Stunden (SB: 136). Die Leitfäden - deren drei, je für eine Befragtengruppe - sind im Schlussbericht nicht enthalten. Ein Teil der Gespräche in der Ostschweiz wurden telephonisch geführt. Diese Gespräche dauerten nur etwa halb so lang wie die persönlichen Gespräche, "waren aber ebenso ergiebig" (SB: 136). Die Gespräche wurden aufgrund von handschriftlichen Notizen ausgewertet. Auf Tonbandprotokolle wurde gänzlich verzichtet. Die Auswahl der Befragten erfolgte in sehr unterschiedlicher Weise:

- Bei den *Vollzugsbeamten* ergab sich die Auswahl quasi von selbst. Es wurden sämtliche eidgenössischen Arbeitsinspektoren befragt. Auf kantonaler Ebene wurden alle Arbeitsinspektoren interviewt, in deren Kantone auch Befragungen von Heimarbeitgebern und Heimarbeitnehmern stattfanden. Man kann hier also von einer Vollerhebung sprechen. In den Kantonen Solothurn und Aargau wurde lediglich ein telephonisches Informationsgespräch geführt. Auf eine eigentliche Befragung wurde verzichtet, da der Vollzug im interessierenden Bereich wegen der kleinen Zahl von Heimarbeitnehmern ruht. Insgesamt wurden mit sieben Vollzugsbeamten Befragungen und mit zwei weiteren Kurzgespräche geführt.
- Die *Heimarbeitgeber* in den ausgewählten Regionen und Branchen wurden aufgrund von Hinweisen der kantonalen Arbeitsinspektorate ausgewählt, wobei die Erreichbarkeit und die Branchenzugehörigkeit zentrale Auswahlkriterien darstellten. Es wurden insgesamt mit dreizehn (oder vierzehn) Heimarbeitgebern Befragungen durchgeführt (SB: 8 und 135).
- Die Bestimmung der Gesprächspartner seitens der *Heimarbeitnehmer* erfolgte in den verschiedenen Kantonen auf unterschiedliche Art. In der Ostschweiz und im Wallis mussten die Kontakte - mangels anderer Möglichkeiten - via die Heimarbeitgeber aufgebaut werden. In anderen Regionen schlugen lokale Gewährspersonen oder andere Heimarbeitnehmer die Gesprächspartner vor. Weiter wurden Heimarbeitnehmer befragt, die in bekannten Heimarbeitsgruppen tätig sind. Gesamthaft sind 34 Heimarbeitnehmer befragt worden (SB: 8 und 135).

Wegen der Auswahlmechanismen war davon auszugehen, dass tendenziell eher eine positive Auswahl zustande kam. Deshalb wurde versucht, auch Gespräche mit nicht-registrierten Heimarbeitsgebern zu führen und Kontakte mit Heimarbeitnehmern nicht nur über die Heimarbeit ausgebende Firma herzustellen. Generell lässt sich sagen, dass sich die Auswahl der Gesprächspartner sehr schwierig gestaltete. Die Evaluatoren haben sich darum bemüht, eine ausgewogene Auswahl zu treffen. Wie die Evaluatoren selbst auch bemerken, ist die Auswahl der Befragten - trotz den getroffenen Vorkehrungen "sicher etwas einseitig" (SB: 137) ausgefallen. Wie relevant der nicht erfasste Bereich ist, kann nicht beantwortet werden. Die Evaluatoren waren sich dieser Problematik bewusst und haben Massnahmen getroffen, um die Verzerrung möglichst klein zu halten (vgl. dazu die Hinweise in SB: 137-138). Trotz diesen Vorsichtsmassnahmen verbleibt eine gewisse Unsicherheit darüber, ob und in welchem Ausmass es im Bereich der Heimarbeit zu Missbräuchen kommt (siehe auch SB: 104).

Zur Abstützung der Erkenntnisse aus den Befragungen wurden weitere Informationsquellen beigezogen. Eine bedeutende Rolle spielten dabei die quantitativen Angaben, die aus verschiedenen Quellen stammen:

- Eidgenössische Industriestatistik (SB: 13, 15, 17, 18, 44, 50 und 58)
- Datenmaterial des Amtes für Heimarbeit des Kantons Uri (SB: 35 und 36)
- Lohnerhebungen des BIGA (SB: 38, 45, 51, 52 und 59)
- Daten aus Iseli (1984) (SB: 13, 21 und 22)

Die Datenlage bezüglich der quantitativen Verbreitung der Heimarbeit, wie auch hinsichtlich der Lohnsituation in diesem Sektor, scheint nicht sonderlich gut zu sein. Die verfügbaren Materialien sind zudem oft nur schwer in zweckmässiger Weise einsetzbar. Diese Probleme liegen nicht im Verantwortungsbereich der Evaluatoren. Grundsätzlich haben die Autoren versucht, die vorhandene - schwache - Datenbasis so gut wie möglich einzusetzen. Die Evaluatoren haben auf die Primärerhebung quantitativen Datenmaterials verzichtet. Wir werden zu einem späteren Zeitpunkt nochmals auf diese Problematik eingehen (vgl. Kapitel 5.1.3.4.8.).

Die Auswertung der Fachliteratur zur Thematik Heimarbeit ist, soweit dies aus dem Schlussbericht hervorgeht, vor allem als Hintergrundinformation verwendet worden. Bis auf jene Stellen des Schlussberichts, die sich stark auf eine historische Seminararbeit abstützen (Iseli 1984, siehe SB: 12-13 und 19-22), sind Hinweise auf die Fachliteratur nur selten anzutreffen. Hierbei von Interesse wären etwa die beiden Dissertationen zur Heimarbeit (Regli 1978 und Stadler 1982). Die Arbeit von Hansruedi Stadler (1982) beschäftigt sich aus juristischer Perspektive mit dem hier zu evaluierenden HArG. Es findet sich

zwar im Forschungsgesuch ein Hinweis auf diese Arbeit (FG: 3), im Schlussbericht ist sie jedoch nicht einmal mehr im Literaturverzeichnis aufzufinden. Die Doktorarbeit von Felix Regli (1978) wiederum ist zwar schon vor dem Inkrafttreten des HArG von 1981 entstanden, enthält jedoch eine Fülle von Informationen aus volkswirtschaftlicher und sozialpolitischer Sicht zur Heimarbeit in der Schweiz und besonders im Kanton Uri. Diese Arbeit findet sich im Literaturverzeichnis der Evaluation (SB: 151), konkrete Hinweise im Text konnten wir jedoch nicht eruieren. Es finden sich diverse weitere Texte im Literaturverzeichnis, bei denen nicht klar wird, ob und wie sie in der Evaluation verwendet wurden. Konsultiert man den Schlussberichts*entwurf*, der von der Expertengruppe des NFP 27 zur Überarbeitung zurückgewiesen wurde, stösst man auf eine Erklärung. Einige Teile des Textes, die sich sehr stark auf die Literatur abstützten, sind bei der Überarbeitung gekürzt oder gestrichen worden. Das Literaturverzeichnis hingegen scheint mehr oder weniger unverändert übernommen worden zu sein.

Die Informationen aus den Botschaften des Bundesrates und den Parlamentsprotokollen finden ihren Niederschlag vor allem in den Kapiteln zu den Zielen staatlicher Massnahmen zur Heimarbeit (SB: 23-27) und zum Lohngleichheitspostulat des HArG von 1981 (SB: 27-29). Der Beizug der Informationsquellen ist zweckmässig erfolgt. Die Hinweise auf die Fundstellen zu den Debatten im Parlament sind jedoch nicht korrekt. Diese finden sich nicht im Bundesblatt, wie dies in der Studie angegeben wird, sondern in den Amtlichen Bulletins der Bundesversammlung (siehe auch Kapitel 5.1.3.4.4.).

Gesamthaft gesehen haben die Evaluatoren Informationen zum Evaluandum beschafft, die der Bearbeitung der Fragestellung dienen und die auch die Interessen der Beteiligten und Betroffenen berücksichtigen. Das Kriterium wird erfüllt.

5.1.3.1.4. Bewertende Interpretationen

Die Argumentationslinien der Evaluatoren werden mehrheitlich klar dargelegt. Teilweise vermittelt die Studie jedoch den Eindruck, dass sich die Autoren nicht auf einen klaren Standpunkt festlegen wollten, um sich 'die Finger nicht zu verbrennen' oder vielleicht auch deshalb, weil sich das Forschungsteam damit schwergetan hat, sich auf eine gemeinsame Position festzulegen. Besonders deutlich zeigt sich dies in jenem Kapitel, in dem die praktischen Schlüsse aus der Evaluation gezogen werden (Kapitel 7, SB: 109-116). Dort stellen die Evaluatoren vier Szenarien vor, die von der 'Beibehaltung des status quo' über die 'Verstärkung des Vollzugs des Lohngleichheitspostulats' bis zur 'Aufhebung des gesamten Heimarbeitsgesetzes' reichen (vgl. dazu auch Kapitel 5.1.3.4.10.). Auch im Schlusskapitel (Kapitel 7.6., SB: 115-116) vermeiden die Autoren

deutliche Positionsbezüge zugunsten des einen oder des anderen Szenarios. Zur Illustration ein Zitat dazu: "Für jede dieser Strategien lassen sich gute Gründe finden - und selbstredend auch dagegen." Das Forschungsteam scheint nicht in der Lage oder nicht willens gewesen zu sein, sich auf eine klare Handlungsempfehlung festzulegen. Es wird auch darauf verzichtet, genauer auf die mit den Szenarien verbundenen Konsequenzen einzutreten. So weisen die verschiedenen Szenarien in unterschiedlichen Bereichen je ihre spezifischen Vor- und Nachteile auf, wie etwa die Kostenfolgen bei einer Intensivierung des Vollzugs oder die politischen Kosten und Risiken eines nur ansatzweise implementierten Gesetzes beziehungsweise einer Gesetzesrevision. Mit einer systematischen Aufarbeitung dieser Konsequenzen hätten durch die Evaluation Entscheidungsgrundlagen für eine (politische) Einschätzung bereitgestellt werden können (vgl. dazu EVD 1995).

Betrachten wir die vier, von den Evaluatoren vorgeschlagenen Strategien im Einzelnen, ist grundsätzlich festzuhalten, dass die - sich widersprechenden - Argumentationen klar dargelegt werden und für den Leser einigermaßen plausibel erscheinen. Dies gilt jedoch nur mit Einschränkungen: Im Szenario 4 (SB: 114) wird dafür plädiert, das Heimarbeitsgesetz zu streichen, da das Lohngleichheitsprinzip heute unbestritten ist und da der Schutz der Heimarbeiter durch das Arbeitsgesetz (ArG, SR 822.11) und das kommende Gleichstellungsgesetz (vgl. BBl 1995 II: 382-388) ausreichend gewährleistet ist. Dazu ist zu bemerken, dass die Evaluatoren in ihrer Untersuchung lediglich einen - wenn auch wichtigen - Teil der Vorschriften des Heimarbeitsgesetzes evaluierten. Deshalb kann die Empfehlung zur Streichung des *gesamten* HArG empirisch nicht gestützt werden. Weiterhin ist darauf hinzuweisen, dass das ArG (zumindest in der bestehenden Form) keinen Beitrag zum Schutz der Heimarbeiter leisten kann, da dieses Gesetz auf die Heimarbeit nicht anwendbar ist (ArG, Art. 3, lit. f). Der Gesetzgebungsprozess zum Gleichstellungsgesetz wiederum, war zum Zeitpunkt der Fertigstellung des Schlussberichts zur Evaluation 'Lohngleichheitspostulat des Heimarbeitsgesetzes' Ende März 1993 noch nicht abgeschlossen. Der Bundesrat verabschiedete Ende Februar 1993 die Botschaft (BBl 1993 I: 1248). Das Parlament hat das Gesetz Ende März 1995 gutgeheissen. Vor kurzem, am 3. Juli 1995, ist die Referendumsfrist unbenutzt abgelaufen. Das Gleichstellungsgesetz ist aber auch heute noch nicht in Kraft gesetzt worden. (vgl. dazu auch Kapitel 5.1.3.4.10.)

Eine Einschätzung, ob die Evaluatoren die Grundlagen für die Werturteile ausreichend klar dargelegt haben, ist nicht möglich, da die Evaluatoren diese Werturteile im Schlussbericht nicht vornehmen.

5.1.3.1.5. Klarheit des Berichts

Der Schlussbericht der Evaluationsstudie von Frohmut Gerheuser und André Schmid ist folgendermassen aufgebaut:

1. *Die Ziele der Arbeit (SB: 1-4)*

2. *Vorgehen (SB: 5-9)*

3. *Der gesetzliche Heimarbeitnehmerschutz im Spannungsfeld unterschiedlicher Ziele und Interessen (SB: 11-31)*

Dieses Kapitel umfasst einen historischen Überblick zur Heimarbeit in der Schweiz und speziell zur Lohnsituation in der Heimarbeit sowie eine Diskussion der Ziele der Gesetzgebung zur Heimarbeit mit einem besonderen Augenmerk auf das Lohngleichheitspostulat im HArG von 1981.

4. *Die Lohn- und Einkommenssituation der Heimarbeitnehmer (SB: 33-70)*

Das Kapitel enthält die vier Fallstudien sowie die für die vier Fallstudien gleichermassen gültigen Untersuchungsergebnisse.

5. *Chancen und Bedingungen zur Verwirklichung des Lohngleichheitspostulates (SB: 71-102)*

Hier wird die Analyse der Perspektiven der verschiedenen Akteurguppen auf dem Hintergrund eines Wirkungsmodells dargelegt.

6. *Bedingt wirksam, bedingt vollziehbar, bedingt nötig. (SB: 103-107)*

Dieses Kapitel enthält einen Zusammenschluss der Untersuchungsergebnisse, die dem Lohngleichheitspostulat des Heimarbeitsgesetzes eine bedingte Wirksamkeit, eine bedingte Vollziehbarkeit und eine bedingte Notwendigkeit attestieren.

7. *Praktische Schlüsse in Form von vier Szenarien (SB: 109-116)*

Hier finden sich die praktischen Schlussfolgerungen, die in der Form von Szenarien dargelegt werden.

ANHANG:

8. *Die wichtigsten gesetzlichen Regelungen (SB: 119-123)*

9. *Methodendiskussion (SB: 125-148)*

10. *Literaturverzeichnis (SB: 149-154)*

Das Inhaltsverzeichnis, oder vielleicht besser die Inhaltsübersicht, wie sie sich im Schlussbericht findet (SB: 1), entspricht den oben kursiv wiedergegebenen Angaben. Der Schlussbericht enthält somit kein detaillierteres Inhaltsverzeichnis. Wie sich der obenstehenden Übersicht entnehmen lässt, ist die Aussagekraft der Überschriften nicht besonders hoch. Der Leser hat Mühe, sich allein anhand der Titelsetzung zu orientieren. Hier hätten sicherlich aussagekräftigere Überschriften verwendet werden können. Zudem ist der Aufbau des Schlussberichts doch relativ flach gestaltet worden. Wir meinen damit den Umstand, dass sich verschiedene Kapitel finden, die durchaus auch in andere Kapitel hätten integriert werden können. So hätten die ersten beiden Kapitel sicherlich zu einem

zusammengefasst werden können (ebenso Kapitel 6 und 7 bzw. 8 und 10). Die Studie hätte unseres Erachtens mit einem hierarchischeren Aufbau an Übersichtlichkeit gewonnen.

Die Leserführung im Schlussbericht hätte gewiss noch ausgebaut werden können. Ansätze dazu finden sich im Kapitel zu den Zielen der Arbeit (SB: 3). Aus diesen Hinweisen wird ersichtlich, was die Kapitel 4, 5, 7, und 9 enthalten. Hinweise zu den anderen Kapiteln fehlen jedoch.

Die Arbeit enthält sowohl ein Literaturverzeichnis als auch ein Abkürzungsverzeichnis. Von der Darstellung her kann die Studie als gelungen bezeichnet werden. Sprachlich ist der Text ebenfalls gut ausgefallen. Die Sprache des Berichts zeichnet sich durch hohe Klarheit aus. Nur selten lässt der Text Fragen offen. Sprachliche Fehler kommen zwar vor, sind jedoch relativ selten.

Das Kriterium wird von der Studie erfüllt, wenn auch besonders bezüglich des Aufbaus der Studie und der Leserführung noch Optimierungspotential vorhanden gewesen wäre.

5.1.3.1.6. Verbreitung des Berichts

Die Studie ist in der Reihe 'Schlussberichte' des NFP 27 publiziert worden. Soweit uns Informationen vorliegen, haben die an der Evaluation beteiligten Personen - also etwa die Interviewpartner - ein Exemplar der Studie erhalten. Der Bericht kann gegen einen Preis von 10.- sFr. beim Schweizerischen Nationalfonds (SNF) bezogen werden.

In der Reihe 'Kurzberichte' des NFP 27 ist auch eine zehnsseitige Zusammenfassung der Studie erschienen. Diese Kurzfassung kann von Interessenten gratis bezogen werden.

Die verwendeten Publikationsformen weisen nicht einen überaus hohen Verbreitungsgrad auf. Sie stellen jedoch auf alle Fälle sicher, dass sich interessierte Personen Zugang zu den Berichten verschaffen können. Dem Kriterium ist deshalb in ausreichendem Masse nachgekommen worden.

5.1.3.1.7. Rechtzeitigkeit des Berichts

Da die Studie im Auftrag des Schweizerischen Nationalfonds (SNF) beziehungsweise der Expertengruppe des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) erstellt wurde, kann die Rechtzeitigkeit einmal hinsichtlich des politi-

schen Umfeldes, in dem sie sich bewegt, und zum zweiten bezüglich des NFP 27 beurteilt werden.

Der für die Studie ausschlaggebende gesetzgeberische Erlass, das Heimarbeitsgesetz (HArG, SR 822.31), ist seit dem 1. April 1983 in Kraft. Es handelt sich somit um eine Massnahme, die sich seit längerer Zeit im Vollzug befindet und sich damit für eine ex-post-Evaluation bestens eignet. Das Heimarbeitsgesetz ist ein Erlass, der seit einiger Zeit durch die schwindende quantitative Verbreitung der herkömmlichen Heimarbeit in der Schweiz immer mehr an Bedeutung verloren hat. Die Gesetzgebung entbehrt einer hohen politischen Brisanz. Davon auszunehmen ist vielleicht der Bereich der neuen Heimarbeitsformen (Tele-Kommunikation), die sich jedoch bisher auch nicht so stark verbreitet haben, wie dies vorerst angenommen worden war. Gerade dieser Aspekt wird aber, wie wir bereits festgestellt haben, in dieser Evaluation nicht behandelt. Sie beschränkt sich auf herkömmliche Heimarbeitsformen. Damit handelt es sich beim evaluierten Gegenstand um einen Teil eines Erlasses, der bereits über eine längere Vollzugstradition verfügt und der gleichzeitig kaum in der Lage ist, in der politischen Debatte hohe Wellen zu schlagen. Kurz nach Projektbeginn der hier diskutierten Evaluation, am 16. Januar 1991, hat der Bundesrat - aufgrund des Schlussberichts der Arbeitsgruppe 'Lohngleichheit für Mann und Frau' (EJPD 1988) - dem EVD (BIGA) den Auftrag erteilt, "in Zusammenarbeit mit dem EJPD (BJ) und dem EDI (Büro für Gleichstellung) zu prüfen, wie der Schutz der Heimarbeiterinnen" ... "verstärkt werden könnte." (EVD 1995: 2, Auslassung TW; vgl. dazu auch BBl 1993 I: 1275) Die zur Prüfung dieser Frage eingesetzte Arbeitsgruppe präsentierte neulich ihren Bericht zuhanden des Bundesrates (EVD 1995). Dieser Bericht stützt sich massgeblich auf Erkenntnisse der Evaluation ab. Der Schlussbericht der Evaluation lag unseres Wissens rechtzeitig vor. Die Fortschritte im Rahmen der Arbeitsgruppe wurden kaum durch die noch nicht abgeschlossene Evaluation verzögert. Die Rechtzeitigkeit ist somit hier sicherlich gegeben. Da der Heimarbeit in der Wirtschaftswelt teilweise eine Pufferfunktion zukommt, ist davon auszugehen, dass die Erkenntnisse der Studie auch konjunkturabhängig sind. Aus dieser Warte betrachtet, hätte die Studie, wäre sie zwei Jahre später in Angriff genommen worden, möglicherweise andere Resultate gezeitigt.

Bezüglich des NFP 27 hingegen ist eine andere Beurteilung vorzunehmen. Das Projekt wurde vom SNF mit einer Projektdauer von 18 Monaten bewilligt, nachdem die Forscher im Forschungsgesuch eine Projektdauer von 14 bis 18 Monaten beantragt hatten. Das Projektende wurde auf Ende Juni 1992 festgelegt. Im Verlauf des Evaluationsprozesses ist die Projektdauer mehrmals verlängert worden, und zwar im August 1992 um drei Monate, im Januar 1993 um sechs Monate und im September 1993 um weitere drei Monate.

Die letzte Verlängerung stand im Zusammenhang mit der Überarbeitung des Schlussberichts und dem dafür gesprochenen Zusatzkredit. Insgesamt hat sich damit die Projektdauer von den vorgesehenen 18 Monaten auf 30 Monate, also um ein ganzes Jahr verlängert. Formell ist das Projekt auf Ende Juni 1993 abgeschlossen worden. Der Schlussbericht lag jedoch bereits Ende Mai 1993 vor. Weiterhin ist anzufügen, dass der Forschungsrat den positiven Entscheid erst Mitte Januar 1991 getroffen hat, das Forschungsprojekt jedoch bereits auf Anfang Januar 1991 beginnen sollte. Unter Beachtung dieser beiden Aspekte wurden für das Projekt statt der geplanten 18 Monate deren 28 benötigt. Wir können somit festhalten, dass das Forschungsteam die vorgesehene Projektdauer nicht einhalten konnte. Bezüglich des SNF ist die Schlussberichterstattung nicht rechtzeitig erfolgt.

Aufgrund dieser Feststellungen kommen wir zum Schluss, dass die Evaluation das Kriterium teilweise erfüllt.

5.1.3.1.8. Wirkung der Evaluation

Da diese Beurteilung im wesentlichen auf einer Dokumentenanalyse beruht und nur gerade punktuell kürzere Gespräche mit potentiellen Adressaten der Studie geführt wurden, liegen uns nur wenige Informationen zu den Wirkungen dieser Evaluation vor.

In der Presse hat der Schlussbericht der Studie kein grosses Echo hervorgerufen. Wir fanden lediglich einen Hinweis in einem Zeitungsartikel (Cash vom 15. März 1994). Dort wird festgehalten, dass Colette Nova, zuständig für das HArG im Bundesamt für Industrie, Gewerbe und Arbeit (BIGA), die Studie nicht für besonders brauchbar hält. Sie zieht aus der Untersuchung - gemäss der Darstellung im Zeitungsartikel - folgende Konsequenz: "nicht viel Brauchbares: Am besten ist es wohl, wenn wir gar nichts machen." (Cash vom 15. März 1994) Damit stimmt sie mit den Autoren der Studie in gewisser Weise überein, wenn sie schreiben: "Ein dringender politischer Handlungsbedarf besteht nach unserer Einschätzung derzeit nicht." (SB: 115) Bei den von uns kontaktierten Personen, die an der untersuchten Thematik interessiert sind, haben wir festgestellt, dass die Studie keine nachhaltigen Eindrücke hinterlassen hat. Weiter ist aus Aussagen von Mitgliedern der Projektbegleitgruppe zu schliessen, dass der Evaluationsprozess nicht als sehr intensiv empfunden wurde. Angesprochene Personen, die sich auf kantonaler Ebene mit dem Vollzug des Heimarbeitsgesetzes befassen, aber nicht direkt in die Evaluation involviert waren, kennen die Studie nicht. Dass die Evaluation grosse Aufmerksamkeit in der Öffentlichkeit erhalten würde, war aber schon von der Ausgangslage her nicht zu erwarten gewesen. Hingegen hat die Studie für die verwaltungsinterne Arbeitsgruppe des

Bundes, bestehend aus Vertretern des BIGA (EVD), des BJ (EJPD) und des Büros für Gleichstellung (EDI), eine wertvolle Grundlage bereitgestellt. Der Bericht dieser Arbeitsgruppe (EVD 1995), an dem Colette Nova (BIGA) massgeblich beteiligt war, bezieht sich sodann auch stark auf die hier untersuchte Evaluationsstudie. Der Bericht nimmt die vier, in der Evaluation erarbeiteten, Szenarien auf und wägt deren Vor- und Nachteile gegeneinander ab. Er kommt zum Schluss, dass die Beibehaltung des status quo zu präferieren ist, da die alternativen drei Szenarien bei namhaften Nachteilen keine wesentlichen Verbesserungen bringen können. Die Evaluation ist damit in der Bundesverwaltung nicht ungelesen in einer Schublade verschwunden, sondern wurde dort sehr wohl zur Kenntnis genommen und diente als Ansatzpunkt für die eigenen Arbeiten. Es zeigt sich, dass die in der Zeitschrift *Cash* zitierte Aussage von Colette Nova (BIGA) einen einseitigen Eindruck hinterlässt. Die Studie wurde auf alle Fälle gebraucht. Auf der kantonalen Ebene, also bei den Kantonalen Ämtern für Industrie, Gewerbe und Arbeit (KIGA), scheint die Evaluation nicht auf ein besonders grosses Interesse gestossen zu sein. Das könnte damit in Zusammenhang stehen, dass auf kantonaler Ebene bei schwächerer Ausdifferenzierung der Strukturen, hoher Problemlast sowie (für das HARG) ungünstiger politischer Prioritätenfestlegung auch ein direkter Adressat für die Studie fehlt.

Inwiefern die Evaluation zu einer Erweiterung der Handlungsalternativen oder zu veränderten Problemperzeptionen führte, kann aufgrund der von uns vorgenommenen Abklärungen nicht ausreichend fundiert entschieden werden. Für eine Bewertung sind weitergehende Abklärungen erforderlich.

5.1.3.2. Anwendbarkeit

5.1.3.2.1. Praktische Verfahren

Die Evaluation arbeitet an zentraler Stelle mit Befragungen. Dabei wurden sowohl Heimarbeitnehmer, Heimarbeitgeber wie auch die zuständigen Vollzugsbeamten befragt. Als Ergänzung dazu wurden die relevante Fachliteratur und offizielle Dokumente (Botschaften des Bundesrates, Parlamentsprotokolle) beigezogen. Daneben wurden die relevanten Statistiken konsultiert.

Grundsätzlich handelt es sich dabei um einen gängigen Methodenmix, der relativ wenig an Risiken beinhaltet. Das verwendete Instrumentarium ist gegen Störungen wenig anfällig. Im Laufe des Forschungsprozesses musste das ursprünglich geplante Vorgehen mehrfach modifiziert werden. So konnten nicht alle vorgesehenen Regionen untersucht werden, da in diesen teilweise praktisch keine Heimarbeit mehr ausgegeben wird. Weiter

zeigte sich relativ bald, dass es nicht zweckmässig ist, auf die Befragung von Heimarbeitgebern zu verzichten, wie dies vorerst geplant war. Auch die Zahl der Gespräche wurde ausgeweitet, um eine breitere empirische Basis zu erhalten. Da die Auswahl der Interviewpartner potentiell eine positive Auswahl darstellen könnte, wurden zu den Heimarbeitgebern und Heimarbeitnehmern alternative Zugänge gesucht. Durch die hohe Flexibilität der Verfahren konnte die Anfälligkeit auf mögliche Störungen reduziert werden.

Die grösste Veränderung im Untersuchungsdesign stellt jedoch eine veränderte Sichtweise dar, die dazu führte, dass auch verfahrensmässige Änderungen nötig wurden. Die Evaluatoren sind im Verlauf des Projektes davon abgekommen, als zentralen Bestandteil ihrer Untersuchung einen direkten Vergleich der Löhne bei Heimarbeit und bei Betriebsarbeit vorzunehmen, wie dies auch durch das Lohngleichheitspostulat im Heimarbeitsgesetz suggeriert wird. Die Evaluatoren kamen zur Ansicht, dass es zweckmässiger sei, Einblicke in das Wirkungsgefüge des Programms zu erhalten, um von daher auf die Wirksamkeit des Programms zu schliessen. Lohnvergleiche sind zwar nicht ganz unterblieben, doch sie stellen nicht den zentralen Bestandteil der Untersuchung dar. Sie wurden auch mit relativ kleinem Aufwand realisiert.

Der Bezug der relevanten Statistiken erwies sich als nicht immer ganz einfach. Vielfach zeigten die zur Verfügung stehenden Statistiken Erfassungslücken, Ungenauigkeiten oder eine mangelnde Differenzierung der Kategorien.

Gesamthaft gesehen ist festzustellen, dass sich das Untersuchungsdesign während des Forschungsprozesses deutlich gewandelt hat. Die Verfahren, besonders die Befragungen, erwiesen sich durch ihre hohe Anpassungsfähigkeit als ausreichend robust gegenüber Störungen und waren - mit Abstrichen - auch in der Lage, die benötigten Informationen zu beschaffen. Die Abstriche betreffen primär die Lohnvergleiche, die nur in rudimentärer Weise erfolgen konnten. Die Untersuchung hat sich äusserst flexibel an die Rahmenbedingungen angepasst und macht diesen Anpassungsprozess im Schlussbericht auch transparent. Das Kriterium wird mehrheitlich erfüllt.

5.1.3.2.2. Politische Überlebensfähigkeit

Wie bereits erwähnt, bewegen sich die Evaluatoren in einem Politikfeld, das sich heute durch eine relativ schwache Politisierung auszeichnet. Durch die abnehmende Bedeutung des Phänomens der (konventionellen) Heimarbeit scheint auch der dazugehörige Politikbereich an Aktualität und Brisanz eingebüsst zu haben. Dieser Umstand lässt erwarten, dass die Resultate der Evaluation über eine vergleichsweise hohe politische Überlebens-

fähigkeit verfügen. Harte Widerstände sind eigentlich nicht zu erwarten. In dieselbe Richtung wirkt auch der Umstand, dass die Evaluation mit den vier Szenarien in der Darstellung der Untersuchungsergebnisse sehr offen bleibt. Die präsentierten Handlungsalternativen offerieren ein sehr weites Feld möglicher Strategien zur zukünftigen Ausgestaltung des Heimarbeitnehmerschutzes. Auch dies trägt dazu bei, dass die Studie nicht auf pauschale Ablehnung stösst. Die vier Szenarien sind im Detail etwas zu wenig präzise in der Argumentation (vgl. Kapitel 5.1.3.4.10.). Im Gegensatz zur Endversion des Schlussberichts (SB) enthält der Entwurf zum Schlussbericht (ESB) noch eine dezidierte Stellungnahme zur zukünftigen Entwicklung. Aufgrund der diesbezüglichen Differenzen zwischen dem Schlussberichtsentswurf und dem Schlussbericht drängt sich an dieser Stelle eine Zusatzbemerkung auf. Es besteht aufgrund der uns zugänglichen Unterlagen die Vermutung, dass der Schlussbericht aufgrund (politisch-motivierter) Vorschläge der Expertengruppe im Bereich der Folgerungen offener ausgestaltet wurde. Weiterhin mag sich hier auch eine Diskrepanz der Positionen innerhalb des Evaluatorenteams artikulieren.

Insgesamt betrachtet ergibt sich bezüglich der politischen Überlebensfähigkeit aber durchaus ein positives Bild. Der Evaluation ist es gelungen, die möglichen Handlungsalternativen aufzuzeigen. Unseres Erachtens hätten die verschiedenen Szenarien hinsichtlich ihrer Grundannahmen und Konsequenzen noch etwas detaillierter und präziser ausgearbeitet werden können. Das Kriterium wird mehrheitlich erfüllt.

5.1.3.2.3. Kostenwirksamkeit

Die Evaluation von André Schmid und Frohmut Gerheuser hat Kosten im Umfang von rund 136'000.- sFr. verursacht. Es handelt sich somit um eine Evaluation der mittleren Preisklasse, wenn man sie mit den anderen Studien unserer Stichprobe vergleicht.

Das von den Evaluatoren gewählte Vorgehen ist im Vergleich zu anderen Untersuchungen nicht besonders aufwendig. Der Untersuchungsgegenstand bildete in der Schweiz noch kaum Gegenstand von Evaluationen. Daher konnten die Evaluatoren nur spärlich von Vorarbeiten profitieren, wie dies bei öfters evaluierten Gegenständen ja zumeist der Fall ist.

Den Wert der produzierten Informationen einzuschätzen, stellt ein schwieriges Unterfangen dar. Aufgrund der Reaktionen der Praxis, insbesondere was die Arbeitsgruppe zur Verbesserung des Schutzes der Heimarbeiterinnen und Heimarbeiter betrifft (siehe EVD 1995), hat die Evaluation zu nützlichen Resultaten geführt. Im Gegensatz dazu sind wir der Ansicht, dass es möglich gewesen wäre, einerseits die Szenarien präzi-

ser auszuarbeiten und andererseits detaillierter auf die mit diesen Handlungsalternativen verbundenen Konsequenzen einzutreten. Diese Arbeit ist im Anschluss an die Evaluation von der genannten Arbeitsgruppe (unter den für die Verwaltung relevanten Restriktionen der politischen Machbarkeit) geleistet worden (EVD 1995: 11 und 13-16). Für den interessierten Leser des Evaluationsberichts wirkt die Situation unbefriedigend, wenn er mit Szenarien konfrontiert wird, eine Bewertung dieser Szenarien fehlt und auch keine weiterreichenden Überlegungen zu den Vor- und Nachteilen der Optionen angeführt wird (siehe dazu Kapitel 5.1.3.4.10.).

Es stellt sich nun die Frage, wie sich das Verhältnis zwischen den Kosten und der Wirksamkeit bemessen lässt. Urs Paul Engeler kommt im bereits zitierten Zeitungsartikel zu folgendem Schluss: "Ein Nulleffekt zum Preis von 136'264 Franken." (Cash vom 15. 3. 1994) Wir können uns diesem vernichtenden Urteil nicht anschliessen. Die Evaluation hält viele wissenswerte Informationen bereit. In den Schlussfolgerungen fehlt zwar unseres Erachtens eine systematische Aufarbeitung der mit den Szenarien verbundenen Konsequenzen, die Szenarien und weitere Resultate der Evaluation stellen aber gewiss einen bestimmten Wert dar. In Anbetracht der verursachten Kosten ist die Kostenwirksamkeit gegeben. In welchem Ausmass dies der Fall ist, bedarf weiterer Abklärungen. Aufgrund unserer Analyse ist eine ausreichend fundierte Bewertung nicht möglich.

5.1.3.3. Korrektheit

5.1.3.3.1. Formale Verpflichtung

Die Evaluationsstudie ist im Auftrag des Schweizerischen Nationalfonds (SNF) erstellt worden. Deshalb wird anstelle des sonst üblichen Evaluationskontraktes das Forschungsgesuch beigezogen.

Finanzen: Das Projekt ist zu Beginn mit finanziellen Mitteln im Umfang von 120'000.- sFr. ausgestattet worden. Da der erste Entwurf des Schlussberichts den Ansprüchen der Expertengruppe nicht genügen konnte, wurde eine Überarbeitung des Schlussberichts notwendig. Dafür wurde ein Zusatzkredit von 16'264.- sFr. (entspricht 13,6%) gesprochen. Gemäss der finanziellen Berichterstattung durch die Evaluatoren wurden aber auch so die Aufwendungen nicht gedeckt. Es verbleibt eine Kostenüberschreitung von 7'660.- sFr. Diese wurde unseres Wissens von den Forschenden selbst getragen.

Zeit: Wie bereits erwähnt, ist die Laufzeit des Projekts mehrmals verlängert worden. Formell belief sich der Zeitbedarf von Projektbeginn bis Projektende auf 30 Monate. Im For-

schungsgesuch ist eine Projektdauer von 14 bis 18 Monaten beantragt worden. Der Forschungsrat hatte eine Projektdauer von 18 Monaten bewilligt. Faktisch standen den Evaluatoren für die Projektarbeit insgesamt 28 Monate zur Verfügung. Diese Verzögerungen sind als massgeblich zu bezeichnen.

Personen: Die beiden hauptsächlich mit der Erarbeitung der Evaluation beschäftigten Personen sind André Schmid und Frohmut Gerheuser. André Schmid kann als Initiator der Evaluation bezeichnet werden, war er doch schon an der ersten Projektskizze beteiligt. Er figurierte auch als Projektleiter und Hauptgesuchssteller. Frohmut Gerheuser stiess erst später - aber noch vor der Einreichung des definitiven Forschungsgesuchs - zum Forschungsteam. Anton Bellwald und Daniela Perrig-Gmünder waren nur marginal am Projekt beteiligt. Die Überarbeitung des Schlussberichts wurde von Frohmut Gerheuser im Alleingang erledigt. Das Projekt wurde durch eine aussenstehende Gruppe von Experten begleitet. Unseres Wissens hat sich diese nur gerade zwei Mal zu einer Sitzung getroffen (zur Zusammensetzung der Begleitgruppe vgl. Kapitel 5.1.3.1.1.). Die Evaluation ist den personellen Verpflichtungen ausreichend nachgekommen.

Inhalt: Hier erscheinen uns drei Punkte erwähnenswert. Zum einen betrifft dies den Lohnvergleich, der, wie wir bereits gesehen haben (siehe oben, Kapitel 5.1.3.2.1.), nur rudimentär erfolgt ist. Im Forschungsgesuch kam diesem noch eine zentrale Bedeutung zu. Die Autoren begründen diese Abweichung explizit im Schlussbericht (SB: 4-5 und ausführlicher 126-136). Der zweite Punkt, der uns hier wichtig erscheint, ist der Umstand, dass die Tätigkeit der Zentralstelle für Heimarbeit, entgegen der Ankündigung im Forschungsgesuch, nur ganz am Rande der Untersuchung Berücksichtigung findet (so etwa SB: 69 und 90). Die im Forschungsgesuch gestellten Fragen zu dieser Institution bleiben im Schlussbericht praktisch nicht behandelt. Drittens wird die praktische Zielsetzung des Projekts, Hinweise für eine zukünftige Handhabung des Lohngleichheitspostulates zu erhalten (SB: 3), nicht ganz erreicht. Die praktischen Schlüsse (SB: 109-116) enthalten keine Wertung der dargelegten Handlungsoptionen. Für den Leser wird zu wenig klar, mit welchen Vor- und Nachteilen die vorgeschlagenen Szenarien verbunden sind (vgl. dazu Kapitel 5.1.3.4.10). Im Gegensatz dazu enthält die Evaluation aber auch Arbeiten, die im Forschungsgesuch nicht angekündigt waren. In erster Linie betrifft dies die Ausweitungen im Bereich der Interviews. Es wurden auch mit Heimarbeitgebern Interviews durchgeführt, obwohl dies gemäss dem Forschungsgesuch nicht vorgesehen war. Die Zahl der Gespräche mit Heimarbeitnehmern wurde im Vergleich zum Forschungsgesuch massgeblich erhöht. Damit wurde eine Vertiefung und breitere empirische Abstützung der Evaluation erreicht.

Methodik/Vorgehen: Wir sind bereits in einem vorangehenden Kapitel (siehe Kapitel 5.1.3.2.1.) darauf zu sprechen gekommen, dass das Untersuchungsdesign im Verlauf des Evaluationsprozesses in mehrfacher Hinsicht geändert wurde. Auf diese Änderungen wird im Schlussbericht deutlich hingewiesen. Zentral erscheint uns dabei der bereits oben unter den inhaltlichen Punkten erwähnte Umstand, dass die Evaluatoren auf einen Lohnvergleich verzichtet haben und sich auf das Wirkungsgefüge des Programms konzentrierten. Dies hatte natürlich auch Implikationen für das Vorgehen der Evaluatoren. Zusammenfassend können wir festhalten, dass die Evaluation in vielfacher Hinsicht nicht dem Vorgehen, wie es im Forschungsgesuch festgehalten wurde, gefolgt ist. Diese Abweichungen sind jedoch im Schlussbericht deutlich gemacht und ausführlich begründet worden. Zudem steht dem Verzicht auf gewisse Analyseteile eine Ausweitung in anderen Bereichen gegenüber.

Gesamthaft gesehen handelt es sich bei dieser Studie um eine Evaluation, die im Verlauf des Evaluationsprozesses einem recht grossen Wandel unterworfen war. Verschiedene Probleme sind erst im Verlauf der Projektbearbeitung deutlich geworden. Die Evaluatoren haben versucht mit diesen schwierigen Bedingungen umzugehen, was in vielen Fällen auch gelungen ist. Aufgrund der angeführten Überlegungen kommen wir zum Schluss, dass die Evaluation das Kriterium teilweise erfüllt.

5.1.3.3.2. Deklaration von Interessenkonflikten

Zentral für diese Evaluation ist der Interessenkonflikt zwischen einer liberalen marktwirtschaftlichen Haltung, die für eine vermehrte Deregulierung eintritt, einerseits und einer sozialen politischen Perspektive, die primär auf eine Optimierung des Schutzes der Heimarbeitnehmer abzielt. Ausgerichtet auf den Untersuchungsgegenstand geht es also - vereinfacht gesagt - um die Frage, ob das Lohngleichheitspostulat trotz veränderten Rahmenbedingungen nach wie vor einem Bedarf entspricht. Diese Fragen werden von den Autoren im Bericht zwar angeschnitten. Im Kapitel mit den praktischen Schlüssen (Kapitel 7, SB: 109-116) fehlt jedoch eine Würdigung dieser Frage. Die Szenariotechnik gestattet den Evaluatoren, auf eine Bewertung der Handlungsalternativen zu verzichten. Unseres Erachtens wäre es wertvoll gewesen, wenn man sich aufgrund der empirischen Untersuchungsergebnisse auf eine Bewertung der Szenarien hätte einigen können. Darauf haben die Autoren jedoch verzichtet (vgl. auch Kapitel 5.1.3.4.10.). Eine weitere Möglichkeit hätte darin bestanden, die Konsequenzen der verschiedenen Szenarien aufzuzeigen, indem die je spezifischen Vor- und Nachteile der vier Szenarien dargelegt worden wären. Dies hätte den Adressaten die Möglichkeit eröffnet, sich aufgrund individueller Präferenzen für eines der Szenarien zu entscheiden. Die Evaluatoren verzichteten auch

auf diese Alternative. Die durch den Bundesrat eingesetzte verwaltungsginterne Arbeitsgruppe musste diese Arbeit selber leisten (EVD 1995). Das Kriterium wird deshalb mehrheitlich nicht erfüllt.

5.1.3.3.3. Vollständige und aufrichtige Offenlegung

Der Evaluation ist bezüglich der Offenlegung des Vorgehens ein gutes Zeugnis auszustellen. Sie behandelt detailliert die Wandlungen, welche das Untersuchungsdesign über den Projektverlauf hinweg durchgemacht hat. Die Autoren legen die Veränderungen im Untersuchungsplan offen dar und diskutieren die problematischeren Aspekte ihres Vorgehens, wie auch die möglichen alternativen Vorgehensweisen. Offen bleibt aufgrund des Schlussberichts die Frage, weshalb kaum Aussagen zur Zentralstelle für Heimarbeit im Bericht enthalten sind (SB: 69 und 90). Trotz dieser Einschränkung haben sich die Autoren sichtlich darum bemüht, dem Leser die gewählte Vorgehensweise plausibel darzulegen und zu begründen. Wir betrachten deshalb dieses Kriterium, zwar nicht als vollumfänglich, aber doch als mehrheitlich erfüllt.

5.1.3.3.4. Herstellung von Öffentlichkeit

Der Schlussbericht der Evaluation ist kurz nach Projektende in einer der Allgemeinheit zugänglichen Form publiziert worden (SB). Die Forscher haben sich bei der Abfassung des Schlussberichts darum bemüht, auch Lesern, denen die behandelte Materie eher fremd ist, den Zugang zum Untersuchungsgegenstand und zu den Ergebnissen zu ermöglichen. Die inhaltlichen Resultate der Studie finden sich zudem in einer ebenfalls vom NFP 27 publizierten Zusammenfassung (Reihe 'Kurzberichte' (KB)). Weitere Möglichkeiten zur Herstellung von Öffentlichkeit (Veranstaltungen, Pressekonferenzen, Artikel etc.) sind unseres Wissens nicht genutzt worden. Trotzdem kommen wir aber zum Schluss, dass die Evaluation dem Kriterium in ausreichendem Masse nachgekommen ist.

5.1.3.3.5. Schutz individueller Menschenrechte

Uns ist im Zusammenhang mit dieser Evaluation keine Verletzung individueller Menschenrechte bekannt. Das Kriterium kann deshalb als erfüllt betrachtet werden.

5.1.3.3.6. Menschliche Interaktionen

Für dieses Kriterium von Interesse ist vor allem das Verhalten der Evaluatoren gegenüber den Interviewpartnern. Besonders kritisch zu betrachten ist dieser Punkt hier deshalb,

weil ein Teil der Interviewpartner - die Heimarbeiternehmer, aber teilweise wohl auch die übrigen Interviewpartner - im Umgang mit Interviews wenig bis gar keine Erfahrungen aufweisen. Zur Einschätzung des Verhaltens der Evaluatoren in den Gesprächen fehlt uns hier der nötige Einblick. Es fehlen jedoch jegliche Anzeichen, dass diesbezüglich gegen die Würde und den Wert der befragten Personen verstossen worden wäre. Gerade ein Hinweis findet sich im Schlussbericht, der uns etwas seltsam erschienen ist. Es handelt sich dabei um die folgende Anmerkung zur Gesprächsbereitschaft der Heimarbeiternehmer: "Nur zwei angefragte Heimarbeiterinnen, *darunter eine Ausländerin*, wollten keine Auskunft geben." (SB: 8, Hervorhebung TW) Der Hinweis auf die Staatsbürgerschaft einer Person steht isoliert im Raum. Es wird nicht ersichtlich, warum dies hier erwähnenswert sein soll, da die Staatsangehörigkeit der Arbeitnehmer ansonsten im Schlussbericht nicht thematisiert wird. Bezüglich der Gesprächsbereitschaft der Heimarbeiternehmer stellt sich die Frage, ob es wirklich nur zu zwei Verweigerungen kam. In Anbetracht der Gesprächsthematik und der (möglichen) Befürchtungen der Gesprächspartner scheint uns diese Zahl doch sehr gering ausgefallen zu sein. In die gleiche Richtung weist auch ein Hinweis im Methodikteil des Schlussberichts. Dort heisst es, dass es *im Kanton St. Gallen* lediglich zu zwei Verweigerungen kam (SB: 139). Es erscheint uns unwahrscheinlich, dass in den anderen Untersuchungsräumen gar keine Verweigerungen festzustellen waren.

Aufgrund des unzureichenden Informationsstandes müssen wir hier auf eine Bewertung verzichten. Dies würde weitergehende Abklärungen bedingen, die den Rahmen der vorliegenden Meta-Evaluation gesprengt hätten.

5.1.3.3.7. Ausgewogene Berichterstattung

Wir kommen bei der Diskussion dieses Kriteriums wiederum auf den Punkt zurück, dass sich die Evaluatoren in ihren Schlussfolgerungen nicht auf einen Standpunkt festlegen wollen oder können. Aufgrund der Darlegung der Stärken und Schwächen des Evaluandums sollte eigentlich ein Entscheid, ob das Lohngleichheitspostulat des Heimarbeitsgesetzes beibehalten werden soll, möglich werden. Die Evaluatoren legen sich jedoch in den praktischen Schlüssen auf keine der möglichen Optionen fest, sondern zeigen diese lediglich auf. Die Evaluation scheint keine Ergebnisse produziert zu haben, die insgesamt für eine der erwähnten Optionen sprechen würden. Auf der Ebene einzelner Dimensionen (wie Schutz der Heimarbeiternehmer, Vollzugsaufwand, Kosten und Risiken einer Gesetzesänderung usw.) hingegen, wäre unseres Erachtens eine Einstufung der Szenarien möglich gewesen. Es werden zwar in der Evaluation sowohl die Stärken wie auch die Schwächen des Evaluandums benannt. Eine Gewichtung der verschiedenen

Anhaltspunkte bleibt jedoch aus. Die Konsequenzen aus den vier Szenarien werden nicht ausreichend dargelegt. Hinweise zu den Vor- und Nachteilen der Szenarien bleiben weitgehend aus (vgl. EVD 1995: 11 und 13-14). Soweit uns bekannt, erfolgte die Präsentation vollständig und auch fair. Die Evaluatoren verzichteten jedoch darauf, Empfehlungen darüber abzugeben, ob eher die Stärken des Programms zu verstärken sind oder ob man sich besser auf die Eliminierung der Problemfelder - im Extremfall würde dies die Abschaffung des Lohngleichheitspostulates bedeuten - konzentrieren sollte.

Trotz dieser Problematik erfüllt die Evaluation das Kriterium, weil sie in der Präsentation der Stärken und Schwächen (soweit wie möglich) vollständig und fair ist.

5.1.3.3.8. Finanzielle Verantwortlichkeit

Soweit uns bekannt ist, erfolgte die Zuweisung und die Ausgabe von Ressourcen seitens des Evaluationsteams in verantwortlicher Weise. Etwas überrascht hat uns der Umstand, dass die Mitwirkung von Anton Bellwald (Konzeption) und Daniela Perrig-Gmünder (rechtliche Fragen) im Schlussbericht keine Erwähnung findet. Auch die finanziellen Zwischen- und Schlussberichte enthalten dazu keine näheren Angaben. Bezüglich der Form der durchgeführten Interviews stellt sich die Frage, ob durch einen vermehrten Einsatz von telephonischen Interviews die Befragungen nicht effizienter hätten gestaltet werden können (siehe SB: 139). Um dies gründlich beurteilen zu können, bedarf es jedoch weiterer Informationen. Unter Vorbehalt der knappen Informationslage kann dieses Kriterium als erfüllt betrachtet werden.

5.1.3.4. Genauigkeit

5.1.3.4.1. Identifikation des Gegenstandes

Die Evaluatoren befassen sich mit den Auswirkungen des Lohngleichheitspostulates im Heimarbeitsgesetz aus dem Jahre 1981 (HArG, SR 822.31). Der erste Absatz des vierten Artikels dieses Gesetzes bestimmt, dass sich der Lohn für Heimarbeit grundsätzlich "nach den im eigenen Betrieb für gleichwertige Arbeit geltenden Ansätzen" (HArG, Art. 4 Abs. 1, SR 822.31) zu richten habe. Die Evaluatoren definieren das Evaluandum klar (siehe SB: 1). Durch die Problematik der Gleichwertigkeit der Arbeit zum einen und durch die unterschiedlichen Rahmenbedingungen von Heim- bzw. Betriebsarbeit zum anderen, wird die Durchführung eines direkten Vergleiches erheblich erschwert. Die Evaluatoren entscheiden sich deshalb dafür, den direkten Lohnvergleich nur am Rande zu behandeln und das Hauptgewicht der Studie auf eine indirekte Erfassung der Auswirkungen des

oben genannten Postulates zu legen. Dies versuchen die Autoren zu erreichen, indem sie Einsichten in das Wirkungsgefüge des Lohngleichheitspostulates zu gewinnen versuchen, um dann daraus auf die Auswirkungen der Regelung zu schliessen.

Mit diesem Forscherentscheid entfernen sich die Evaluatoren von der ursprünglichen Zielsetzung des Projekts (vgl. dazu oben Kapitel 5.1.3.2.1. und 5.1.3.3.1.). Da die Evaluatoren detailliert auf die Beweggründe für diese Abkehr von der projektierten Untersuchungsanlage eingehen und den Wandel damit transparent und plausibel werden lassen, kann der fehlende Lohnvergleich nicht als eigentliche Lücke bezeichnet werden. Die Frage nach der Lohngleichheit bleibt aber trotzdem unbeantwortet. Dieser Umstand reduziert die Aussagekraft der Studie.

Die Evaluation untersucht das Evaluandum aber gründlich, wodurch eine Einschätzung des Gegenstandes möglich wird. Deshalb erfüllt die Evaluation das Kriterium vollständig.

5.1.3.4.2. Kontextanalyse

Um untersuchen zu können, ob der Kontext von den Evaluatoren ausreichend untersucht worden ist, muss zuerst bestimmt werden, was als Kontext der Untersuchung zu bezeichnen ist. Zum Kontext gehörig definieren wir einerseits das wirtschaftliche und andererseits das rechtliche Umfeld, in dem das Lohngleichheitspostulat lokalisiert werden kann.

Der wirtschaftliche Kontext der Heimarbeit wird im Schlussbericht ziemlich ausführlich dargelegt. Dominant ist dabei sicherlich die abnehmende Bedeutung herkömmlicher Heimarbeit in der Schweiz. Das rechtliche Umfeld des Lohngleichheitspostulates wird zwar im Anhang, der ein Verzeichnis der wichtigsten gesetzlichen Regelungen zum Heimarbeitslohn umfasst (siehe Kapitel 8, SB: 119-123), dargelegt. Teilweise findet jedoch im Schlussbericht das rechtliche Umfeld zu wenig Beachtung. Am deutlichsten tritt dies in den Schlussfolgerungen hervor. Dort wird im vierten Szenario für eine Streichung des Heimarbeitsgesetzes (HArG, SR 822.31) Stellung genommen. Dabei versäumen es die Autoren aber erstens, auf jene Regelungen im Heimarbeitsgesetz einzutreten, die damit zusammen mit dem Lohngleichheitspostulat auch gleich gestrichen werden würden. Zweitens geht die Evaluation von der Annahme aus, dass das Arbeitsgesetz (ArG, SR 822.11) einen ausreichenden Schutz der Heimarbeitsnehmer gewährleisten würde. Das ArG ist aber nicht auf den Bereich der Heimarbeit anwendbar (ArG, Art. 3, lit. f). Drittens wird an demselben Ort darauf verwiesen, dass die Lohngleichheit auch durch den Verfassungsartikel zur Gleichstellung von Mann und Frau (Art. 4 BV) geschützt sei. Zwar handelt sich beim Lohngleichheitsprinzip nach Art. 4 Abs. 2 BV um eine direkt anwendbare Norm. Der Vollzug dieser Rechtsnorm bereitet jedoch grosse Probleme. Das

im Verfassungsartikel zur Präzisierung geforderte Gesetz ist heute, zwei Jahre nach Abschluss der Evaluation, noch nicht in Kraft. Aus diesen Gründen kann das Lohngleichheitspostulat im HARG zum Zeitpunkt der Fertigstellung des Schlussberichts noch nicht als obsolet bezeichnet werden, da ja die Formulierung des Gleichstellungsgesetzes noch nicht in definitiver Form vorlag und zudem der Zeitpunkt des Inkrafttretens noch offen war (zum aktuellen Stand vgl. Kapitel 5.1.3.1.4.). Weiterhin ist deutlich darauf hinzuweisen, dass Art. 4 BV gleichen Lohn für gleichwertige Arbeit *zwischen* den beiden Geschlechtern vorschreibt. Das heisst, dass unter Arbeitnehmern desselben Geschlechts grundsätzlich kein Anspruch auf Lohngleichheit besteht (nach Müller 1987: 61, Rz. 144). Umgesetzt auf die hier zur Debatte stehende Situation würde dies bedeuten, dass dieser Anspruch nur geltend gemacht werden könnte, wenn der für den Lohnvergleich herangezogene Betriebsarbeitnehmer einem anderen Geschlecht zugehört als der Heimarbeiter.

Im Zusammenhang mit dem rechtlichen Kontext ist auf einen weiteren Aspekt hinzuweisen. Der in der Diskussion um die Einflüsse eines Gesamtarbeitsvertrages (GAV) auf die Lohnsituation der Heimarbeiter, wie sie sich im Kapitel zur Walliser Uhrenindustrie findet (SB: 43-48), angeführte Vergleich (SB: 47-48) von Unternehmen, die dem GAV unterstellt sind, mit solchen, die dies nicht sind, entbehrt einer fundierteren empirischen Grundlage. Die für die Analyse beigezogenen Unternehmen sind nämlich ausnahmslos nicht dem GAV unterstellt (vgl. SB: 44, Anm. 26).

Gesamthaft gesehen erfüllt die Evaluation das Kriterium nur mit Einschränkungen.

5.1.3.4.3. Beschreibung von Zielen und Vorgehen

Bezüglich dieses Kriteriums schneidet die Evaluation gut ab. Es wird ausführlich und auf nachvollziehbare Weise dargelegt, wie sich die Evaluation im Forschungsprozess gewandelt hat. Die Evaluatoren versäumen es auch nicht, auf die diversen Schwierigkeiten und Widrigkeiten bei der Erarbeitung der Evaluation hinzuweisen. Auch eher technische Aspekte, wie beispielsweise die Art und Weise der Durchführung der Interviews oder die Auswahl der Gesprächspartner werden ausreichend beschrieben. Die hier relevanten Ausführungen finden sich im Schlussbericht zum einen in einem kurzen einleitenden Kapitel zur Vorgehensweise (SB: 5-9) und andererseits in einer Methodendiskussion im Anhang des Berichts (SB: 125-148). Zwischen dem Kapitel 5 ("Chancen und Bedingungen zur Verwirklichung des Lohngleichheitspostulates", SB: 71-102) und der Methodendiskussion im Anhang (Kapitel 9) kommt es stellenweise zu Überschneidungen (etwa SB: 71 und 143), die jedoch inhaltlich durchaus zu rechtfertigen sind. Bezüglich dem Vorgehen

bleibt eine Frage offen. Da die Evaluation ja im empirischen Teil massgeblich aus vier Fallstudien zu verschiedenen Branchen in verschiedenen Kantonen besteht, kommt sofort die Frage auf, weshalb nicht ein systematischer Fallvergleich eingesetzt wurde. Auf diese Frage bleiben die Evaluatoren eine Antwort schuldig. Dies ist jedoch nicht als allzu schwerwiegend zu gewichten, da praktisch bei jeder Evaluation Fragen dieser Art unbeantwortet bleiben müssen.

Ein kritischer Punkt ist an dieser Stelle trotz dem generell sehr positiven Eindruck anzumerken. Dies betrifft die Beschreibung der vier Fallstudien. Hier finden sich im Schlussbericht teilweise widersprüchliche Angaben. So etwa bei der Fallstudie im Bereich des Maschinen- und Apparatebaus. An einer Stelle des Schlussberichts wird festgehalten, dass dazu dieser Wirtschaftszweig im Kanton St. Gallen untersucht wurde (SB: 5). In einer Übersichtstabelle wird neben dem Kanton St. Gallen auch der Kanton Uri erwähnt (Tabelle 2.1., SB: 8). In der empirischen Fallstudie sodann wird von der Ostschweiz gesprochen (siehe Kapitel 4.5., SB: 57-62). In der Methodendiskussion im Anhang der Studie wird angeführt, dass die Maschinen- und Apparatebaubranche anhand von Gesprächen in den Kantonen St. Gallen und Wallis untersucht worden ist (SB: 134). Ein zweites Beispiel dieser Art bilden die Angaben zur Zahl der durchgeführten Gespräche. Auch hier finden sich widersprüchliche Angaben im Schlussbericht. Dies betrifft etwa die Zahl der befragten Heimarbeitgeber, die an einer Stelle mit 14 (SB: 8) und an einer anderen mit 13 (SB: 135) angegeben wird. Es ist zu hoffen, dass diese Inkonsistenzen lediglich auf redaktionelle Unsorgfältigkeiten zurückzuführen sind.

Dieser Aspekt verändert jedoch die generell positive Einschätzung der Evaluationsstudie in bezug auf dieses Kriterium nicht erheblich. Das Kriterium wird grösstenteils erfüllt.

5.1.3.4.4. Verlässliche Informationsquellen

Die verschiedenen, von den Evaluatoren beigezogenen Informationsquellen, sind einer differenzierten Analyse zu unterziehen.

Die Durchführung der *Befragungen* wird im Schlussbericht relativ genau beschrieben. Die Auswahlverfahren, die zur Bestimmung der Gesprächspartner zur Anwendung gelangten, sind deutlich gemacht worden. Die Umstände, wie die Gespräche zustande gekommen sind, werden dargelegt. Die Zahl der Interviews ist ausreichend hoch. Es sind insgesamt 55 bzw. 54 längere Gespräche geführt worden (SB: 8 und 135). Die Auswahl der Gesprächspartner erscheint uns - soweit dies anhand der Dokumentenanalyse beurteilt werden kann - sachdienlich erfolgt zu sein. Die Evaluatoren haben sich darum bemüht, dass die Auswahl der Interviewpartner nicht allzu einseitig erfolgte. Trotz der getroffenen

Gegenmassnahmen ist jedoch nicht ganz auszuschliessen, dass die Gesprächspartner tendenziell eher eine positive Auswahl darstellen.

Die von den Evaluatoren hinzugezogenen *statistischen Materialien* stammen aus unterschiedlichen Quellen, die bereits weiter oben dargelegt worden sind (siehe Kapitel 5.1.3.1.3.). Aus dem Schlussbericht geht deutlich hervor, dass das Datenmaterial nur ausnahmsweise über eine ausreichende Qualität verfügt. Die Evaluatoren verzichteten selbst auf eine systematische Primärerhebung quantitativer Daten. Dieser Entscheid wird im Schlussbericht ausführlich begründet (siehe SB: 139-141). Das quantitative Datenmaterial ist von den Evaluatoren nur am Rande eingesetzt und eher zurückhaltend interpretiert worden. Dessen Einsatz führt zudem zu einer Fülle von Schwierigkeiten, welche die Interpretation behindern. Exemplarisch sei hier auf die diversen Definitionsprobleme hingewiesen. Die Heimarbeit etwa wird unterschiedlich abgegrenzt. Die Definitionen der Wirtschaftsbranchen (siehe Tabelle 3.2., SB: 17) werfen Probleme auf, da sie nicht mit den von den Evaluatoren verwendeten Festlegungen der Untersuchungsbranchen übereinstimmen. So finden sich in dieser Tabelle keine Angaben zur Heimarbeit für das Militär. Die nach Kantonen aufgeschlüsselten Angaben in einer anderen Tabelle (Tabelle 3.3., SB: 18) beziehen sich scheinbar nicht nur auf die herkömmliche Heimarbeit im sekundären Sektor, sondern umfassen auch die neuen Heimarbeitsformen im tertiären Sektor. Die zur Verfügung stehenden Angaben zu den Vergleichslöhnen aus der Lohnerhebung des Bundesamtes für Industrie, Gewerbe und Arbeit (BIGA) sind für die Zwecke der Evaluation regional und arbeitsspezifisch zuwenig detailliert. Angesichts der schwierigen Quellenlage haben sich die Autoren der Studie klugerweise dafür entschieden, den direkten Lohnvergleich nur am Rande der Untersuchung durchzuführen und das Hauptgewicht der Studie auf die qualitativen Analysen zu legen. Verschiedentlich finden sich im Schlussbericht quantitative Angaben, zu denen ein Quellenhinweis fehlt. So etwa die Angaben zur Bruttolohnsumme der Heimarbeitnehmer oder die geschlechterspezifische Strukturierung der Heimarbeitnehmerschaft (siehe SB: 16).

Die von den Evaluatoren verwendeten *Dokumente* waren mehrheitlich offizieller Natur. Namentlich sind dies die gesetzlichen Grundlagen, die Botschaften des Bundesrates und die Protokolle zu den Verhandlungen in den eidgenössischen Räten. Die Fundstellen dieser Texte sind im Schlussbericht teilweise nur mangelhaft angeführt worden. Die Ratsprotokolle beispielsweise finden sich nicht (wie dies die Autoren angeben) im Bundesblatt, sondern in den Amtlichen Bulletins (siehe SB: 153-154, vgl. dazu Kapitel 5.1.3.1.3.). Die Gesetzestexte werden ohne entsprechende Hinweise auf die Fundstellen in den Gesetzessammlungen angeführt (siehe SB: 152). Neben diesen formalen Unzulänglichkeiten sind die beigezogenen Dokumente ausreichend genau beschrieben worden.

Aufgrund dieser Überlegungen kommen wir zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt.

5.1.3.4.5. Glaubwürdigkeit und Übertragbarkeit

a) Glaubwürdigkeit

Die Autoren verzichten im Schlussbericht zur Evaluationsstudie darauf, ihre Rekonstruktionen der individuellen Konstruktionen der Akteure offenzulegen. Die Akteurkonstruktionen werden in kollektiver Weise dargelegt. Für jede der drei hier relevanten Akteurgruppen (Heimarbeitgeber, Heimarbeitnehmer und Akteure des Vollzugs) findet sich eine ausführliche Behandlung der Zielsetzungen, der Rahmenbedingungen und der möglichen Handlungsoptionen (siehe SB: 71-102). In diesem Rahmen wird auch der Versuch unternommen, den Stellenwert des Lohngleichheitspostulats für die Akteurgruppen zu erfassen. Diese ganze Diskussion baut auf einem Denkmodell zur Wirkungsweise des Lohngleichheitspostulates auf, das im Bericht auch in der Form einer Graphik veranschaulicht wird (siehe SB: 73). Entgegen dem Wirkungsmodell haben die Gerichte, die demnach neben den Arbeitsinspektoraten Vollzugsorgane sind, keine Aufnahme ins empirische Design der Studie gefunden. Unseres Wissens sind mit Vertretern der zuständigen Gerichte keine Befragungen durchgeführt worden, da die Zahl der hier relevanten gerichtlichen Auseinandersetzungen äusserst klein ist (siehe SB: 91-92, vgl. Kapitel 5.1.3.1.1.).

Inwiefern die dargelegten Rekonstruktionen der Evaluatoren korrekte Abbildungen der Konstruktionen der Akteure darstellen, ist von unserer Seite (und ohne die Durchführung von weiteren Interviews mit den Akteuren im Feld) kaum einzuschätzen. Dies hängt auch damit zusammen, dass die Evaluatoren darauf verzichten, die verwendeten Auswertungsverfahren im Schlussbericht darzustellen. Dem Leser fehlt aufgrund der Lektüre des Berichts die Einsicht in die interpretative Arbeitsweise der Evaluatoren. Es fehlt im Schlussbericht eine Verbindung der empirischen Daten mit den Aussagen im Schlussbericht. Zwar enthält der Schlussbericht auch direkte Bezüge auf Gesprächsinhalte, meist in Form von zusammenfassenden Paraphrasen der Äusserungen verschiedener Interviewpartner. Doch haben diese Hinweise mehrheitlich illustrativen Charakter. Vor allem aber ist es dem Leser nicht möglich zu erfassen, wie diese Textstellen produziert (das heisst selektiert, gebündelt, verwesentlich) worden sind. Aufgrund dieses nur mangelhaften Einblicks in den Entstehungsprozess der Aussagen fällt eine Beurteilung schwer (vgl. dazu auch unten, Kapitel 5.1.3.4.11). Die Rekonstruktionen erscheinen soweit aber

durchaus als plausibel. Auf alle Fälle sind wir auf keine Hinweise gestossen, die andere Schlüsse nahelegen würden.

An dieser Stelle drängt sich noch eine Randbemerkung auf: In der Diskussion um die Validität der Studienergebnisse kommen die Autoren aufgrund des in sich stimmigen Gesamtbildes zu folgendem Schluss: "So gibt es nur zwei Möglichkeiten: Die Aussagen sind alle in die gleiche Richtung verzerrt oder sie sind gültig und entsprechen der Realität. Das erste halten wir für unwahrscheinlich. Daher kommen wir zum Schluss, dass die Validität der Resultate hinreichend gewährleistet ist." (SB: 139) Dieser Argumentation können wir uns nicht anschliessen. Allein schon deshalb nicht, weil wir eine gleichgerichtete Verzerrung der Ergebnisse nicht für unwahrscheinlich halten. So besteht keine Gewähr dafür, dass die Auswahl der Gesprächspartner den Untersuchungsbereich vollständig abdeckt, dass sich also keine 'schwarzen Schafe' der Evaluation entzogen haben. Ebenso kann in den Gesprächen selbst eine Verzerrung wirken, wenn es von seiten der Interviewpartner strukturell zu beschönigenden Auskünften kommt (vgl. dazu auch Cook 1985: 48-57). Ebenso ist an dieser Stelle zu bedenken, dass durchaus divergierende Perspektiven - in oder zwischen den Akteurgruppen - bestehen können. Dies würde heissen, dass nicht eine gleichgerichtete Verzerrung bei allen Gesprächspartnern vorliegen muss, um die interne Validität der Ergebnisse in Frage zu stellen.

b) Übertragbarkeit

In der Evaluation werden vier Fallstudien erarbeitet, die sich auf unterschiedliche Wirtschaftsbranchen in ausgesuchten Schweizer Kantonen beziehen. Die Fallstudien verzichten auf eine ausführliche Beschreibung der kantonalen Rahmenbedingungen für die Wirkungsweise des Lohngleichheitspostulates. Die Fallstudien sind recht knapp gehalten und orientieren nur rudimentär über den Kontext in den ausgewählten Regionen und Branchen. Weiterhin bleibt offen, ob die Studie, aufgrund der zum Zeitpunkt der Evaluation herrschenden Rahmenbedingungen (etwa eine bestimmte Konjunkturlage) Resultate erzeugte, die nicht auf andere Zeitperioden übertragen werden können. Dadurch ist es dem Leser der Studie nicht möglich, zu beurteilen, ob die Schlussfolgerungen der Studie auch in anderen Kontexten relevant sein würden. Hier wäre seitens der Evaluatoren sicherlich mehr zu leisten gewesen. Die Übertragbarkeit der Evaluationsergebnisse kann durch den Leser nicht ausreichend eingeschätzt werden. Dafür fehlt es den Fallstudien an deskriptiver Substanz. Das Kriterium wird nicht erfüllt.

5.1.3.4.6. Verlässlichkeit

Die Evaluatoren bemühen sich im Schlussbericht erfolgreich darum, das gewählte Vorgehen transparent darzustellen und argumentativ zu begründen. Es bleiben zwar auch nach intensiver Lektüre des Schlussberichts noch einige Fragen bestehen, so etwa, in welcher Form die Interpretation der Interviews erfolgte oder weshalb nur ein rudimentärer Vergleich der vier implementierten Fallstudien durchgeführt wurde. Im übrigen ist es jedoch den Autoren gelungen, den Forschungsprozess von aussen erkennbar darzustellen. Dies ist besonders deshalb sehr wertvoll, da das Projekt verschiedentlich und zum Teil grundlegenden Änderungen unterworfen war. Das Kriterium wird durch die Evaluation erfüllt.

5.1.3.4.7. Systematische Datenüberprüfung

Für eine eingehendere Analyse der Evaluation 'Lohnleichheitspostulat des Heimarbeitsgesetzes' fehlen uns in bezug auf dieses Kriterium die nötigen Einsichten. Dafür müssten weitere Informationen (zum Beispiel in Form von vertiefenden Interviews) zugänglich gemacht werden. Da wir uns hier lediglich auf eine Dokumentenanalyse und auf eine kleine Zahl von Gesprächen stützen, können wir lediglich festhalten, dass keine Hinweise erkennbar sind, die darauf hindeuten würden, dass die Datenüberprüfung nicht in systematischer Weise erfolgt ist. Die Qualität des Schlussberichts deutet an, dass der Text systematisch überprüft wurde. Er enthält nur wenige formale und inhaltliche Unstimmigkeiten. Der Schlussbericht vermittelt jedoch keine expliziten Angaben darüber, ob und wie eine Datenüberprüfung stattfand. In gewisser Weise kann die Arbeit der Begleitgruppe und jene der Expertengruppe des NFP 27, die den Entwurf zum Schlussbericht (ESB) kommentiert haben, als Datenüberprüfung betrachtet werden.

Aufgrund der Qualität des Berichts gehen wir - obwohl uns keine expliziten Hinweise dazu bekannt sind - davon aus, dass die Evaluation einer systematischen Datenüberprüfung unterzogen wurde. Das Kriterium wird erfüllt.

5.1.3.4.8. Analyse quantitativer Informationen

Die Evaluation setzt ihr Hauptgewicht auf die Analyse qualitativer Informationen. Entgegen dem ursprünglichen Vorhaben, wie es im Forschungsgesuch festgehalten worden ist, werden quantitative Angaben nur punktuell beigezogen. Dies geschieht folgendermassen:

- Im historischen Überblick (SB: 12-22) dienen quantitative Angaben aus der Eidgenössischen Industriestatistik zur *Rekonstruktion der Entwicklung der Heimarbeit*. Diese Zahlen werden auch aufgliedert nach Branchen bzw. nach Kantonen vor-

- gelegt. Ergänzt werden die Angaben durch Informationen, die aus einer historischen Seminararbeit stammen (Iseli 1984). Das Datenmaterial ist teilweise nur sehr schwer zu interpretieren, da die statistische Erfassung hinsichtlich der Definition der Heimarbeit Strukturbrüche aufweist, weil nur zu ausgewählten Jahren Daten zur Verfügung standen oder weil die verschiedenen Quellen teilweise sich widersprechende Angaben enthalten. Die Evaluatoren vermeiden eine vertiefte Diskussion dieser Schwierigkeiten im Datenmaterial und wählen einen pragmatischen Analysezugang. Die Evaluation ist in ihren Interpretationen vorsichtig und zurückhaltend. Trotzdem treten dabei aber punktuell Schwierigkeiten auf. Exemplarisch sei hier auf die graphische Darstellung hingewiesen, welche die Entwicklung der industriell-gewerblichen Heimarbeit nach Branchen von 1905 bis 1988 darstellt (Abbildung 3.1., SB: 13). Erstens stammen die Daten aus zwei verschiedenen Quellen (Eidgenössische Betriebszählungen für die Jahre 1905, 1929, 1939 und 1955 (nach Iseli 1984) sowie Eidgenössische Industriestatistik für die Jahre 1982 und 1988). Deshalb ist die Vergleichbarkeit der Daten möglicherweise nicht gegeben. Da die Daten zweitens nur in unregelmässigen Zeitabschnitten vorliegen, ist es nicht zweckmässig, die Ordinatenachse der Darstellung in gleichmässige Abstände aufzuteilen, wie dies die Autoren im Schlussbericht tun. Drittens stellt sich die Frage, weshalb die Datenbasis nicht durch weitere Angaben verdichtet worden ist. Wir denken dabei etwa an die Eidgenössische Industriestatistik, die nicht nur für die beigezogenen Jahre 1982 und 1988 zur Verfügung steht (siehe etwa für die Jahre 1977 und 1978 die Hinweise in der Botschaft zur Revision des Heimarbeitsgesetzes (BBl 1980 II: 287-288 und 301-302)). Bei der Nachzeichnung der Entwicklung der Heimarbeit seit 1980 ist den Evaluatoren ein Fehler unterlaufen. Sie gehen davon aus, dass das revidierte Heimarbeitsgesetz 1982 bereits in Kraft war (siehe SB: 15). Dies ist jedoch falsch. Das revidierte Heimarbeitsgesetz ist per 1. April 1983 in Kraft getreten (Bundesratsbeschluss von 20. Dezember 1982; AS 1983: 113).
- Das Datenmaterial aus der Eidgenössischen Industriestatistik wird im Schlussbericht in einem anderen Zusammenhang nochmals aufgegriffen. Im Rahmen der *Fallstudien zu den Wirtschaftsbranchen* Uhrenindustrie, Textil- und Bekleidungsindustrie sowie Maschinen- und Apparatebau werden diese quantitativen Angaben beigezogen, um einerseits die Bedeutung der Heimarbeit in diesen Wirtschaftszweigen zu zeigen und um die Entwicklung von 1982 bis 1988 nachzuzeichnen (siehe Tabellen 4.3., 4.4. und 4.5., SB: 44, 50 bzw. 58). Da diese Daten nur für die gesamte Schweiz vorzuliegen scheinen, ergibt sich in der Interpretation das Problem, dass die dargelegten Aspekte für die in den Fallstudien behandelten Regionen nicht zwangsläufig zutreffen müssen. In der Darstellung der Entwicklung in der Textil- und Bekleidungsindustrie (vgl. Tabelle 4.4., SB: 50) ist den Evaluatoren zudem ein Fehler bei der Aufarbeitung des Daten-

materials unterlaufen. Während sich das Zahlenmaterial für das Jahr 1982 auf die drei Wirtschaftszweige 'Herstellung von Textilien', 'Herstellung von Bekleidungen, Wäsche' und 'Herstellung von Lederwaren und Schuhen' bezieht, umfassen die Angaben für das Jahr 1988 lediglich die ersten zwei Kategorien. Da im Jahre 1988 doch immerhin 546 Heimarbeiter mit der Herstellung von Lederwaren und Schuhen beschäftigt waren (Tabelle 3.2., SB: 17), führt dies doch zu einer stärkeren Verzerrung. Der Rückgang der Zahl der Heimarbeiter in diesen Branchen beträgt nicht 49, sondern lediglich 38 Prozent.

- Zur quantitativen Erfassung der *Heimarbeit für das Militär im Kanton Uri* konnten die Evaluatoren auf Datenmaterial des Kantonalen Amtes für Heimarbeit zurückgreifen. Dieses gibt detailliert Auskunft über die Einkommensverteilung bei den vom Amt beschäftigten Heimarbeitern und über die Kalkulation der Stundenansätze (siehe Tabellen 4.1. und 4.2., SB: 35 und 36). Auf die systematische Wiedergabe von weiterem Datenmaterial wird in der Studie verzichtet. Die Interpretation der Daten ist äusserst knapp gehalten. Es finden sich punktuell auch weitere quantitative Angaben im Text. Teilweise leisten diese jedoch nur sehr wenig Klärungshilfe. Beispielsweise wird erwähnt, dass im Kanton Uri ca. 315 Heimarbeiter tätig sind und dass davon etwa 60 Prozent (also 189 Personen) direkt durch das Amt für Heimarbeit beschäftigt werden (SB: 34). Vergleicht man diese Angaben mit jenen in einer Tabelle (Tabelle 4.1., SB: 35), fällt auf, dass dort eine Zahl von 254 durch das Amt für Heimarbeit beschäftigte Heimarbeiter vermerkt ist. Unseres Erachtens hätte bei dieser Fallstudie die vorliegende Datenbasis besser genutzt werden können. Zudem sind die Kommentare der Evaluatoren nicht in allen Teilen sehr klar ausgefallen.
- Zum *Lohnvergleich* werden die Angaben aus der Lohnerhebung des Bundesamtes für Industrie, Gewerbe und Arbeit (BIGA) beigezogen. Diese liegen in Form von Stundenlöhnen entweder differenziert nach Region *oder* nach Branche und Ausbildungsstand vor. Dies ergibt für den Vergleich mit Lohnzahlungen für Heimarbeit zweierlei Probleme. Erstens werden in der Heimarbeit üblicherweise nicht Stundenlöhne sondern Stücklöhne bezahlt. Zweitens müssten die Vergleichslöhne *gleichzeitig* arbeits- und auch regionenspezifisch differenziert sein, um einen stichhaltigen Vergleich zu ermöglichen. Weiterhin kommt hinzu, dass die Lohnerhebungen des BIGA zumindest potentiell dazu neigen, die Stundenlohnansätze zu überschätzen, weil sich tendenziell eher Arbeitgeber, die gute Löhne zahlen, an der Erhebung beteiligen. Aber auch die Angaben zu den Lohnzahlungen bei der Heimarbeit in den vier Branchen und den vier Regionen sind nicht unproblematisch. Wir haben bereits an anderer Stelle (vgl. Kapitel 5.1.3.1.3.) auf die Auswahl der Gesprächspartner hingewiesen. Die

Lohnangaben zur Heimarbeit stammen aus diesen Gesprächen und können nicht als repräsentativ betrachtet werden. Erschwerend wirkt der Umstand, dass die Bezahlung im Akkordlohnsystem, die Verschiedenartigkeit der vergebenen Heimarbeit und die unterschiedlichen Leistungsfähigkeiten der Heimarbeitnehmer dazu führen, dass in Stundenansätze umgerechnete Heimarbeitslöhne interpersonell und intertemporal zum Teil sehr hohe Varianzen aufweisen. Aus diesen Gründen beschränken sich die Autoren auf eine zurückhaltende Analyse und vermeiden weitergehende Interpretationen, die sich auf die problematische Datenbasis stützen. In jeder der vier Fallstudien wird auf den Lohnvergleich eingegangen (SB: 38, 45, 51, 52 und 59). In Anbetracht der schwierigen Rahmenbedingungen kann von einer angemessenen und systematischen Analyse der Lohnsituation gesprochen werden. Zur Erhöhung der Übersichtlichkeit hätten die entsprechenden Daten in Tabellen zusammengefasst werden können.

Gesamthaft gesehen sind wir der Ansicht, dass es den Evaluatoren gelungen ist, die schwergewichtig qualitativ ausgerichtete Evaluation in zweckmässiger Form mit quantitativen Informationen zu ergänzen. Das quantitative Datenmaterial erlaubt eine Verortung und Gewichtung des Untersuchungsgegenstandes. Teilweise erhält man den Eindruck, dass im quantitativen Bereich noch Optimierungspotential bestanden hätte. Punktuell sind Unstimmigkeiten festzustellen. Das Kriterium wird mehrheitlich erfüllt.

5.1.3.4.9. Analyse qualitativer Informationen

Die qualitativen Informationen, die in der Evaluation 'Lohnleichheitspostulat des Heimarbeitsgesetzes' ausgewertet werden, stammen einerseits aus den Befragungen und andererseits aus schriftlichen Dokumenten und Unterlagen. Diese beiden Typen sind hier getrennt zu betrachten.

- Die *Befragungen* erscheinen aufgrund der uns zugänglichen Informationen in zweckmässiger Form erfolgt zu sein. Die daraus entstandenen Interpretationen zu den Perspektiven der drei hauptsächlich relevanten Akteurgruppen (Heimarbeitnehmer, Heimarbeitgeber, Vollzugspersonen) wirken plausibel und sachgerecht (siehe dazu SB: 76-102). Etwas problematisch ist jedoch der Bezug zwischen den Interviews und den Aussagen im Schlussbericht (vgl. dazu Kapitel 5.1.3.4.5.). Hier fehlen Angaben, wie die Gespräche analysiert und wie diese in den Bericht eingegangen sind. Die Verbindung zwischen empirischem Material und Berichterstattung ist nicht genügend transparent gemacht worden. Aufgrund der Erhebungstechnik - es sind keine Tonbandprotokolle der Gespräche erstellt worden, sondern lediglich "thematisch geordnete Gesprächsprotokolle" (SB: 136), die aufgrund von handschriftlichen Notizen entstanden

- sind - ist diese Verbindung jedoch auch äusserst schwer herzustellen. Die Möglichkeit, direkte Zitate aus Interviews als Belege oder zur Illustration in den Text des Schlussberichts zu integrieren, entfällt weitgehend. Weiterhin fehlen im Schlussbericht die verwendeten Gesprächsleitfaden für die drei verschiedenen Gruppen von Gesprächspartnern (Arbeitgeber, Arbeitnehmer und Vollzugsbeamte). Ebenso vermissen wir im Schlussbericht genauere Angaben, nach welchen Verfahren die Auswertung der Interviews erfolgte. Hier besteht ein relativ hoher Klärungsbedarf. Eine eher produktorientierte Einschätzung kommt zum Schluss, dass die Analysen zu in sich stimmigen, plausiblen und nachvollziehbaren Rekonstruktionen führten. Orientiert man sich zur Beurteilung der Evaluation hingegen eher am wissenschaftlichen Produktionsprozess, vermisst man Nachvollziehbarkeit und Transparenz. Gleichzeitig muss hier aber auch betont werden, dass die Etablierung einer vollständigen Nachvollziehbarkeit einen deutlich höheren Aufwand mit sich gebracht hätte, wodurch das bestehende Projektvolumen markant überschritten worden wäre (vgl. dazu auch SB: 139).
- Zur Interpretation der beigezogenen *schriftlichen Dokumente*, also etwa die Gesetzestexte, die Parlamentsprotokolle, die amtlichen Botschaften und Berichte, die Fachliteratur oder andere Unterlagen, müssen je nach Art der Schriftstücke unterschiedliche Anmerkungen erfolgen. Bei der Fachliteratur ist auffällig, dass eine relativ lange Liste von Titeln im Literaturverzeichnis figuriert. Betrachtet man jedoch den Text des Schlussberichts, sind Literaturhinweise nur sehr spärlich angebracht worden. Ob und - wenn ja - wie Informationen aus der Fachliteratur in die Studie eingeflossen sind, kann daher kaum beurteilt werden. Weiterhin ist zu beobachten, dass die schriftlichen Dokumente vor allem für einzelne Kapitel beigezogen worden sind (so etwa Kapitel 3. 'Der gesetzliche Heimarbeitnehmerschutz im Spannungsfeld unterschiedlicher Interessen', SB: 11-31). Bei der Durchsicht des Literaturverzeichnisses fällt auf, dass einige der dort angeführten Titel für die Studie nur marginal von Bedeutung sein konnten. Andererseits sucht man im Literaturverzeichnis vergebens nach der Dissertation von Hansruedi Stadler (1982), welche den Autoren bekannt war (siehe FG: 3) und sich mit dem Schutz des Heimarbeitnehmers nach dem HArG aus dem Jahre 1981 - also dem Untersuchungsgegenstand der Studie - auseinandersetzt. Die Verarbeitung der Parlamentsprotokolle in den Schlussbericht leidet etwas darunter, dass konsequent falsche Fundstellenangaben verwendet wurden. Eher schwach ist die Behandlung der Aufgaben und Tätigkeiten der Zentralstelle für Heimarbeit, über deren Jahresberichte und Merkblätter die Autoren gemäss dem Literaturverzeichnis verfügten. Die im Forschungsgesuch (FG: 3d) aufgeführten Fragen zur Zentralstelle werden praktisch nicht behandelt (für Ausnahmen, vgl. SB: 69 und 90). Die gesetzlichen Vorschriften, welche das Heimarbeitsverhältnis regeln, sind in einem Anhang zusammengefasst (siehe SB:

119-123). Auffallend ist dabei, dass verschiedene Passagen nicht durch Hinweise auf die einschlägigen rechtlichen Erlasse belegt werden. Dies wirkt stossend, da ja das Kapitel mit der Überschrift 'Die wichtigsten gesetzlichen Regelungen zum Heimarbeitslohn' versehen ist.

Für eine gesamthafte Beurteilung ist sicherlich von zentraler Bedeutung, dass das Ergebnis der qualitativen Analysen überzeugt. Als negativer Punkt ist dem sicherlich beizufügen, dass die Transparenz des Analyseprozesses nicht vollständig gegeben ist. Das Kriterium wird durch die Evaluation mehrheitlich erfüllt.

5.1.3.4.10. Gerechtfertigte Folgerungen

Wir konzentrieren uns hier auf die Folgerungen der Evaluatoren, wie sie im Kapitel 'Praktische Schlüsse in Form von vier Szenarien' (SB: 109-116) vorgelegt werden. Wie bereits schon früher angesprochen, enthalten die Schlussfolgerungen keine klare Präferenzierung eines der vier Szenarien. Die Evaluatoren begründen dieses Vorgehen folgendermassen:

"Je nach Blickwinkel treten unterschiedliche Argumente in den Vordergrund, ohne dass sich sofort entscheiden lässt, welches der 'richtige' Blickwinkel ist. Wir formulieren daher unsere praktischen Schlussfolgerungen in Form von vier alternativen Szenarien, die die weiteren Schritte beschreiben und begründen." (SB: 109) Die vier Szenarien enthalten folgende Perspektiven:

- Szenario 1: Beibehaltung des status quo
- Szenario 2: Gezielterer Vollzug unter Beibehaltung der gesetzlichen Grundlagen
- Szenario 3: Anpassung und Konkretisierung der gesetzlichen Grundlagen (Richt- und Mindestlöhne)
- Szenario 4: Streichung des Heimarbeitsgesetzes

Auch in der nachfolgenden Zusammenfassung (7.6. Schluss, SB: 115-116) fehlt eine deutliche Stellungnahme zu den vier Szenarien. Vergleicht man den Schlussbericht mit dem Schlussberichtsentswurf, der von der Expertengruppe des NFP 27 zur Überarbeitung zurückgewiesen wurde, fallen die grossen Differenzen auf. Im Entwurf nehmen die Autoren klar Stellung und favorisieren ganz klar eine Handlungsoption. Es kommt der Verdacht auf, dass die unverbindliche Darstellung der Schlussfolgerungen mittels der vier Szenarien aufgrund eines Vorschlags der Expertengruppe des NFP 27 zustande kam. Dies würde natürlich automatisch auch in Frage stellen, ob die im endgültigen Schlussbericht präsentierten Schlussfolgerungen überhaupt gerechtfertigt sein können. Abgesehen von diesem Aspekt tauchen aber bei einer genaueren Analyse der vier vorgebrachten Szenarien

rien weitere Schwierigkeiten auf. Das vierte Szenario plädiert für eine Abschaffung des Heimarbeitsgesetzes (SB: 114, vgl. auch SB: 116 und oben Kapitel 5.1.3.1.4.). Dies kann durch die vorgenommene Untersuchung in keiner Weise empirisch belegt werden. Die Evaluation beschäftigt sich lediglich mit *einer* Regelung des Heimarbeitsgesetzes, dem Lohngleichheitspostulat. Alle übrigen Vorschriften - das Gesetz umfasst immerhin 23 Artikel - werden durch die Untersuchung nicht abgedeckt. Weiterhin gehen die Evaluatoren davon aus, dass das Arbeitsgesetz (ArG, SR 822.11) auch auf Heimarbeitsverhältnisse anwendbar ist, was nicht zutrifft (ArG, Art. 3, lit. f; vgl. dazu auch EVD 1995: 11 und 14 sowie Kapitel 5.1.3.1.4.). Das dritte Szenario enthält den Vorschlag, dass Mindestlöhne und Richtlöhne festgelegt werden können. Die Evaluatoren schreiben: "Zusätzlich wird im Gesetz verankert, dass Richt- und Mindestlöhne festgesetzt werden können. Richtlöhne sind Orientierungsgrößen, die ein durchschnittlich leistungsfähiger Heimarbeitnehmer erreichen sollte. Auf den Mindestlohn besteht dagegen ein Rechtsanspruch, wobei ebenfalls vom Arbeitspensum eines durchschnittlich leistungsfähigen Heimarbeitsnehmers ausgegangen wird." (SB: 113) Auffällig bei dieser Vorgehensweise ist der Umstand, dass sowohl bei der Festsetzung der Richt- wie auch der Mindestlöhne von der durchschnittlichen Leistungsfähigkeit eines Heimarbeitnehmers ausgegangen werden soll. Der Vorschlag, Mindestlöhne einzuführen, erstaunt zudem etwas, wenn man sich die Geschichte der Heimarbeitsgesetzgebung vor Augen führt. Vor der Revision des heute gültigen Heimarbeitsgesetzes bestand nämlich bereits einmal die Möglichkeit der Festlegung von Mindestlöhnen. Die Evaluatoren schreiben dazu: "Die Mindestlohnregelung des 1. Heimarbeitsgesetzes erwies sich auf Dauer als nicht praktikabel; zu unterschiedlich waren die Arbeits- und Einkommensbedingungen von Region zu Region; zu gegensätzlich die Interessen von Arbeitgebern und Arbeitnehmern, auch innerhalb dieser Gruppen. Um die Ausgabe von Heimarbeit nicht zu gefährden, wurden zudem die Mindestlöhne vielfach so tief festgesetzt, dass das Schutzziel ausgehöhlt wurde." (SB: 26) Aufgrund dieser Erfahrungen mit einer Mindestlohnregelung erscheint der Vorschlag zu deren Wiedereinführung einem starken Rechtfertigungsdruck ausgesetzt. Diesem vermag die Argumentation der Autoren kaum standzuhalten. Das dritte Szenario umfasst auch die Entschlackung des Lohngleichheitspostulates. "Es wird entschlackt, indem die Mehr- und Minderaufwendungen gestrichen werden." (SB: 113, vgl. HArG, Art. 4, Abs. 1, siehe auch oben, Kapitel 5.1.1.1.) Auch eine derartige Regelung war bereits einmal gültig, nämlich im Uhrenstatut aus dem Jahre 1961, das 1971 ausser Kraft gesetzt wurde (Art. 9 Abs. 2, AS 1961: 1089). Erstaunlicherweise wird diese Regelung in der Studie überhaupt nicht erwähnt, obwohl sich ja eine der vier Fallstudien mit der Uhrenindustrie befasst.

Ein weiterer Aspekt, der hier anzufügen ist, betrifft die Existenz von Missbräuchen im Bereich der Heimarbeit. Aufgrund der Vorgehensweise in der Evaluation - zu der es

kaum valable Alternativen gab - ist nicht völlig auszuschliessen, dass die Bedeutung von Missständen durch die Studie unterschätzt wird (vgl. dazu Kapitel 5.1.3.4.5.). Die Autoren äussern sich im Kapitel zur Methodik dazu folgendermassen: "Ob es einzelne 'schwarze Schafe' unter den Arbeitgebern und einzelne systematisch ausgebeutete Heimarbeitnehmer gibt, lässt sich nicht beurteilen, aber auch nicht ausschliessen." (SB: 140-141) Auf diesem Hintergrund erscheint die Aussage in der Zusammenfassung des Evaluationsberichts, dass ausbeuterische Arbeitgeber verschwunden seien (SB: 107), etwas zu absolut.

Der Gesamteindruck der Schlussfolgerungen ist nicht sehr positiv. Sie erscheinen teilweise etwas zu wenig durchdacht, teilweise fehlt eine empirische Abstützung. Erschwerend kommt hinzu, dass die Formulierung der Folgerungen unseres Erachtens zu offen geblieben ist. Zur Einschätzung der vier Szenarien vermisst der Leser die Diskussion der mit den vier Szenarien verbundenen, je spezifischen Vor- und Nachteile. Dies würde es ihm erlauben, aufgrund seiner persönlichen Präferenzen eine Bewertung (und Entscheidung) zu den vier Szenarien vorzunehmen. Die Evaluation erfüllt das Kriterium aus den angeführten Gründen nicht.

5.1.3.4.11. Bestätigbarkeit

Wir sind bereits früher (siehe Kapitel 5.1.3.4.5.) auf das Problem gestossen, dass die Darstellung des Schlussberichts nur selektiv Einblicke in den Entstehungsprozess der Ergebnisse zulässt. Besonders schwierig ist dies bei der Interpretation der Interviews. Die eingesetzten Gesprächsleitfaden sind im Schlussbericht nicht enthalten. Die Autoren sind, da auf eine Erhebung mit Tonband verzichtet wurde, nicht in der Lage, Interpretationen in transparenter Weise auf Gesprächssequenzen rückzuführen (vgl. Kapitel 5.1.3.4.9.). Ansätze zur Herstellung der Rückführbarkeit sind zwar sehr wohl in der Arbeit enthalten, diese vermögen jedoch nicht voll zu befriedigen. Angesichts der Schwierigkeit des Unterfangens muss jedoch den Evaluatoren zugute gehalten werden, dass wenigstens der Versuch dazu unternommen worden ist. An einigen Stellen ist dies auch gelungen. Aus diesen Überlegungen heraus ist es nicht gerechtfertigt, der Evaluation hier ein allzu schlechtes Zeugnis auszustellen. Das Kriterium wird deshalb trotz erheblicher Mängel, als knapp erfüllt betrachtet.

5.1.4. Gesamteindruck

Die Evaluation zum 'Lohnleichheitspostulat des Heimarbeitsgesetzes' - eine Thematik, der in der politischen Diskussion keine hohe Beachtung zukommt - hinterlässt gesamthaft gesehen einen positiven, sehr ausgeglichenen Eindruck. Es lassen sich zu den vier Bereichen (Nützlichkeit, Anwendbarkeit, Korrektheit, Genauigkeit) keine ausgeprägten Stärken oder Schwächen finden. Auf der Ebene der Einzelkriterien finden sich dagegen sehr wohl Unterschiede. Die Evaluation zeichnet sich dadurch aus, dass sie sich, auch wenn sich ihr zum Teil im Evaluationsprozess grössere Probleme entgegenstellten, im positiven Sinn sehr anpassungsfähig zeigte. Es ist gelungen, in pragmatischer und praktikabler Weise mit diesen Schwierigkeiten umzugehen. Dabei spielte sicher auch die gewählte Methodik eine gewichtige Rolle. Die Evaluation vermittelt aussenstehenden Dritten zudem einen guten Einblick in die Thematik, was nicht zuletzt auch durch eine (im zweiten Anlauf) gelungene Schlussberichterstattung ermöglicht wurde. Eher negativ fällt bei der Bewertung durch die Kriterien ins Gewicht, dass die Schlussfolgerungen ausgesprochen zurückhaltend präsentiert werden. Auf eine prononcierte Stellungnahme und Bewertung der vorgestellten Szenarien wird verzichtet. Dies führte in unserer Analyse dazu, dass sich die Evaluation beim Kriterium 'Bewertende Interpretationen' einer Bewertung entzog (siehe Kapitel 5.1.3.1.4.) und dass sie andere Kriterien, wie zum Beispiel jenes zu 'Gerechtfertigte Folgerungen' (siehe Kapitel 5.1.3.4.10.) nicht oder nur unvollständig erfüllen konnte. Schwächen zeigen sich primär bei den betont vorsichtigen Schlüssen, die aus der interessanten Analyse gezogen wurden. Gesamthaft betrachtet liegt aber eine gelungene Evaluation vor.

5.1.5. Anhang

5.1.5.1. Unterlagen zur Evaluationsstudie

Die Untersuchung dieser Evaluationsstudie beruht auf der Analyse von schriftlichen Materialien und auf Gesprächen mit einigen relevanten Akteuren (siehe Kapitel 5.1.5.2.). Zur Erarbeitung dieser Fallstudie sind folgende schriftlichen Unterlagen verwendet worden:

- Forschungsskizze vom 10/1/1990
- 2. Forschungsskizze vom 29/6/1990
- Schreiben BIGA an Präsident der Expertengruppe vom 26/9/1990
- Schreiben Schmid an Programmleitung vom 26/9/1990
- Forschungsgesuch vom 27/11/1990 (FG)
- Bewilligungsschreiben Abt. IV des SNF vom 17/1/1991
- Projektkurzbeschrieb vom Februar 1991
- Schreiben Projektteam an Mitglieder der Projektkommission vom 19/6/1990
- Wissenschaftlicher Zwischenbericht vom 25/2/1992 (WZB)
- Finanzieller Zwischenbericht vom 25/2/1992 (FZB)
- Aktennotiz zu einer Besprechung vom 31/3/1992
- Entwurf des Schlussberichts mit Methodendiskussion und Zusammenfassung vom 18/8/1992 (ESB)
- Schreiben des Revisorats SNF zur Projektverlängerung vom 24/8/1992
- Stellungnahme Bussmann zum Entwurf Schlussbericht vom 1/9/1992
- Stellungnahme Buschor zum Entwurf Schlussbericht vom 3/9/1992
- Aktennotiz zu einer Besprechung vom 24/9/1992
- Neuer Gliederungsentwurf vom 24/9/1992
- Schreiben von Gerheuser an Programmleitung vom 28/11/1992
- Aktennotiz zu einer telephonischen Besprechung vom 16/12/1992
- Schreiben des Revisorats SNF zur Projektverlängerung vom 20/1/1993
- Überarbeiteter neuer Gliederungsentwurf vom 15/2/1993
- 2. Entwurf Schlussbericht von Ende März 1993
- Schreiben Gerheuser an Programmleitung vom 24/6/1993
- Kostenaufstellung für Überarbeitung Schlussbericht vom 27/6/1993
- Schlussbericht in der NFP27-Reihe (SB)
- Brief Gerheuser an Programmleitung vom 6/8/1993
- Bewilligung des Revisorats SNF zur Projektverlängerung und zum Zusatzkredit vom 28/9/1993
- Finanzieller Schlussbericht vom 30/9/1993 (FSB)
- Kurzbericht (NFP27-Broschüre) (KB)

5.1.5.2. Liste der kontaktierten Personen

- Werner Bussmann, Programmleitung NFP 27, Bern
- Frohmut W. Gerheuser, Brugg
- Ueli Leupin, Geschäftsleiter, Schweiz. Zentralstelle für Heimarbeit, Bern
- Colette Nova, ehem. Abt. Arbeitnehmerschutz und Arbeitsrecht, BIGA, Bern
- André Schmid, Brig
- Anton Wydler, Abt. Arbeitsinspektorat, Amt für Industrie, Gewerbe und Arbeit des Kantons Zürich, Zürich

5.1.5.3. Literaturverzeichnis

- ABT, R./BELLWALD, Anton/ZURSCHMITTEN, Klaus (1981): Entwicklungsengpässe und Innovationsverhalten bestehender Betriebe im Berggebiet. NFP 5: Regionalprobleme in der Schweiz, Arbeitsbericht Nr. 21, Bern: NFP 5.
- BELLWALD, Anton (1983): "Die Ansprüche des Tourismus an die Landwirtschaft." In: Walther, Bernhard/Popp, Hans W. (Hrsg.) (1983): Erwerbskombinationen in der Berglandwirtschaft. Grösch: Rügger: 165-175.
- BELLWALD, Anton et al. Der touristische Arbeitsmarkt: Angebot, Nachfrage und Steuerungsmöglichkeiten. NFP 5: Regionalprobleme in der Schweiz, Arbeitsbericht Nr. 51, Bern: NFP 5.
- CASPAR, Reta/GERHEUSER, Frohmüt W./HENZ, Hans-Rudolf (1985): "Die Bedeutung regionaler Entwicklungsträger." In: Hänni, Heinz D. (Hrsg.): Raumordnungspolitik im Vollzug: Anspruch und Wirklichkeit. Grösch: Rügger: 359-384.
- COOK, Thomas D. (1985): "Postpositivist critical multiplism." In: Shotland, Robert Lance/Mark, M. (eds.): Social science and social policy. Beverly Hills: Sage: 21-62.
- EIDG. BÜRO FÜR DIE GLEICHSTELLUNG VON FRAU UND MANN (Hrsg.) (1991): Arbeitsbewertung und Lohndiskriminierung von Frauen. Bern: EDMZ.
- EIDG. JUSTIZ- UND POLIZEIDEPARTEMENT (EJPD) (Hrsg.) (1988): Lohngleichheit für Mann und Frau. Schlussbericht der vom Eidg. Justiz- und Polizeidepartement eingesetzten Arbeitsgruppe Lohngleichheit. Oktober 1988. Bern: EJPD.
- EIDG. VOLKSWIRTSCHAFTSDEPARTEMENT (EVD) (Hrsg.) (1995): Die Verbesserung des Schutzes der Heimarbeiterinnen und der Heimarbeiter im Sinne des Schlussberichtes der Arbeitsgruppe "Lohngleichheit für Mann und Frau". Bericht des EVD in Erfüllung eines Auftrages des Bundesrates. 12. Juli 1995. Bern. (mimeo)
- GAGG, Margareta (1927): "Weibliche Heimarbeit in der Schweiz." In: Zeitschrift für Schweiz. Statistik und Volkswirtschaft, 63. Jahrgang, Heft 1: 104-138 und Heft 2: 149-200.
- GERHEUSER, Frohmüt W. (1971): Soziologische und sozialpsychologische Hinweise zum Mietwohnungsbau. Brugg: Metron.
- GERHEUSER, Frohmüt W. (1984): "Der Arbeitsmarkt der mittleren Kader in peripheren Regionen." In: Brugger, Ernst A. (Hrsg.): Arbeitsmarktentwicklung: Schicksalsfrage der Regionalpolitik? Grösch: Rügger: 227-242.
- GERHEUSER, Frohmüt W. (1988): "Pragmatische Soziologie - oder die unpraktische Soziologie." In: Schweiz. Zeitschrift für Soziologie, Vol. 3: 535-548.
- GERHEUSER, Frohmüt W. (1991): Die Wirkungen der beruflichen Vorsorge auf den Arbeitsmarkt. Chur: Rügger.
- GERHEUSER, Frohmüt W. et al. (1988): "Agglomerationen - Problemgebiete ohne Handlungskompetenz?" In: Bassand, Michel/Joye, Dominique/Schuler, Martin (Hrsg.): Agglomerationsprobleme in der Schweiz. Bern: Lang: 233-257.
- GERHEUSER, Frohmüt W./CASTELAZZI, Eveline (1982): Grundlagenbeschaffung für die Planung der Quartiererneuerung. Bewohner und Hauseigentümer. Schriftenreihe Wohnungswesen, Band 25. Bern: EDMZ.
- GERHEUSER, Frohmüt W./MANGOLD, Hans (1981): Form, Struktur und Funktionalität schweizerischer Stadtsysteme. Der Wandel der schweizerischen Zentrenhierarchie. Ursachen - Entwicklungen - regionalpolitische Konsequenzen. NFP 5: Regionalprobleme in der Schweiz, Arbeitsbericht Nr. 19, Bern: NFP 5.

- GERHEUSER, Frohmut W./MANGOLD, Hans (1982): Periphere Arbeitsmärkte für mittlere Kader in der Zentrenhierarchie. NFP 5: Regionalprobleme in der Schweiz, Arbeitsbericht Nr. 26, Bern: NFP 5.
- GERHEUSER, Frohmut W./PERRET-GENTIL, Jean-Claude (1993): Gewerbeverdrängung. Gewerbeentwicklung und Gewerbepolitik in den schweizerischen Grossstädten. Bericht 19 des NFP 'Stadt und Verkehr'. Zürich: NFP 25.
- GERHEUSER, Frohmut W./SARTORIS, Elena (1989): Städtische Liegenschaftsmärkte im Spannungsfeld privater und institutioneller Anleger. Entwicklungen auf dem Liegenschaftenmarkt für Anlageobjekte, insbesondere für Altbauten, 1970-1985. Arbeitsberichte Wohnungswesen, Nr. 19. Bern: Bundesamt für Wohnungswesen.
- GERHEUSER, Frohmut W./SCHUMANN, Carola (1973): Kommunikatives Wohnen. Brugg: Metron.
- HENZ, Hans-Rudolf/GERHEUSER, Frohmut/CASPAR, Reta (1984): Die Bedeutung der regionalen Entwicklungsträger. Tätigkeitsfelder, Leistungsfähigkeit und Funktionsweise der Regionen. NFP 5: Regionalprobleme in der Schweiz, Arbeitsbericht Nr. 47, Bern: NFP 5.
- MÜLLER, Georg (1987): "Art. 4." In: Aubert, Jean-François et al. (Hrsg.): Kommentar zur Bundesverfassung der Schweizerischen Eidgenossenschaft vom 29. Mai 1874. Basel/Zürich/Bern: Helbing & Lichtenhahn/Schulthess/Stämpfli.
- PARLAMENTARISCHE VERWALTUNGSKONTROLLSTELLE (Hrsg.) (1994): Evaluation der regionalpolitischen Koordination von Bundespolitiken. Schlussbericht zuhanden der Geschäftsprüfungskommission des Nationalrates. Verfasser: Frohmut W. Gerheuser. Bern: EDMZ.
- REGLI, Felix (1979): Heimarbeit in der Schweiz. Eine volkswirtschaftliche und sozialpolitische Studie mit besonderer Berücksichtigung des Kantons Uri. Diss. Hochschule St. Gallen.
- SPINATSCH, Markus/GERHEUSER, Frohmut W. (1994): "Die regionalpolitische Koordination von Bundespolitiken." In: NZZ, 5. 9. 1994, Nr. 206: 19.
- STADLER, Hansruedi (1982): Der Schutz des Heimarbeitnehmers (Nach dem Heimarbeitsgesetz vom 20. März 1981). Diss. Universität Bern.
- TANZER, Franz/BÖHM, Christian (1986): Heimarbeit in Österreich. Schriftenreihe über Arbeit und Arbeitsbeziehungen, 6. Wien: Bundesministerium für soziale Verwaltung.
- VORJANS, Bernd (1987): Tele-Heimarbeit. Möglichkeiten - Probleme - Perspektiven. Frankfurt: Lang.
- ZURSCHMITTEN, Klaus (1984): "Arbeitsmarktentwicklung: Schicksalsfrage der Regionalpolitik?" In: Brugger, Ernst A. (Hrsg.): Arbeitsmarktentwicklung: Schicksalsfrage der Regionalpolitik? Grösch: Rüeegg: 199-217.

5.1.6. Stellungnahme der Evaluatoren

5.1.6.1. Vorbemerkung

Diese Anmerkungen zur Meta-Evaluation unseres Projektes geben ausschliesslich meine persönliche Sicht als ein Mitarbeiter am Projekt und Verfasser des schliesslich akzeptierten Schlussberichtes wieder, auf den sich die Meta-Evaluation im wesentlichen bezieht.

5.1.6.2. Generelle Bemerkungen zum vorliegenden Meta-Evaluationsbericht

Die Arbeit von T. Widmer, einen Kriterienkatalog für Meta-Evaluationen auszuarbeiten und auf eine Reihe von Fallstudien anzuwenden, ist aus meiner Sicht sehr verdienstvoll. Sie entspricht einem Bedürfnis, in diesem Metier professionelle Standards zu etablieren.

Soweit ich anhand der mir vorliegenden Teile beurteilen kann, liefert der Bericht eine gute Grundlage für eine fruchtbare Diskussion zur quality control von Evaluationen. Dies gilt auch dann, wenn ich nicht mit allem, was darin ausgeführt ist, einverstanden bin. Dieser Diskussion, die kommen muss, will ich nicht vorgreifen, jedoch auf einzelne Punkte hinweisen, die aus meiner Sicht zu reden geben:

- a) die Länge und Konsistenz des Kriterienkatalogs: Die vier Grunddimensionen scheinen mir richtig gewählt; die Zahl der Unterkriterien ist dagegen unübersichtlich gross. Aus der Wiederholung von gleichen Argumenten unter mehreren Rubriken schliesse ich, dass die Abgrenzung zwischen den Kriterien nicht trivial ist.
- b) Anwendbarkeit eines einheitlichen Kriterienkatalogs auf Studien mit unterschiedlichen konzeptionellen und methodischen Ansätzen: Um der Vergleichbarkeit willen ist ein einheitlicher Kriterienkatalog unabdingbar. Zugleich sollten jedoch die gewählten Kriterien den jeweiligen konzeptionellen und methodischen Ansätzen der einzelnen Evaluationen gerecht werden. Wie mit diesem Widerspruch umzugehen ist, scheint mir weiterer Überlegungen wert.

5.1.6.3. Zur Meta-Evaluation unserer Studie

5.1.6.3.1. Kritik am Bearbeitungsteam

Der Meta-Evaluator hat richtig erkannt, dass das Bearbeitungsteam in der Schlussphase nicht mehr harmonierte. Ob und wie sich das hätte vermeiden lassen, ist mir auch heute

noch nicht ganz klar. Der Initiator eines Projektes ist in einer Arbeitsgemeinschaft Projektleiter, daran lässt sich intern nicht rütteln.

5.1.6.3.2. Redaktionelle Kritikpunkte am Schlussbericht

Die festgestellten redaktionellen Mängel wie falsche Quellenangaben, wenig prägnante Kapitelüberschriften, inkonsistente Zahlenangaben, nicht revidierte Bibliographie, fehlender Gesprächsleitfaden etc. lassen sich nicht bestreiten. Das sind Schönheitsfehler, die uns aus Finanz- und Zeitdruck entgangen sind. Sie hätten sich mit geringem Aufwand beheben lassen.

5.1.6.3.3. Methodische Kritikpunkte

a) Fallstudien

Angesichts der Heterogenität der in regionalisierten Branchenstudien erfassten Heimarbeitsverhältnisse ist es kaum gerechtfertigt, von eigentlichen Fallstudien zu reden. Die Unterscheidung nach Branchen ist für den Lohnvergleich bis zu einem gewissen Grade relevant, andere Variablen, die quer dazu liegen und die zugrundeliegenden Mechanismen beeinflussen, sind ungleich gewichtiger.

b) Fehlender Einbezug weiterer Akteure

Es wird moniert, dass entgegen den Ankündigungen im Forschungsgesuch eine Reihe von Akteuren nicht einbezogen wurden, so insbesondere die Arbeitgeber- und Arbeitnehmerverbände, die Gerichte, die Zentralstelle für Heimarbeit und vom Gegenstand her naheliegend Vertreter von Frauenorganisationen.

Der Organisationsgrad der Heimarbeiter ist äusserst gering; ebensowenig sind nach unseren Auskünften Heimarbeiter systematisch organisiert, in einzelnen Branchen allenfalls auf regionaler Ebene, jedoch ohne eigentliche heimarbeitsbezogene Verbandstruktur. Diese Verbände sind somit keine relevanten Informationslieferanten; ihre Vernachlässigung ist daher begründet.

Die Vernachlässigung der Gerichte ergibt sich aus der einfachen Tatsache, dass nur vereinzelte Gerichtsfälle aktenkundig waren, die zudem Jahre zurücklagen und bereits aufgelöste Arbeitsverhältnisse betrafen. Die Gerichte spielen praktisch keine Rolle. Warum

das nicht der Fall ist und welche Konsequenzen sich daraus ergeben, sind die hier relevanten Erkenntnisse.

Die Zentralstelle für Heimarbeit war in der Begleitgruppe vertreten. Ihre Funktion wurde bewusst nicht eingehend analysiert, weil sie in den einschlägigen Kreisen nicht unumstritten ist. Das hätte einen Nebenschauplatz eröffnet, der vom Hauptthema abgelenkt hätte.

Der Kontakt zu Frauenorganisationen wurden z. B. über das St. Galler Gleichstellungsbüro gesucht. Doch waren die Probleme der Heimarbeiterinnen dort nicht bekannt. Heimarbeiterinnen scheinen auch hier im Schatten der Aufmerksamkeit zu stehen.

Widmers Bemängelungen sind zudem einseitig, weil sie die Ausweitungen unerwähnt lassen. Unsere Wahl, mit beschränkten Mitteln mehr Heimarbeitnehmer und vor allem zusätzlich Heimarbeitgeber einzubeziehen, war richtig. Indirekt wird hier auf eine Lücke in der Bewertungskriterien hingewiesen, die die Forschungsökonomie betrifft.

c) Erhebungs- und Auswertungsmethode

Idealiter hätten wir von allen Gesprächen Tonbandprotokolle gemacht, diese ganz oder teilweise transkribiert und anschliessend nach den Regeln der Kunst ausgewertet. Dieses Vorgehen ist aufwendig und war mit unseren Mitteln nicht machbar. Seine Transparenz und Nachvollziehbarkeit sind daher kaum ein Massstab. Unser Vorgehen war einfacher, pragmatischer und so systematisch wie möglich, naturgemäss eher induktive Kunst als deduktive Wissenschaft. Seine methodischen Abkürzungen bergen gewisse Risiken für die Zuverlässigkeit der Resultate und auch für die Transparenz und Nachvollziehbarkeit des Analyseprozesses; das ist nicht zu leugnen. Forschungsökonomie ist jedoch auch hier ein relevanter Gesichtspunkt, dessen Einbezug die Akzente verschieben könnte. Die anerkannte Stimmigkeit, Plausibilität und Nachvollziehbarkeit der inhaltlichen Rekonstruktionen dienten uns als methodisches Kontrollinstrument.

d) die Validität der Resultate (siehe Kapitel 5.1.3.4.5 a)

Widmer erkennt die Plausibilität unserer Erkenntnisse an; er zweifelt jedoch unser Validitätskriterium an, das ich weiterhin für gut halte. Entscheidend ist nach meinem Dafürhalten die Tatsache, dass trotz den sehr unterschiedlichen Interessenlagen die Aussagen der einbezogenen Gruppen alle in die gleiche Richtung weisen. Eine systematische, gleichgerichtete Verzerrung ihrer Aussagen erscheint mir als unwahrscheinlich.

5.1.6.3.4. Inhaltliche Kritikpunkte

a) Positionslosigkeit und Unverbindlichkeit der Schlussfolgerungen

Widmer behauptet (Kapitel 5.1.3.3.2.): "Zentral für diese Evaluation ist der Interessengegensatz zwischen einer liberalen marktwirtschaftlichen Haltung, die für eine vermehrte Deregulierung eintritt, einerseits und einer sozialen politischen Perspektive, die primär auf eine Optimierung des Schutzes der Heimarbeiter abzielt. Ausgerichtet auf den Untersuchungsgegenstand geht es also - vereinfacht gesagt - um die Frage, ob das Lohngleichheitspostulat trotz veränderten Rahmenbedingungen nach wie vor einem Bedarf entspricht."

Dieser von Widmer hypostasierte zentrale Interessengegensatz betrifft jedoch nur die Regulierungsebene, ob und welche gesetzlichen Massnahmen ergriffen werden sollen. Er vernachlässigt die Bedingtheit und Widersprüchlichkeit der zugrundeliegenden Interessenlagen von Heimarbeitgebern, Heimarbeitnehmern und Aufsichtsorganen. Auf deren Interessenlagen samt deren Inkonsistenzen wurde ausführlich eingegangen. Daraus leitet sich ab, welche gesetzlichen Massnahmen auf der Regulierungsebene wirksam sein können, unabhängig von ideologischen Positionen.

Unter günstigen Rahmenbedingungen ist der Interessenausgleich zwischen Heimarbeitgebern und -nehmern möglich, der gleichzeitig einen fairen Lohnsatz und Beschäftigung und damit Einkommen sichert. Unter ungünstigen wirtschaftlichen Voraussetzungen, wenn Heimarbeit knapp, das Einkommen benötigt und der gesetzliche Schutz erforderlich werden, bricht dieser Interessengegensatz auf; Lohnansprüche lassen sich nicht mehr wirksam durchsetzen; denn Ausgabe und Zuteilung von Heimarbeit entziehen sich staatlicher Kontrolle, faktisch ebenfalls die Arbeits- und Lohnverhältnisse der HeimarbeiterInnen; Lohnansprüche sind schliesslich nicht einklagbar, ohne die Beschäftigung aufs Spiel zu setzen. Der Meta-Evaluator scheint die Schlüsselfunktion dieser Erkenntnisse nicht hinreichend gewürdigt zu haben.

Aus dieser Sachlage heraus ergibt sich, dass es keinen Königsweg gibt und keine Patentlösung, die grosse Wirksamkeit entfalten könnte. Die vier vorgeschlagenen Alternativen sind zweit- und drittbeste Auswege, ohne dass das zugrundeliegende Dilemma gemildert oder gar beseitigt werden könnte. Sich diesem Dilemma mit allen Konsequenzen bewusst zu stellen, hat nichts mit Positionslosigkeit zu tun.

b) Zu den Szenarien

Es trifft zu, dass sich frühere (landesweite) Mindestlohnregelungen nicht bewährt haben. Die Gründe dafür liegen sowohl in der Handhabung als auch in den Interessen, die damit verbunden waren. Unter heutigen Umständen können nach meinem Dafürhalten flexibel gehandhabte Richt- und Orientierungswerte für Lohnsätze ein gangbarer Weg sein, vor allem weil der Vollzug von solchen impliziten Mindestwerten ohnehin ausgeht und aus praktischen Gründen ausgehen muss.

Ebenso trifft der Einwand zu, dass aufgrund unserer Arbeit Schlüsse, das ganze Heimarbeitsgesetz (HArG) zu streichen, an sich nicht zulässig sind. Wie man uns von kundiger Seite versichert hat, hätten bei einem Wegfall des Lohngleichheitspostulates die anderen Teile des HArG durch geringfügige Anpassungen anderer Gesetze abgedeckt werden können. Unter den Voraussetzungen hätte das ganze HArG tatsächlich gestrichen werden können.

Rückblickend hätte es sich dennoch gelohnt, die Szenarien noch etwas weiter auszuarbeiten und systematisch gegeneinander abzuwägen, was eine zweite Evaluation in sich wäre. Ob wir vom Schreibtisch aus zu eindeutigen Schlüssen und Empfehlungen gelangt wären, wage ich zu bezweifeln. Dazu wäre es sinnvoll und notwendig gewesen, unmittelbar mit der zuständigen Arbeitsgruppe der Bundesverwaltung zusammenzuarbeiten, um deren Überblick und praktische Erfahrung zu nutzen.

Nach meiner Erfahrung als Evaluator ist es in unklaren politischen Entscheidungssituation fruchtbarer, diskutabile und begründete Alternativen vorzuschlagen und damit auch das Entscheidungsspektrum mit all seinen Vor- und Nachteilen, Gewissheiten und Unwägbarkeiten zu umreißen, als eine scheinbar eindeutige und umso fragwürdigere Position zu beziehen. Drum halte ich daran fest, dass in unserem Fall des HArG-Lohngleichheitspostulates und beim gewählten Forschungsansatz die alternativen Szenarien trotz Mängeln im Detail die bestmöglichen Schlussfolgerungen darstellen.

Frohmut W. Gerheuser

5.2. FALLSTUDIE 'SCHNEERÄUMUNG VON URNER PASSSTRASSEN' (STUDIE NR. 6)

5.2.1. Die Evaluation

5.2.1.1. Das Thema

Die Evaluation (Sommer/Suter 1993; Schlussbericht (SB)) befasst sich mit der Schneeräumung auf Urner Passstrassen im Frühling. Der Kanton Uri setzt dabei jeweils erhebliche Mittel ein, um eine relativ frühe Passöffnung zu ermöglichen. Von einer frühen Öffnung der fünf Urner Pässe - Gotthard, Klausen, Susten, Furka und Oberalp - erhofft man sich positive Effekte, wie die Förderung des Tourismus, die Sicherung der Alpwirtschaft, die Ermöglichung von Unterhaltsarbeiten im Gotthard-Strassentunnel und die Förderung sozialer Kontakte. Die Kurzevaluation untersucht, ob und inwiefern diese Zielsetzungen durch die Schneeräumung erreicht werden.

5.2.1.2. Die Fragestellung

Die Evaluation behandelt die folgenden Fragen:

- "- Werden mit der Schneeräumung tatsächlich die erwarteten Wirkungen ausgelöst? (Effektivitätskontrolle)
- Werden die Ziele (z.B. Förderung des Tourismusaufkommens) im festgelegten Umfang erreicht? (Zielerreichungskontrolle)
- Könnte die Schneeräumung effizienter ausgestaltet werden, also das Verhältnis von Kosten zu Nutzen durch eine Änderung im Vollzug der Massnahme verbessert werden? (Effizienzkontrolle)" (SB: 11)

Explizit weisen die Autoren darauf hin, dass folgende Fragen in dieser Evaluation nicht untersucht werden:

- Wären alternative Massnahmen geeigneter zur Erreichung der Ziele? (nach SB: 11)
- "Ist das Detailziel 'Förderung des Tourismus' sinnvoll, um das Grobziel 'Wirtschaftsförderung (Sicherung von bestehenden und Ansiedlung von neuen Arbeitsplätzen)' zu erreichen?
- Sollte die Festlegung der Detailziele geändert werden?" (SB: 11)

5.2.1.3. Das Vorgehen

Die Evaluatoren folgen in dieser Studie jener Vorgehensweise, welche sie an anderer Stelle entwickelt haben (vgl. dazu Neuenschwander/Sommer/Walter 1990). Unterschieden werden insgesamt sechs Untersuchungsphasen (vgl. SB: 3-8):

1. Vorarbeiten (Festlegung der Massnahme(n) und der Fragestellung(en), welche zu untersuchen sind; Definition des Bestimmungsgrads der Evaluation; organisatorische Fragen, wie zeitliche Planung, Datenzugang usw.)
2. Zieldefinition (Definition der Zieldimensionen, wenn möglich durch messbare Zielgrössen und mit Soll-Werten)
3. Vollzugskontrolle (Untersuchung des Vollzugs, evtl. Festhalten von Abweichungen von der vorgesehenen Implementationsweise; Eruiierung von Vollzugsproblemen, Erheben von Datenmaterial zum Vollzug)
4. Wirkungsanalyse (Erarbeitung der zu erwartenden Wirkungsmechanismen; Messung der Wirkungen; Bewertung der Messergebnisse)
5. Auswertung (wie Effektivitäts-, Effizienz- oder Zielerreichungskontrolle, Abschätzung der 'Vorteilhaftigkeit', Vergleich mit alternativen Massnahmen(varianten), Überprüfung des Zielsystems)
6. Umsetzung (Formulierung von Empfehlungen oder Vorschlägen)

Zur Erhebung des Datenmaterials stützen sich die Evaluatoren zum einen auf bestehende Datengrundlagen der kantonalen Verwaltung (Bauamt, KIGA, Polizeidirektion) und anderer Quellen (BfS, Verkehrsvereine usw.). Zum anderen werden mit verschiedenen beteiligten und/oder betroffenen Personen Gespräche durchgeführt. Zur Auswertung quantitativer Daten verwenden die Evaluatoren primär die graphische Analyse, die Berechnung einfacherer statistischer Kennziffern und die Regressionsanalyse. Ergänzt werden diese Analysen mit qualitativen Erkenntnissen, welche die Autoren vor allem aus den Gesprächen gewonnen haben.

5.2.1.4. Der Auftraggeber

Auftraggeber dieser Evaluationsstudie ist das Bauamt des Kantons Uri. Begleitet wurde die Evaluation seitens des Auftraggebers durch drei Personen aus dem Bauamt Uri, nämlich durch den Leiter der Hauptabteilung Strassen (Edoardo Frei), durch den für die Schneeräumung verantwortlichen Betriebsleiter (Josef Brand) und durch den beim Bauamt Uri für das Rechnungswesen Betrieb Zuständigen (Josef Zurfluh).

Die Studie ist im Rahmen des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) unter dem Titel 'Kurzevaluation' mit einem finanziellen Beitrag unterstützt worden.

5.2.1.5. Die Forscher

Das Forschungsteam besteht aus Heini Sommer (Projektleiter) und Stefan Suter.

Heini Sommer (Jahrgang 1958) hat an der Universität Bern Volks- und Betriebswirtschaft studiert. Während des Studiums war Sommer als Werkstudent während zweier Jahre bei IBM-Schweiz als PC-Koordinator tätig. Nach Abschluss des Studiums arbeitete er als Assistent bei Prof. Ernst Baltensperger, der auch seine Dissertation betreute. Diese Arbeit trägt den folgenden Titel: 'Kantonale Nationalstrassen-Rechnung. Die Kosten und Erträge des Kantons Uri, ihr Vergleich mit anderen Kantonen und eine ökonomische Analyse der bundesstaatlichen Regelungen im Nationalstrassenbereich' (Sommer 1990). Nach Abschluss der Doktorarbeit (Dr. rer. pol., 1989) kam Sommer zum Beratungs- und Planungsbüro Ecoplan, welches 1988 als Kollektivgesellschaft gegründet worden war. Seither arbeitet Sommer bei Ecoplan in Altdorf. Er hat dort an der Erarbeitung verschiedener Studien mitgewirkt, die sich schweremässig mit Fragen zur Wirtschafts-, Finanz-, Verkehrs- und Energiepolitik befassen. Weiter hat Sommer auch einige Projekte zu ökonomischen Aspekten der Wasserversorgung und Abwasserentsorgung bearbeitet (vgl. dazu ausführlicher Kapitel 5.2.3.1.2.). Beim hier zur Diskussion stehenden Evaluationsprojekt hat Sommer den Hauptteil der Arbeit geleistet und war zudem verantwortlicher Projektleiter.

Stefan Suter (Jahrgang 1962), das andere Mitglied des Forschungsteams, hat an der Universität Bern Volkswirtschaft, Politologie und Staatsrecht studiert (lic. rer. pol.). Seit 1990 arbeitet Suter ebenfalls bei Ecoplan. Seither hat er bei verschiedenen Forschungsprojekten mitgewirkt (vgl. Kapitel 5.2.3.1.2.). In der hier zur Diskussion stehenden Evaluation hatte Suter lediglich einen relativ kleinen Anteil an den Forschungsarbeiten. Er übernahm vor allem die Rolle des kritischen Erstlesers.

5.2.2. Beschreibung des Evaluationsprozesses

Das Nationale Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen' (NFP 27) führte im Juni 1992 zuhanden der Kantone und grösserer Städte eine Ausschreibung durch, in der auf die Möglichkeit hingewiesen wurde, dass kleinere Evaluationsprojekte ('Kurzevaluationen') öffentlicher Stellen mit einem Kostenbeitrag des NFP 27 versehen werden können. Das Büro Ecoplan schrieb daraufhin verschiedene Stellen in der Zentralschweiz an, um darauf hinzuweisen, dass es in der Lage wäre, derartige Projekte zu bearbeiten. Am 7. Oktober 1992 fand zwischen dem Bauamt des Kantons Uri und Ecoplan eine erste Kontaktaufnahme statt. Daraufhin entschloss sich der Kanton Uri, beim NFP 27 einen Kostenbeitrag zu beantragen und Ecoplan mit der Ausarbeitung einer Offerte/Projektskizze zu betrauen. Ende Oktober lagen die Offerte und das Gesuch zur Mitfinanzierung vor und wurden bei der Programmleitung des NFP 27 eingereicht. Am 3. November 1992 bestätigte der Programmleiter des NFP 27, Werner Bussmann, den Eingang des Gesuchs. Der Kantonsingenieur, Peter Püntener (Leiter des Bauamts Uri), teilte Bussmann in einem Schreiben vom 17. November 1992 mit, dass die Finanzierung seitens des Kantons Uri gesichert sei. Bussmann informierte Püntener am 5. Januar 1993, dass die Expertengruppe des NFP 27 dem Beitragsgesuch des Kantons Uri zugestimmt habe. Ecoplan wurde am 22. Januar 1993 durch das Bauamt Uri über diesen Entscheid informiert, mit der Ausarbeitung einer definitiven Offerte beauftragt und zur Erarbeitung eines Katalogs zum Datenbedarf aufgefordert. Der unterzeichnete Vertrag über die Mitfinanzierung durch das NFP 27 wurde am 25. Januar 1993 vom Bauamt Uri der Programmleitung des NFP 27 zugesandt. An einer Besprechung zwischen Ecoplan und Edoardo Frei (Leiter der Hauptabteilung Strassen beim Bauamt Uri) vom 11. Februar 1993 lagen die definitive Offerte und der Katalog zum Datenbedarf vor. Einige Tage später, am 16. Februar 1993, erteilte das Bauamt Uri offiziell den Auftrag an Ecoplan zur Durchführung der Studie gemäss der Offerte. Am 1. März 1993 trafen sich Ecoplan und Vertreter des Kantons zu einer Besprechung, an der vor allem über den Datenbedarf diskutiert wurde.

Im März 1993 nahm Ecoplan die Arbeiten für die Studie auf, wobei sich die Evaluatoren vorerst schwerpunktmässig mit organisatorischen Fragen und der Datenbeschaffung beschäftigten. Am 12. Mai trafen sich Ecoplan und der Kantonsingenieur zu einer Kurzbesprechung, an der die Evaluatoren über den Stand der Arbeiten informierten. In den Monaten Juni, Juli und August 1993 haben die Evaluatoren mit verschiedenen Personen Gespräche durchgeführt. Unter den Gesprächspartnern fanden sich Vertreter der ernerischen Verwaltung (Bau-, Landwirtschafts- und Volkswirtschaftsdirektion), des Bundesamtes für Statistik (BfS), des Eidgenössischen Instituts für Schnee- und Lawinenforschung, der

Gemeinde Spiringen (UR), der Verkehrsvereine Urserental und Meiringen-Haslital (BE) und eines Transportunternehmers. Die Gespräche dienten einerseits der Beschaffung von Informationen und Daten, andererseits aber auch zur Erfassung der unterschiedlichen Interessenlagen der Gesprächspartner bzw. der von ihnen vertretenen Akteure. Die Datenbeschaffung erwies sich zum Teil als nicht ganz einfach. So mussten die Evaluatoren aus Datenschutzgründen für die Sekundärnutzung des beim Kantonalen Amt für Industrie, Gewerbe und Arbeit (KIGA) bereits vorliegenden Datenmaterials zu den Umsatzzahlen der Gastwirtschaftsbetriebe eigens diese Betriebe anfragen, um deren Einwilligung dafür zu erhalten. Diese Umfrage fand in der zweiten Hälfte Juli 1993 statt. Im August 1993 befassten sich die Evaluatoren zur Hauptsache mit der Wirkungsanalyse, der Auswertung und der Umsetzung der Resultate in Empfehlungen. Am 3. September 1993 lag ein Entwurf zum Schlussbericht vor. Dieser wurde am 8. September 1993 mit Peter Püntener (Kantonsingenieur) und Edoardo Frei (Leiter Hauptabteilung Strassen) besprochen. Nach der Durchführung einiger kleinerer Änderungen konnte der definitive Schlussbericht am 10. September 1993 an den Auftraggeber und an die Programmleitung des NFP 27 versandt werden. Am 12. November 1993 legte Peter Püntener in einem Schreiben zuhanden des NFP 27 die Einschätzung der Studie aus der Sicht des Auftraggebers dar. Denselben Schreiben war auch ein Journal/Logbuch zur Evaluation angefügt. Am 26. November erfolgte die Rechnungsstellung durch Ecoplan an die Baudirektion Uri. Am 1. Dezember 1993 wurde dem Programmleiter des NFP 27 die Schlussrechnung des Projektes und die Rechnung zur Begleichung des Kostenbeitrages des NFP 27 übermittelt.

5.2.3. Bewertung entlang der Kriterien

Für die Evaluation der Evaluation von Heini Sommer und Stefan Suter verwenden wir die konventionellen Kriterien. Es bestehen keine Hinweise auf paradigmatische Positionen, welche eine Bewertung entlang der alternativen Kriterien nahelegen würden.

5.2.3.1. Nützlichkeit

5.2.3.1.1. Identifikation der beteiligten Gruppen

Im Rahmen der Evaluation haben die Forscher mit Vertretern folgender Institutionen Gespräche durchgeführt:

- *Baudirektion, Kanton Uri*
 - Kantonsingenieur, Bauamt
 - Leiter Hauptabteilung Strassen, Bauamt
 - Betriebsleiter Kantonsstrassen (verantwortlich für Schneeräumung), Bauamt
 - Chef betriebliches Rechnungswesen, Bauamt
 - Leiter Verkehrswesen, Bauamt
 - Sekretariat, Bauamt
- *Übrige Verwaltungsstellen, Kanton Uri*
 - KIGA, Volkswirtschaftsdirektion
 - Land- und Forstwirtschaftsdirektion
- *Bundesstellen:*
 - Bundesamt für Statistik
 - Eidg. Institut für Schnee- und Lawinenforschung
- *Gemeinden:*
 - Gemeinde Spiringen
- *Fremdenverkehr:*
 - Verkehrsverein Andermatt
 - Verkehrsverein Meiringen-Haslital
- *Transporteure:*
 - Transportunternehmung in Spiringen

Weiterhin haben die Evaluatoren eine Befragung bei den Urner Gastwirtschaftsbetrieben durchgeführt, um bei diesen die Erlaubnis zur Verwendung des bereits früher vom KIGA für andere Zwecke erhobenen Datenmaterials zu erhalten.

Entgegen den Angaben in der Offerte (EK: 4) wurde für diese Evaluation keine Begleitgruppe konstituiert. Die Studie ist durch drei Mitarbeiter des Bauamts Uri begleitet worden. Eine eigentliche Begleitgruppe, die auch Vertreter der Verkehrsvereine oder

betroffener Regionen umfasst hätte, wurde aber nicht etabliert. Die Offerte sah vor, dass der Auftraggeber für die Bildung der Begleitgruppe besorgt ist (EK: 4). Die Gründe für den Verzicht sind uns nicht bekannt (vgl. SB: 13).

In der Diskussion um das Zielsystem einer frühen Passöffnung äussern sich die Autoren zu dieser Frage wie folgt: "Obwohl diese Ziele [also etwa die Tourismusförderung] wie bereits erwähnt nicht schriftlich fixiert sind, kann davon ausgegangen werden, dass damit die Anliegen der wichtigsten Interessenskreise abgedeckt sind." (SB: 15, Ergänzungen TW) Die Programmziele haben die Autoren aus mündlichen Angaben der zuständigen Personen in der kantonalen Verwaltung abgeleitet (vgl. SB: 14).

Insgesamt betrachtet ist festzuhalten, dass die Studie sicherlich einen Grossteil der Interessen in irgendeiner Form in den Evaluationsprozess zu integrieren vermochte. Es verbleiben jedoch einige Parteien, welche am evaluierten Programm beteiligt oder von diesem betroffen sind, deren Interessen in der Studie zumindest nicht explizit zum Ausdruck kommen. Beispielsweise betrifft dies ökologische Gruppierungen oder die privaten Unternehmen, welche im Auftrag des Kantons einen Teil der Schneeräumung übernehmen. Durch die Beschränkung des Untersuchungsgebietes auf den Kanton Uri besteht die Gefahr, dass die Interessen der angrenzenden Kantone zu wenig berücksichtigt werden. Da vier der fünf Urner Pässe die Kantonsgrenze überschreiten und auch der fünfte Pass, der Klausen, die Verbindung in den Kanton Glarus gewährleistet, kann dies zu einer einseitig ernerischen Perspektive führen, die gesamthaft betrachtet wenig zweckmässig erscheint.

Inwiefern die Evaluation die Interessen berücksichtigt hat, ist ohne Gespräche mit den relevanten Akteuren nicht zu entscheiden. Erschwerend kommt hinzu, dass aufgrund des Schlussberichts stellenweise zu wenig klar wird, welche Akteure welche Interessen vertreten. Ein Zitat aus dem Schlussbericht soll dies verdeutlichen: "Von verschiedener Seite wird bezweifelt, ob diese Wirkungen im erhofften Ausmass eintreten und ob der Aufwand ökonomisch vertretbar ist." (SB: K-1, vgl. auch SB: 1) Es finden sich aber auch Passagen im Schlussbericht, in welchen die unterschiedlichen Interessenlagen der Akteure deutlich herausgearbeitet werden (vgl. SB: 15, 17).

Aufgrund des doch reichhaltigen Einbezugs von Interessen im Kanton Uri und unter Berücksichtigung der knappen Ressourcen, die für die Evaluation zur Verfügung standen, kommen wir zum Schluss, dass die Evaluation dieses Kriterium erfüllt.

5.2.3.1.2. Glaubwürdigkeit des Evaluators

Der hauptsächlich an dieser Studie beteiligte Evaluator, Heini Sommer, bringt in das Projekt verschiedene Kompetenzen ein. Er kennt, da er in Altdorf wohnt und arbeitet, den regionalen Kontext gut. Aus dem Studium bringt er weitgehende ökonomische Kenntnisse mit. Seine Dissertation befasst sich mit finanziellen Aspekten von Bau, Unterhalt und Betrieb der Urner Nationalstrassen - wozu übrigens auch die Gotthardpassstrasse gehört. Die Doktorarbeit wurde mit finanzieller Unterstützung des Kantons Uri realisiert (Sommer 1990). Sommer arbeitet seit 1989 bei Ecoplan und hat in dieser Zeit an einer Vielzahl verschiedener Forschungsaufträge gearbeitet. Die Forschungsprojekte von Ecoplan, die er leitete oder an denen er massgeblich beteiligt war, befassen sich schwerge- wichtig mit Wirtschafts- und Finanzpolitik, Verkehr, Energie und mit der Wasserversorgung bzw. der Abwasserentsorgung. Unter den Auftraggebern dieser Projekte finden sich verschiedene Bundesstellen, Kantone und Gemeinden (vgl. Ecoplan 1994). Daneben sind auch Studien im Auftrag des Nationalen Forschungsprogramms 'Stadt und Verkehr' (NFP 25) erstellt worden (vgl. Sommer/Neuenschwander/Walter 1993a und 1993b, Neuenschwander/Suter/Walter/Sommer 1992a&b, Neuenschwander/Sommer/Suter/Walter 1992). Stefan Suter, der auch an der Studie mitgewirkt hat, ist seit 1990 bei Ecoplan tätig, und hat seither einige Projekte bearbeitet (so bspw. Neuenschwander/Sommer/Suter/Walter 1992, Neuenschwander/Suter/Walter/Sommer 1992a&b, Walter/Suter/Nieuwkoop 1993). Zudem leitete er im Jahre 1993 ein anderes Projekt, das im Rahmen des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) als 'Kurzevaluation' mitfinanziert wurde (Suter/Walter 1993).

Aufgrund dieser Bemerkungen kommen wir zum Schluss, dass die Mitglieder des Forschungsteams zur Bearbeitung der hier zur Diskussion stehenden Evaluation sicherlich ausreichende Erfahrung und ausgewiesene Expertise einbringen und dass aufgrund unserer Kenntnisse keinerlei Veranlassung besteht, deren Glaubwürdigkeit in Frage zu stellen. Das Kriterium wird erfüllt.

5.2.3.1.3. Informationsrahmen und -auswahl

Im Rahmen der Studie haben die Evaluatoren verschiedene Informationsquellen genutzt.

- *Gespräche*: Zur Erhebung von Informationen und Interessen sind mit Personen aus verschiedensten Institutionen Gespräche geführt worden. Unter den Gesprächspartnern finden sich Vertreter der kantonalen Verwaltung Uri, von Bundesstellen, Urner Gemeinden, Verkehrsvereinen, Gastbetrieben und Transportunternehmern (für eine ausführliche Liste der Institutionen vgl. oben, Kapitel 5.2.3.1.1.).

- *Quantitatives Datenmaterial:* Zur Durchführung der Evaluation verwendeten die Autoren statistische Angaben aus verschiedensten Quellen und unterzogen diese einer Sekundäranalyse. Zentrale Daten lieferte das Bauamt des Kantons Uri, das Bundesamt für Statistik und das KIGA Uri.
- *Dokumente/Literatur:* Die Evaluatoren stützen ihre Argumentation verschiedentlich auf die im sog. Quellenverzeichnis aufgeführte Literatur ab. Dies vor allem in zwei Teilen der Studie. Zum einen betrifft dies die gewählte methodische Vorgehensweise, die Ecoplan im Auftrag der Finanzdirektion des Kantons Bern für die Erfolgskontrolle von Subventionen entwickelt hat (Neuenschwander/Sommer/Walter 1990, vgl. SB: 3-8). Zweitens ziehen die Evaluatoren zur Legitimierung ihrer Vorgehensweise bei der Abschätzung der Einflüsse auf den Tourismusbereich verschiedentlich Fachliteratur aus dem Bereich der ökonomischen Inzidenzanalyse (Dürsteler 1984, Häusel 1985, Masuhr 1988, Ecoplan 1989, vgl. SB: 36-40) und der Tourismusforschung bei (Rütter 1990, vgl. SB: 35 und 36, bes. Anm. 13 und 16). Der Hinweis im Quellenverzeichnis auf ein Ökonometrie-Lehrbuch von Andrew C. Harvey wirkt eher etwas als Fremdkörper (siehe SB: 57, vgl. Harvey 1990).

Mittels dieser Informationsauswahl erhalten die Evaluatoren ausreichende Informationen zur Behandlung sachdienlicher Fragen zum Evaluandum. Sicherlich gäbe es neben den erfassten Informationen weitere Informationsbedürfnisse, die jedoch im Rahmen dieser Evaluation nicht befriedigt werden konnten. Darauf weisen auch die Evaluatoren im Schlussbericht deutlich hin (so SB: 11, 13, 30, 35 und 46-47). Thematisch liegen die Felder für zusätzliche Erhebungen in den Bereichen Nebenwirkungen (Umweltbelastung durch Verkehrsaufkommen und Schneeräumung, direkte Beschäftigungs- und Einkommenseffekte durch die Schneeräumung etc.), Verkehrsfrequenzen auf den Passstrassen, Interessen Betroffener in den angrenzenden Kantonen (zur Region Haslital vgl. SB: 52), Sparpotential einer Verschiebung der Passöffnung und bei intervenierenden Faktoren (wie Wetterverhältnisse, Schneehöhen etc., vgl. SB: 35).

Die Interessen der Beteiligten und Betroffenen im Kanton Uri können aufgrund der beschafften Informationen zwar nicht vollständig (weitere Orte, Tourismusbranche neben Gastgewerbe etc.), aber hinreichend berücksichtigt werden. Eine Ausweitung der Informationsbasis, insbesondere auch auf die relevanten Regionen jenseits der Kantons-grenzen, wäre wünschenswert (vgl. dazu Kapitel 5.2.3.1.1.).

In Anbetracht der knappen zeitlichen und finanziellen Ressourcen, die zur Erarbeitung dieser Evaluation zur Verfügung standen, können wir jedoch das Kriterium als erfüllt betrachten.

5.2.3.1.4. Bewertende Interpretationen

Die durch die Evaluatoren eingenommenen Perspektiven sind im Schlussbericht ausreichend klar dargelegt worden. Bei den Verfahren gilt dies im Grundsatz ebenso, wobei sich hier aber einige Einschränkungen aufdrängen:

- Das Vorgehen bei der Befragung der Gastbetriebe ist etwas verwirrend, da sich die Angaben auf unterschiedliche Grundgesamtheiten beziehen. Zum einen wird die Zahl der vom KIGA erfassten Urner Gastbetriebe angeführt, auf der die Erhebung der Umsatzzahlen basiert. Andererseits werden jedoch auch Angaben von der Polizeidirektion Uri und der massgeblichen Verkehrsvereine benutzt. Dies erschwert es dem Leser, die Übersicht zu bewahren. Dies ist insbesondere deshalb der Fall, weil im Bericht keine Definition der unterschiedlichen Grundgesamtheiten angeführt wird. Dies hätte klarer dargestellt werden können. (vgl. dazu auch Kapitel 5.2.3.4.3.)
- Die Durchführung der Gespräche mit den Exponenten der unterschiedlichen Institutionen wird in der Studie zu wenig erläutert. Es fehlen systematische Angaben zur Ausgestaltung der Gespräche, zur Auswahl der Gesprächspartner usw. Es finden sich lediglich vereinzelte Angaben im Bericht, die jedoch nicht ausreichen, um dem Leser eine klare Vorstellung über die Vorgehensweise in diesem Bereich zu vermitteln (vgl. dazu Kapitel 5.2.3.4.9).
- Im Schlussbericht entstehen aufgrund von knappen Erläuterungen verschiedentlich kleinere Unklarheiten oder Unsicherheiten.

Bezüglich der Begründungen erhalten wir aufgrund der Angaben im Schlussbericht grundsätzlich ebenfalls ein positives Bild. Etwas problematisch erscheint uns jener Teil der Studie, welcher sich mit den Auswirkungen auf den Tourismus befasst. Hier fehlen verschiedentlich Begründungen, welche dem Leser den Nachvollzug des Vorgehens der Evaluatoren erleichtern würden. Aspekte, welche genauer zu begründen wären, sind zum Beispiel die folgenden:

- Warum ist hier der Wertschöpfungsausfall relevant (siehe SB: 31-40)?
- Warum wird im Fall der Gemeinden Hospental und Realp auf die Logiernächte in den Monaten Mai, Juni und Juli abgestützt (vgl. SB: 36), währenddem im Falle von Andermatt lediglich die Zahlen zu den Monaten Mai und Juni beigezogen werden (vgl. SB: 34, vgl. Kapitel 5.2.3.4.3.)?
- Warum werden lediglich die Umsatzzahlen des Gastgewerbes in die Analyse miteinbezogen und andere Betriebe (Bahnen, Autogewerbe, Detailhandel, Sportanlagen, Parahotellerie etc.) lediglich als Zulieferer der Gastbetriebe berücksichtigt (SB: 32-40)?

- Warum kann aufgrund der Angaben zur verspäteten Öffnung des Sustenpasses im Jahre 1989 (Felssturz) auf Umsatzausfälle bei anderen Pässen geschlossen werden (SB: 37-39, vgl. dazu ausführlicher Kapitel 5.2.3.4.6.)?
- Wie wurden die Orte ausgewählt, welche bei einer Verschiebung der Passöffnung möglicherweise einen Umsatzverlust im Gastgewerbe in Kauf zu nehmen hätten (SB: 32-40)?

Trotz diesen Einschränkungen kommen wir jedoch zum Schluss, dass dieses Kriterium erfüllt ist. Der Schlussbericht stellt die Grundlagen für die Werturteile grösstenteils ausreichend klar dar.

5.2.3.1.5. Klarheit des Berichts

Die Studie gliedert sich folgendermassen:

Kurzfassung (SB: K-1-K-11)

Inhaltsverzeichnis (SB: I-III)

1. Einleitung (SB: 1-2)

1.1. Umfeld (SB: 1)

1.2. Auftrag und Zielsetzung (SB: 1-2)

1.3. Projektorganisation (SB: 2)

2. Methodischer Ansatz (SB: 3-8)

2.1. Evaluationen - was ist darunter zu verstehen? (SB: 3)

2.2. Evaluationsmethode (SB: 3-8)

2.3. Kurzevaluation (SB: 8)

3. Kurzevaluation der Schneeräumung (SB: 9-54)

3.1. Vorarbeiten (SB: 9-13)

3.2. Zieldefinition (SB: 14-17)

3.3. Vollzugskontrolle (SB: 18-26)

3.4. Wirkungsanalyse (SB: 27-47)

3.5. Auswertung und Umsetzungsvorschläge (SB: 48-54)

4. Methodische Auswertung (SB: 55)

Quellenverzeichnis (SB: 57-58)

Der Bericht verfügt über eine Inhaltsübersicht (SB: I[a]) und ein Inhaltsverzeichnis (SB: I-III). Eine Kurzfassung mit einem Umfang von elf Seiten bildet einen integralen Bestandteil der Studie und ist zur Unterscheidung von den übrigen Teilen des Berichts auf gelbes Papier gedruckt (SB: K-1-K-11). Die Kurzfassung ermöglicht dem eiligen Leser eine gute Übersicht über das Projekt. Die Zusammenfassung gibt den wesentlichen Inhalt der ausführlichen Fassung in geraffter Form wieder. Lediglich an einer Stelle weicht die

Kurzfassung inhaltlich von den Aussagen im Hauptbericht ab. Bei der Auflistung der vier Hauptwirkungen des zu evaluierenden Programms fehlt die Zieldimension 'Förderung sozialer Kontakte' (siehe dazu K-5 und SB: 14, 29, 46, 48 sowie 53). Der Bericht ist in einer gut verständlichen, klaren Sprache verfasst und zweckmässig strukturiert. Sprachliche Mängel finden sich im Bericht nur wenige, wobei sich in den Anmerkungen eine gewisse Häufung zeigt. Die Evaluatoren verwenden graphische Darstellungen zur Verdeutlichung und Vermittlung der Inhalte. Die Graphiken sind zweckmässig gestaltet und sinnvoll eingesetzt. Zur Beschreibung der Absichten und Verfahren haben wir uns schon im vorangehenden Kapitel ausführlich geäussert (siehe Kapitel 5.2.3.1.4.). Die Ergebnisse der Evaluation sind klar formuliert, wobei sich bei deren Begründung einige inhaltliche Probleme ergeben, die wir jedoch an anderer Stelle diskutieren werden (siehe Kapitel 5.2.3.4.10.).

Insgesamt betrachtet erhalten wir aufgrund der angeführten Aussagen einen sehr positiven Eindruck. Das Kriterium wird erfüllt.

5.2.3.1.6. Verbreitung des Berichts

Der Schlussbericht zum betrachteten Evaluationsprojekt ist vom Auftragnehmer nach Abschluss der Studie auftragsgemäss an den Auftraggeber (Bauamt Uri) und an die Programmleitung des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) gesandt worden.

Die Baudirektion Uri hat die Studie im Zusammenhang mit der Vernehmlassung zu einem 'Öffnungs- und Schliessungskonzept der Urner Alpenpässe' (Bauamt 1994) im Mai 1994 als Beilage den Baudirektionen der angrenzenden Kantone (Bern, Glarus, Graubünden, Tessin und Wallis) zukommen lassen. Inwiefern die kantonale Baudirektion die Studie weiteren Stellen innerhalb und ausserhalb des Kantons Uri weitergegeben hat, entzieht sich unseren Kenntnissen. Der Kanton Uri hat sich - gemäss den uns vorliegenden Informationen - zwar nicht aktiv darum bemüht, die Studie weiteren Kreisen zugänglich zu machen, hingegen ist der Bericht aber auch keineswegs unter Verschluss gehalten worden.

Der Programmleiter des NFP 27, Werner Bussmann, hat den Schlussbericht umgehend an die Mitglieder der Expertengruppe des NFP 27 weitergeleitet.

Zur Berichterstattung über die Studie in den Medien liegen uns keine Hinweise vor.

Die Bewertung der Verbreitung des Berichts kann aufgrund der vorliegenden Informationen nicht abschliessend erfolgen. Aufgrund unseres Wissensstandes erfüllt diese Evaluation das Kriterium mehrheitlich.

5.2.3.1.7. Rechtzeitigkeit des Berichts

Gemäss dem Gesuch um Mitfinanzierung der Kurzevaluation durch das Nationale Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen' (NFP 27) sollte sich die Projektdauer von anfangs Februar bis Ende Juli 1993 erstrecken (GES/OF: 4). In der definitiven Offerte (EK: 4) wurde die Projektdauer unverändert übernommen. Ecoplan hat am 16. Februar 1993 den offiziellen Auftrag zur Durchführung der Studie erhalten und die eigentlichen Projektarbeiten anfangs März 1993, also einen Monat später als vorgesehen, aufgenommen. Am 3. September 1993 lag ein Entwurf zum Schlussbericht vor. Eine Woche später wurde der definitive Schlussbericht versandt. Unter Berücksichtigung des verspäteten Projektbeginns haben die Auftragnehmer somit rund zehn Tage mehr für die Untersuchung benötigt. Diese Verspätung ist trotz der vergleichsweise kurzen Projektdauer von sechs Monaten als geringfügig zu werten. Dies gilt auch unter Berücksichtigung der von der Sache her gegebenen terminlichen Restriktionen. Die Erkenntnisse aus der Studie lagen rechtzeitig vor, um für die Schneeräumung im Frühling 1994 handlungsrelevant werden zu können (vgl. Kapitel 5.2.3.1.8.).

Ziehen wir zur Beurteilung der Rechtzeitigkeit des Berichts die erfolgten Umsetzungsmassnahmen bei, so zeigt sich auch hier, dass der Abschluss der Studie nicht zu spät erfolgt ist. Das (unter anderem) aufgrund der Studie vom Bauamt Uri erstellte Konzept über die Öffnung und Schliessung der Urner Passstrassen lag im März 1994 vor (Bauamt 1994). Die Vernehmlassung dieses Konzepts bei den angrenzenden Kantonen wurde am 18. Mai 1994 gestartet.

Hinsichtlich der Bedürfnisse des NFP 27 zeigt sich aufgrund der Ausschreibungsunterlagen, dass ein Abschluss der Kurzevaluationen spätestens auf den 1. August 1993 gefordert wurde. Da die Auftragserteilung durch den Kanton Uri an das Büro Ecoplan erst Mitte Februar 1993 erfolgte, kann diese Verzögerung den Evaluatoren jedoch nicht angelastet werden.

Aufgrund der gemachten Aussagen kommen wir zum Schluss, dass die Berichterstattung durch die Forscher nur mit einer unwesentlichen Verzögerung erfolgte und dass damit dieses Kriterium als erfüllt gelten kann.

5.2.3.1.8. Wirkung der Evaluation

Zur Abschätzung der Wirkung der Evaluation betrachten wir zunächst einmal, inwiefern die in der Studie enthaltenen Empfehlungen Eingang in ein Konzept gefunden haben, welches durch das Bauamt Uri erarbeitet wurde. Im Konzept wird explizit die Studie von Ecoplan als Grundlage bezeichnet (Bauamt 1994: 3). Das Konzept bezieht sich sowohl auf die Öffnung der Passstrassen im Frühling, wie auch auf deren Schliessung im Herbst. Zum Konzept wurde eine Vernehmlassung bei den angrenzenden Kantonen durchgeführt. Ein Entscheid des Regierungsrates des Kantons Uri zum Konzept steht noch aus.

- Die Evaluatoren empfehlen aufgrund der vorgenommenen Analysen, dass durch die politischen Entscheidungsträger ein klar definiertes *Zielsystem* auszuarbeiten sei. Dieses soll um Soll-Werte ergänzt werden, damit eine angemessene Erfolgsüberprüfung möglich wird (SB: 49). Das nach Abschluss der Evaluation durch das Bauamt Uri ausgearbeitete Konzept enthält jene Faktoren, welche zu berücksichtigen sind. Diese werden - im Gegensatz zur Studie - für jeden Pass separat festgelegt und sind als Bedürfnisse bezeichnet. Den verschiedenen Aspekten sind auch sogenannte Wunschanforderungen beigefügt, welche als Soll-Werte interpretiert werden können.
- Die Evaluatoren halten im Schlussbericht zur *Vollzugsproblematik* folgendes fest: "Der administrative und technische Vollzug der Schneeräumung funktioniert einwandfrei." (SB: 49) Aufgrund dieser Erkenntnisse besteht kein Handlungsbedarf und es erübrigen sich auch diesbezügliche Aussagen im Konzept bzw. deren Vergleich.
- Die Evaluatoren fordern, dass die Festlegung der Öffnungsdaten vermehrt aufgrund von Kostenüberlegungen getroffen werden sollten. Neben den Schnee- und Witterungsverhältnissen und der Sicherheit sollte bei der Entscheidungsfindung vermehrt ein *Kostenbewusstsein* zum Tragen kommen (vgl. SB: 49-50). Dieser Aspekt findet sich im Konzept des Bauamts wieder. Dort heisst es: "Klar definierte Vorgaben sollen eine kosten- und ressourcenoptimierte Planung und Ausführung durch das Bauamt ermöglichen ..." (Bauamt 1994: 3).
- Der Schlussbericht macht klare Aussagen zur Frage, zu welchem *Zeitpunkt* die Urner Passstrassen im Frühling geöffnet werden sollten. Die Evaluatoren empfehlen, zur Terminierung der Schneeräumung vermehrt auf Kostenaspekte Rücksicht zu nehmen. Aufgrund ihrer Analysen kommen sie zum Schluss, dass für alle Pässe - mit Ausnahme des Klausenpasses - ein Öffnungsdatum anzustreben sei, welches im Vergleich zum langjährigen Mittel rund zwei Wochen später festgesetzt werden sollte. Weiterhin soll - in Abkehr zur bisherigen Praxis - für jeden Pass ein anzustrebendes, fixes Öffnungsdatum festgelegt werden (SB: 50-54). Das Konzept des Bauamts folgt den Emp-

fehlungen in der Evaluation praktisch vollständig (vgl. Bauamt 1994: 4, 6, 8, 10 und 12). Eine Abweichung ergibt sich lediglich dadurch, dass die Termine in der Studie als anzustrebende Fixdaten gemeint sind, im Gegensatz dazu jedoch das Konzept die Daten als frühest mögliche Öffnungstermine definiert.

Da das Konzept in der Vernehmlassung durch die angrenzenden Kantone teilweise heftiger Kritik ausgesetzt war und weil dieses vom Urner Regierungsrat bisher nicht verabschiedet wurde, müssen für eine definitive Bewertung der Wirksamkeit die weiteren Ereignisse abgewartet werden. Doch schon zum jetzigen Zeitpunkt kann festgehalten werden, dass die Empfehlungen der Evaluation im Umsetzungsprozess eine wesentliche Rolle gespielt haben.

Zur Ergänzung der vorangehenden Betrachtungen scheint es zweckmässig zu verfolgen, ob die Folgerungen der Evaluatoren bereits in der Vollzugspraxis aufgenommen wurden. Uns liegen hierzu jedoch lediglich Angaben zu den Passöffnungsdaten im Frühling 1994 vor. Diese deuten darauf hin, dass die Evaluationsempfehlungen bisher nicht vollzugsrelevant geworden sind. Alle Pässe wurden deutlich früher geöffnet, als dies durch die Studie vorgeschlagen wurde. Mit Ausnahme der Furka erfolgte die Öffnung im Frühling 1994 vor dem langjährigen Durchschnittstermin des entsprechenden Passes.

Obwohl in der Vollzugspraxis bisher keine Verhaltensänderungen festgestellt werden konnten, scheint die Evaluation bei der Ausarbeitung des Konzepts des Bauamts (1994) grosse Beachtung erhalten zu haben. Obwohl die Erkenntnisse der Evaluation seitens angrenzender Kantone auf teilweise massiven Widerstand stossen (vgl. Kapitel 5.2.3.2.2., 5.2.3.3.2., 5.2.3.4.2. und 5.2.3.4.11.), hat die Evaluation erreicht, dass die verantwortlichen Stellen beim Kanton Uri bisher der Evaluation gefolgt sind. Deshalb kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt. Inwiefern die Evaluation damit neue Aspekte erschlossen hat, oder ob die Evaluation lediglich bereits vorhandene Pläne bestätigte, kann aufgrund der uns zugänglichen Dokumente und der wenigen Gespräche nicht festgestellt werden. Auf alle Fälle sind die im Evaluationsbericht formulierten Empfehlungen ausreichend praxisnah und auch praxisfähig, so dass ihnen gefolgt werden kann.

5.2.3.2. Anwendbarkeit

5.2.3.2.1. Praktische Verfahren

Aufgrund der knapp bemessenen zeitlichen und finanziellen Ressourcen zur Durchführung der Evaluation waren die Auftragnehmer gezwungen, bei der Datenerhebung und Auswertung einen pragmatischen Ansatz zu wählen. Dies führte einerseits dazu, dass einige der durchgeführten Analysen methodische Probleme aufwerfen, auf die wir an anderer Stelle noch zu sprechen kommen werden (siehe unten, Kapitel 5.2.3.4.8., 5.2.3.4.9.). Andererseits wurde aus denselben Gründen auch auf die Überprüfung einiger Grundannahmen verzichtet, auf denen die Evaluation ihre Argumentation aufbaut. So wird nicht überprüft, ob und in welchem Ausmass durch eine Verschiebung der Passöffnung um einige Wochen eine Kostenersparnis eintreten würde. Problematisch ist weiter die Annahme der Autoren, dass die Einflüsse der Passöffnungen auf das Tourismusgewerbe in den untersuchten Regionen miteinander vergleichbar seien.

Die Datenerhebung durch die Evaluatoren im Rahmen der vorgegebenen Restriktionen ist aufgrund der uns zur Verfügung stehenden Informationen zweckmässig erfolgt. Es wurden zwar im Projektverlauf auch Datenquellen konsultiert, die für die Analyse nicht verwendet werden konnten (zum Beispiel die Verkehrsstatistik oder die Daten zu den Schneehöhen). Dies war jedoch nicht voraussehbar und damit auch nicht zu vermeiden.

Bei der Datenanalyse sind verschiedene Aspekte hervorzuheben:

- Die durchgeführten Regressionsanalysen zur Untersuchung des Zusammenhangs zwischen dem Zeitpunkt der Passöffnung und den Übernachtungszahlen in den ausgewählten Orten erweisen sich als wenig robust (vgl. Kapitel 5.2.3.4.8.).
- Der Vergleich der Kosten der Schneeräumung mit dem Wertschöpfungsausfall im Gastgewerbe der konsultierten Gemeinden ist aus mehreren Gründen in Frage zu stellen:
 - a) Auswirkungen auf andere Gemeinden im Kanton Uri und in angrenzenden Kantonen sind nicht auszuschliessen.
 - b) Zur Schätzung des Wertschöpfungsausfalls sind nicht nur die Gastbetriebe zu berücksichtigen, sondern auch weitere Branchen.
 - c) Die Hochrechnungen der Umsatzzahlen basieren teilweise auf sehr wenigen Fällen.

- d) Aufgrund des erfolgten Vergleichs lassen sich die in der Evaluation formulierten Empfehlungen zur Terminierung der Passöffnungen nicht rechtfertigen. Es wurde nicht abgeklärt, welche Kosteneinsparungen bei der Schneeräumung eintreten würden, falls diese beispielsweise um zwei Wochen verschoben würde (vgl. dazu Kapitel 5.2.3.4.10. und 5.2.5.4.).

Aufgrund dieser Feststellungen muss auch dieser Analyseteil als störungsanfällig betrachtet werden.

Es zeigt sich bei der Bewertung dieses Kriteriums, dass die Evaluation - unter anderem aufgrund der vorgegebenen Restriktionen - gewisse Mängel aufweist. Die Evaluation erfüllt das Kriterium trotzdem teilweise, da wir die Ansicht vertreten, dass die Auftragnehmer - unter Berücksichtigung der Rahmenbedingungen - bezüglich der Praktikabilität die Möglichkeiten zu einem massgeblichen Teil ausgeschöpft haben.

5.2.3.2.2. Politische Überlebensfähigkeit

Wir haben bereits an anderer Stelle auf den zur Zeit hängigen Umsetzungsprozess im Nachgang zu dieser Evaluation hingewiesen (siehe Kapitel 5.2.3.1.8.). Aus den Beobachtungen lässt sich bezüglich der in der Evaluation enthaltenen Empfehlungen festhalten,

- dass sie bei der zuständigen Baudirektion des Kantons Uri auf hohe Akzeptanz sties-
sen und
- dass sie bei den zur Vernehmlassung eingeladenen, angrenzenden Kantonen sehr unterschiedlich aufgenommen wurden. Die Stellungnahmen der Kantone enthalten Reaktionen, welche von einem völligen Einverständnis bis zur totalen Ablehnung reichen.

Wie sich der Regierungsrat des Kantons Uri zur Sache stellen wird, bleibt abzuwarten. Aufgrund der uns vorliegenden Informationen ist eine Einschätzung der Einstellungen der beteiligten und/oder betroffenen Kreise im Kanton Uri nicht zu bewerkstelligen. Zur Bearbeitung dieser Frage wären weitergehende Abklärungen erforderlich. Es ist jedoch davon auszugehen, dass auch hier divergierende Meinungen vertreten werden.

Es stellt sich aufgrund dieser Feststellungen die Frage, inwiefern im Rahmen der Evaluation die Möglichkeit bestanden hätte, die Akzeptanz der Studie zu verbessern. Bezüglich der opponierenden Kantone liegt es auf der Hand, dass für den Einbezug dieser Interessen in den Evaluationsprozess das Untersuchungsgebiet über die Urner Kantonsgrenzen hin-

weg hätte ausgedehnt werden müssen. Dieser Ausweitung standen natürlich die beschränkten Mittel der Evaluation und möglicherweise auch die Interessen des Auftraggebers entgegen. Andererseits muss jedoch auch festgehalten werden, dass eine lediglich auf den Kanton Uri begrenzte Untersuchungsanlage im Widerspruch zur grenzüberschreitenden Charakteristik der Urner Alpenübergänge steht. Bedenkenswert erscheint uns die Frage, ob nicht eine vertieftere Untersuchung ausgewählter Pässe (unter Einbezug der angrenzenden Regionen ausserhalb des Kantons Uri) dem effektiven Vorgehen vorzuziehen gewesen wäre. Gemäss dem Evaluationskontrakt war eigentlich auch eine Beschränkung der zu untersuchenden Pässe "auf zwei bis maximal drei" (EK: 2) vorgesehen (vgl. Kapitel 5.2.3.3.1.). Durch die realisierte Ausgestaltung des Untersuchungsdesigns setzt sich die Evaluation dem Vorwurf aus, die Thematik lediglich aus Urner Perspektive abgehandelt zu haben. Dies kann die politische Überlebensfähigkeit der Aussagen in der Studie (ausserhalb des Kantons) gefährden.

Für eine abschliessende Bewertung dieses Kriteriums sind weitere Abklärungen notwendig. Insbesondere wäre dabei zu untersuchen, wie es zur Definition des Untersuchungsgebietes gekommen ist, wer diesen Entscheid getroffen hat (Auftraggeber und/oder Evaluator) und welches die Gründe dafür waren. Zudem ist der Fortgang des Entscheidungsprozesses zum Konzept des Urner Bauamts abzuwarten. Wir verzichten deshalb an dieser Stelle auf eine Bewertung des Kriteriums.

5.2.3.2.3. Kostenwirksamkeit

Die Durchführung der Evaluation hat 39'973.- sFr. gekostet. Die Betreuungsleistungen durch den Kanton Uri halten sich im üblichen Rahmen. Das Nationale Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen' (NFP 27) hat im Rahmen des Projektes 'Erprobung von Kurzevaluationen' einen Kostenanteil von 50% übernommen. Die hier betrachtete Studie ist damit im Vergleich mit den übrigen untersuchten Evaluationen zu einem sehr tiefen Preis erarbeitet worden.

Die Wirkung der Evaluation haben wir bereits an anderer Stelle ausführlich diskutiert (vgl. Kapitel 5.2.3.1.8.). Aufgrund eines Vergleichs der Evaluationsergebnisse mit dem 'Öffnungs- und Schliessungskonzept der Urner Alpenpässe' (Bauamt 1994) zeigt sich, dass die durch die Evaluation produzierten Informationen zumindest in der Baudirektion Uri als nützlich erachtet wurden.

Aufgrund dieser Feststellungen ist die Kostenwirksamkeit bei dieser Studie gegeben. Das Kriterium wird erfüllt.

5.2.3.3. Korrektheit

5.2.3.3.1. Formale Verpflichtung

Zur Beurteilung dieses Kriteriums ziehen wir die definitive Offerte (EK) bei, welche durch Ecoplan im Februar 1993 erstellt worden ist. Mit der offiziellen Auftragserteilung durch das Bauamt des Kantons Uri an Ecoplan vom 16. Februar 1993 übernahm die Offerte die Rolle eines Evaluationskontraktes.

Finanzen: Im Evaluationskontrakt wird festgehalten, dass der effektive Zeitaufwand verrechnet wird. Weiter wird ein Kostendach von 40'000.- sFr. vereinbart (EK: 5). Dieser Kostenrahmen wird von den Evaluatoren eingehalten. Nach Abschluss des Projektes wird von Ecoplan ein Betrag von 39'973.- sFr. in Rechnung gestellt. Die finanziellen Verpflichtungen sind von der Evaluation eingehalten worden.

Zeit: Der Evaluationskontrakt legte den zeitlichen Rahmen folgendermassen fest: Im Februar 1993 sollen die Projektarbeiten aufgenommen und im Juli 1993 abgeschlossen werden (siehe EK: 4). Wie wir bereits an anderer Stelle festgehalten haben (vgl. Kapitel 5.2.3.1.7.), erteilte das Bauamt des Kantons Uri den offiziellen Auftrag zur Durchführung der Evaluation erst am 16. Februar 1993, worauf die Evaluatoren die Projektarbeiten am 1. März 1993 in Angriff nahmen. Am 3. September 1993 konnten die Evaluatoren dem Auftraggeber einen Entwurf zum Schlussbericht präsentieren, der nur noch geringfügigen Änderungen unterzogen werden musste, um am 10. September 1993 als definitiver Schlussbericht vorzuliegen. Damit haben die Evaluatoren zur Bearbeitung der Studie einige Tage mehr benötigt, als dies vereinbart worden ist. Hauptsächlich ist jedoch die Verzögerung der Schlussberichterstattung auf die späte Auftragserteilung zurückzuführen. Die in der Evaluation entstandene Verzögerung halten wir für geringfügig.

Personen: Im Evaluationskontrakt wird festgehalten, dass das Projekt seitens Ecoplan von Heini Sommer geleitet und voraussichtlich (EK: 4) von Felix Walter und Stefan Suter bearbeitet werden wird. Effektiv an den Projektarbeiten beteiligt waren Heini Sommer, der den Grossteil der Arbeiten geleistet hat, und Stefan Suter. Felix Walter (Jahrgang 1963, lic. rer. pol., bei Ecoplan seit 1988) hat an diesem Projekt nicht mitgewirkt. Da im Evaluationskontrakt lediglich von einer voraussichtlichen Mitarbeit von Felix Walter gesprochen wird, erachten wir diesen Vertragsbestandteil als erfüllt. Der Vertrag sieht die Bildung einer Projektbegleitgruppe durch den Auftraggeber vor. Für die Mitgliedschaft in diesem Gremium wurden im Evaluationskontrakt Edoardo Frei (Leiter Hauptabteilung Strassen, Bauamt Uri) und Peter Püntener (Urner Kantonsingenieur), sowie Vertreter von Verkehrsvereinen und von betroffenen Regionen (zum Beispiel 'Pro

Meien') vorgeschlagen. Diese Projektbegleitgruppe wurde jedoch - aus uns nicht bekannten Gründen - gar nie gebildet. Stattdessen wurde das Projekt von drei Mitarbeitern des Urner Bauamts begleitet. Gesamthaft gesehen stellen wir fest, dass bis auf die fehlende Begleitgruppe die personellen Verpflichtungen eingehalten wurden.

Inhalt: Wir untersuchen die Einhaltung anhand der im Evaluationskontrakt festgelegten Fragestellung der Untersuchung (vgl. EK: 2):

- *"Welche übergeordneten Ziele werden mit Eröffnung der Passstrassen verfolgt?"* (EK: 2) Diese Frage wird in der Evaluation behandelt (siehe SB: 14-17). Da keine schriftlich festgehaltene Zielsetzung zur Schneeräumung bestand, musste diese "aus den mündlichen Angaben der zuständigen Personen in der kantonalen Verwaltung" (SB: 14) abgeleitet werden. An anderer Stelle im Schlussbericht wird darauf hingewiesen, dass das Zielsystem "aus den Angaben der verschiedenen Interessengruppen" (SB: 17) entwickelt wurde. Es wurde für jeden Pass ein individuelles Zielsystem erstellt.
- *"Inwieweit sind die Ziele operationalisiert (Messgrössen, Sollwerte)?"* (EK: 2) Die Evaluation behandelt diese Frage (siehe SB: 17).
- *"Wie hoch sind die gesamten Kosten der Schneeräumung?"* (EK: 2) Die Kosten der Schneeräumung werden durch die Evaluation untersucht (siehe SB: 23-26).
- *"Was sind die Wirkungen einer zeitlich verspäteten Öffnung der Passstrassen (z.B. Umsatzausfall im Tourismusgewerbe pro Tag)?"* (EK: 2) Die Evaluation geht auf diese Fragestellung ein (vgl. SB: 27-47), wobei sie sich dabei auf die Hauptwirkungen (Tourismus, Alpwirtschaft, Unterhalt Gotthard-Strassentunnel, Soziale Kontakte) beschränkt und allfällige Nebenwirkung (Verkehrsaufkommen, Umwelt, Unterhalt Passstrassen, Beschäftigung und Einkommen durch die Schneeräumung, Staatshaushalt) nur am Rande berücksichtigt.
- *"Werden die Ziele tatsächlich erreicht?"* (EK: 2) Da für die Schneeräumung auf Urner Passstrassen keine eigentlichen Ziele in der Form von Soll-Werten oder Zielvorgaben in den Wirkungsdimensionen vorliegen, erübrigt sich eine Untersuchung des Zielerreichungsgrades (vgl. SB: 17 und 53).
- *"Wie sieht das Verhältnis zwischen Aufwand und Ertrag aus?"* (EK: 2) Die Evaluation untersucht, inwiefern die Wirkungen die für die Schneeräumung getätigten Aufwendungen rechtfertigen.
- *"Welche Änderungen sind vorzunehmen (z.B. bei der Definition der Ziele, beim Vollzug, usw.)?"* (EK: 2) Der Bericht enthält Änderungsvorschläge (SB: 48-54).

- "*Könnten die Ziele auch mit anderen Massnahmen erreicht werden?*" (EK: 2) Die Evaluatoren verzichteten aufgrund der knappen Mittel, die für die Evaluation zur Verfügung standen, auf eine "Berücksichtigung und Beurteilung von alternativen Massnahmen" (SB: 13). Diese Frage wird somit nicht bearbeitet.

Neben der Fragestellung ist bezüglich des Inhalts auch die Abgrenzung des Untersuchungsgegenstandes von Interesse. Der Evaluationskontrakt legt fest, dass von den insgesamt fünf Urnerpässen lediglich "zwei bis maximal drei" (EK: 2) untersucht werden sollen. Im Verlaufe des Forschungsprozesses haben die Evaluatoren jedoch den Entscheid gefällt, alle fünf Pässe in die Untersuchung miteinzubeziehen. Weiterhin wird im Vertrag festgehalten, dass der Untersuchungszeitraum auf "die Jahre 1988 bis 1993" (EK: 2) festgelegt werden soll. "Im Lauf der Arbeiten hat sich gezeigt, dass dieser Zeitraum zu knapp bemessen ist. Er wurde daher je nach Verfügbarkeit der Daten bis ins Jahr 1975 bzw. 1961 erweitert." (SB: 13)

Zusammenfassend lässt sich festhalten, dass der Untersuchungsgegenstand im Verlaufe der Untersuchung zeitlich wie räumlich ausgedehnt wurde. Die im Evaluationskontrakt formulierte Fragestellung wurde weitgehend behandelt.

Methodik/Vorgehen: Die Evaluatoren folgen sowohl im Evaluationskontrakt wie auch in der Untersuchung selbst einem methodischen Ansatz, der an anderer Stelle entwickelt worden ist (vgl. Neuenschwander/Sommer/Walter 1990: 20-73, vgl. dazu auch oben, Kapitel 5.2.1.3.). Zu den einzelnen Arbeitsschritten ergeben sich folgende Bemerkungen:

- *Zieldefinition:* Die Evaluatoren leiten aufgrund von Äusserungen von Akteuren aus der Verwaltung (und möglicherweise auch von interessierten Kreisen, vgl. oben 'Inhalt') ein Zielsystem ab. Entgegen den Angaben im Evaluationskontrakt (EK: 2) verzichten sie jedoch auf die Entwicklung von messbaren Zielgrössen.
- *Vollzugskontrolle:* Die Evaluatoren untersuchen den Vollzug, wenn auch nur in knapper Form (vgl. dazu Kapitel 5.2.3.3.7., 5.2.3.4.1., 5.2.3.4.10. und 5.2.3.4.11.). Die wichtigsten, im Evaluationskontrakt angegebenen Daten (EK: 2) werden erhoben.
- *Wirkungsanalyse:* Die Evaluation behandelt die im Evaluationskontrakt festgehaltenen Aspekte weitgehend (vgl. EK: 3). Es bestehen jedoch einzelne Lücken. Beispielsweise fehlt in der Studie eine Antwort auf die Frage, wie gross die Einsparungen bei einer späteren Passöffnung sein könnten. Die angegebenen Verfahren zur Datenerhebung (Befragungen und Recherchen) werden in der Evaluation eingesetzt. Auch die Verfahren zur Ermittlung der Beschäftigungs- und Einkommenseffekte (Inzidenzanalyse und

Kennzahlen zur Wertschöpfung) kommen, wie im Kontrakt festgehalten, zur Anwendung.

- *Auswertung*: Die im Evaluationskontrakt für diesen Arbeitsschritt vorgesehenen Fragen (vgl. EK: 3) konnten in der Evaluation nur teilweise untersucht werden. Mangels festgelegter Soll-Werte erübrigte sich eine Analyse des Zielerreichungsgrades. Aus finanziellen Gründen wurde auf die Analyse von Nebenwirkungen weitgehend verzichtet.
- *Umsetzung*: Die Umsetzung der Ergebnisse ist, soweit dies im Evaluationskontrakt festgehalten wurde (vgl. EK: 3-4), durch die Evaluation angestrebt worden. Die Evaluation enthält Vorschläge zur Umsetzung der Evaluationsergebnisse. Auch eine methodische Auswertung ist im Schlussbericht enthalten (siehe SB: 55).

Es bleibt an dieser Stelle anzufügen, dass die methodischen Festlegungen im Evaluationskontrakt relativ offen formuliert wurden. Diese wurden in der Evaluation praktisch vollständig erfüllt.

Aus diesen Überlegungen heraus kommen wir zum Schluss, dass die Evaluation die formalen Verpflichtungen praktisch gänzlich erfüllt hat. Die festgestellten Abweichungen sind von geringfügiger Bedeutung. Das Kriterium wird deshalb erfüllt.

5.2.3.3.2. Deklaration von Interessenkonflikten

Aufgrund der Angaben im Schlussbericht bestehen bezüglich der Schneeräumung auf Urner Passstrassen vor allem Konflikte zum Ausmass der Auswirkungen sowie zu deren Einschätzung und Gewichtung. Verschiedentlich finden sich im Schlussbericht Textstellen, die deutlich auf die Diskrepanzen der Interessenlagen hinweisen. Beispielsweise wird hervorgehoben, dass die kantonale Verwaltung und der Verkehrsverein Andermatt unterschiedliche Gewichtungen der verschiedenen Zielsetzungen der Schneeräumung am Gotthard-Pass vornehmen (SB: 15, 17 und 48). Auch die von den Untersuchungsergebnissen abweichenden Auffassungen der Interessenvertreter im Haslital (BE) werden im Bericht - wenn auch nur sehr knapp - angeführt (siehe SB: 52).

Im Gegensatz dazu finden sich im Schlussbericht jedoch auch Stellen, wo die Evaluatoren die vorhandenen Interessenkonflikte nur in einer Form anführen, die es dem Leser verunmöglicht, eine Verbindung zu den zugehörigen Akteuren herzustellen. Dies betrifft insbesondere folgende Aussage in der Einleitung des Schlussberichts: "Von verschiedener Seite werden diese Punkte zwar nicht grundsätzlich in Frage gestellt. Es wird jedoch bezweifelt, ob die Wirkungen tatsächlich im erhofften Ausmass eintreten und ob der da-

mit verbundene Aufwand ökologisch und ökonomisch vertretbar ist." (SB: 1) Der Evaluationsbericht enthält keine weiteren Präzisierungen, wer diese Einwände äussert.

Weil wir unsere Analyse lediglich aufgrund der uns vorliegenden Dokumente und kurzen Gespräche vornehmen, ist die Frage, inwiefern weitere, im Schlussbericht nicht erwähnte Interessenkonflikte bestehen, schwer auszumachen. Hinweise bestehen dahingehend, dass nicht nur im Kanton Bern, sondern auch in angrenzenden Regionen anderer Kantone Widerstände bezüglich der Evaluation und ihrer Vorgehensweise bestehen. Weiter ist zu vermuten, dass möglicherweise Interessen weiterer Akteure existieren, welche in der Evaluation keine Berücksichtigung fanden. Wir denken hier beispielsweise an die Bedürfnisse des Militärs (Zugang zu Truppenübungsplätzen usw.) oder des öffentlichen Verkehrs (Postautobetriebe). Weiter ist zu beachten, dass sich eine frühe Passöffnung nicht nur positiv sondern auch negativ auswirken kann. Die Evaluatoren weisen in der Studie deutlich darauf hin, dass Schneeräumung und Passöffnung Nebenwirkungen verursachen, die auch negativer Art sein können (siehe SB: 29-30). Die Evaluatoren verzichten aufgrund der knappen Ressourcen auf eine eingehende Untersuchung dieser Nebenwirkungen (SB: 30). Daraus ergibt sich jedoch das Problem, dass die damit tangierten Interessen in der Studie nicht weiter berücksichtigt werden. Wir denken dabei etwa an die Interessen des Skitourismus' (eine frühe Passöffnung könnte beispielsweise die Skisaison im Skigebiet Nätschen-Oberalp tangieren) oder jene ökologisch orientierter Kreise (eine Erhöhung des Verkehrsaufkommens beim motorisierten Individualverkehr wird kaum Begeisterung auslösen). Durch eine spätere Öffnung der Passstrassen wiederum könnten sich auch Umsatzeinbussen bei jenen Privatunternehmern ergeben, welche bis anhin Teile der Schneeräumungsarbeiten im Auftrag der kantonalen Verwaltung erledigten. Der Bericht enthält zwar "erste grobe Anhaltspunkte" (SB: 46) zur Einschätzung dieser Nebenwirkungen (siehe SB: 29-30 und 46-47). Jene Kreise, welche hinter diesen Interessen stehen, werden jedoch im Evaluationsprozess nicht oder nur am Rande berücksichtigt. Das führt in der Folge dazu, dass die in diesem Zusammenhang auftretenden Interessenkonflikte in der Studie nicht thematisiert werden.

Weil die uns vorliegenden Unterlagen nicht ausreichen, um eine Bewertung dieses Kriteriums vorzunehmen, werden wir darauf verzichten. Zur weiteren Klärung dieser Thematik wäre die Durchführung von weiteren Interviews mit relevanten Akteuren unumgänglich.

5.2.3.3.3. Vollständige und aufrichtige Offenlegung

Der Schlussbericht zeichnet sich generell durch eine hohe Transparenz aus. Die Evaluatoren legen in der Abfassung des Berichts hohes Gewicht auf eine vollständige und aufrichtige Offenlegung. Von diesem, sehr positiven Eindruck weichen nur einige wenige Aspekte ab, auf die wir an dieser Stelle kurz eintreten möchten.

- Die Evaluatoren verzichten im Schlussbericht darauf, jene Datenerhebungen darzulegen, welche im Evaluationsprozess durchgeführt worden sind, sich jedoch nachträglich für die Evaluation als wenig zweckmässig erwiesen haben (zum Beispiel Verkehrsaufkommen oder Schneehöhen auf den Passstrassen). Dadurch wird zwar der Schlussbericht von diesen Um- oder Irrwegen im Forschungsprozess entschlackt. Es wird damit aber auch in Kauf genommen, dass die Darstellung des Evaluationsprozesses an Transparenz verliert.
- Bei der Berechnung, Schätzung und Hochrechnung der monatlichen Umsatzzahlen im Gastgewerbe legen die Evaluatoren lediglich die Ergebnisse für den Ort Andermatt ausführlich dar (vgl. Grafik 3-16, SB: 33). Für die anderen untersuchten Ortschaften, wird auf eine Darstellung des Jahresgangs der Umsatzzahlen verzichtet.
- Das Vorgehen bei der Erhebung qualitativen Datenmaterials - wir denken hier insbesondere an die durchgeführten Gespräche - ist aufgrund der Angaben im Schlussbericht nicht nachvollziehbar. Die Autoren verzichten auf Angaben zur Durchführung der Datenerhebung wie auch der Datenauswertung (Gesprächspartner, Interviewtyp, Datenerfassung, Auswertungsverfahren etc.) (vgl. Kapitel 5.2.3.4.3., 5.2.3.4.4. und 5.2.3.4.9.).

Verschiedentlich weisen die Autoren auf die Grenzen der Aussagekraft der Evaluationsergebnisse hin. So enthält der Schlussbericht deutliche Anmerkungen zur beschränkten Aussagekraft der Studie hinsichtlich der Nebenwirkungen der Schneeräumung (SB: 46). Auch zur Abschätzung des Wertschöpfungsausfalls im Gastgewerbe (SB: 35), zu bestimmten Datenmaterialien (SB: 30) oder bestimmten Datenauswertungen (SB: 35) finden sich Hinweise dieser Art. Darüberhinaus enthält die Studie im Kapitel, das sich mit der methodischen Auswertung der Kurzevaluation befasst, die folgende Aussage: "Die Wirkungen staatlicher Massnahmen können auch mit einem beschränkten Budget einer *ersten groben Beurteilung* unterzogen werden." (SB: 55, ähnlich auch K-11 und 13, Hervorhebungen TW)

Obwohl sich bei der Offenlegung einige punktuelle Mängel gezeigt haben, ergibt sich gesamthaft gesehen ein durchaus positives Bild. Die Evaluation berücksichtigt dieses Kriterium in ausreichendem Masse.

5.2.3.3.4. Herstellung von Öffentlichkeit

Der Schlussbericht ermöglicht auch einer mit der Materie wenig vertrauten Leserschaft ein gutes Verständnis (vgl. Kapitel 5.2.3.1.5.).

Der Bericht ist unseres Wissens öffentlich zugänglich. Inwiefern der Bericht im Kanton Uri eine weite Verbreitung gefunden hat, ist uns nicht bekannt.

Die Baudirektion Kanton Uri hat bei den angrenzenden Kantonen Bern, Glarus, Graubünden, Tessin und Wallis ein Vernehmlassungsverfahren zu dem auf der Grundlage der Evaluationsstudie erarbeiteten 'Öffnungs- und Schliessungskonzept der Urner Alpenpässe' (Bauamt 1994) durchgeführt. Die für die Passstrassen zuständigen Stellen dieser Kantone haben in diesem Zusammenhang auch den Schlussbericht der Studie erhalten.

Der Bericht ging zudem an die Programmleitung und die Expertengruppe des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27), welches die Studie mit einem Kostenbeitrag unterstützt hat.

Auch wenn der Auftraggeber unseres Wissens bisher darauf verzichtet hat, die Publizität der Studie aktiv zu fördern, sind wir der Ansicht, dass die Evaluation das Kriterium erfüllt.

5.2.3.3.5. Schutz individueller Menschenrechte

Verstösse gegen individuelle Menschenrechte sind uns im Zusammenhang mit dieser Evaluation nicht bekannt geworden. Das Kriterium wird erfüllt.

5.2.3.3.6. Menschliche Interaktionen

Zu diesem Kriterium liegen uns kaum Informationen vor. Da wir uns hier lediglich auf eine Dokumentenanalyse stützen, fehlen uns Hinweise, welche durch die Führung von Interviews mit den durch die Evaluatoren kontaktierten Personen gewonnen werden könnten.

Der Schlussbericht zur Studie enthält kaum Hinweise, wie das qualitative Datenmaterial erhoben wurde. Besonders im Bereich der Gespräche bestehen verschiedentlich Informationslücken. Der Schlussbericht erwähnt folgende Personen, mit denen die Forscher im Evaluationsprozess in Kontakt getreten sind:

- Edoardo Frei, Bauamt Uri (Projektbegleitung) (SB: 2, 15 Anm. 2),
- Josef Brand, Bauamt Uri (Projektbegleitung) (SB: 2, 15 Anm. 2, 19 Anm. 3,),
- Josef Zurfluh, Bauamt Uri (Projektbegleitung) (SB: 2) und
- A. Arnold, Land- und Forstwirtschaftsdirektion Uri (Inform. zur Alpwirtschaft) (SB: 42 Anm. 31)

Weitere Personen werden im Bericht nicht namentlich genannt. Aufgrund der Angaben im Logbuch zum Projekt wurden mit einer Vielzahl weiterer Personen im Zusammenhang mit dem Projekt Interviews durchgeführt. Es stellt sich nun die Frage, wie die Evaluatoren jene Personen zur namentlichen Nennung im Schlussbericht ausgewählt haben. Während dies bei den drei Vertretern des Bauamtes wohl daran lag, dass diese in die Projektbegleitung involviert waren, ist dies für den Mitarbeiter der Land- und Forstwirtschaftsdirektion weniger klar. Die Auswahl erscheint eher zufällig erfolgt zu sein.

Weil die Informationen zu diesem Kriterium nur sehr spärlich vorhanden sind, verzichten wir hier auf eine Bewertung. Dies würde weitere Abklärungen erforderlich machen.

5.2.3.3.7. Ausgewogene Berichterstattung

Die Darstellung des Evaluandums erfolgt aus einer Perspektive, die stark von ernerischen Überlegungen geprägt ist. Dies steht im Zusammenhang mit der räumlichen Abgrenzung des Untersuchungsgebietes. Bis auf eine Ausnahme (Verkehrsverein Meiringen-Haslital) beschränken sich die Ausführungen im Schlussbericht auf den Kanton Uri. Die Evaluatoren weisen im Schlussbericht auch mehrfach darauf hin, dass die Bewertung aus ernerischer Sicht erfolgt (siehe etwa SB: 17 und 52). An einer Stelle findet sich auch der Hinweis, dass ausserhalb des Kantons Uri eine frühe Passöffnung höhere Priorität zu haben scheint und dass daher interkantonale Absprachen nötig seien (SB: 52).

Bezüglich der generellen Erkenntnisse der Studie, welche sich nicht auf die einzelnen Pässe beziehen, fällt vor allem der Umstand auf, dass der Vollzug der Schneeräumung als unproblematisch bezeichnet wird. Die Autoren schreiben: "Der administrative und technische Vollzug der Schneeräumung funktioniert einwandfrei." (SB: 49) Aufgrund der Hinweise im Kapitel 'Vollzugskontrolle' (SB: 18-26) entsteht jedoch ein etwas anderes Bild. Erstens wiesen die Autoren dort darauf hin, dass eine Überprüfung der Informationspra-

xis angebracht ist, dass kein ausgeprägtes Kostenbewusstsein bestehe und dass eine Vergabe der Schneeräumungsarbeiten an private Räumungsunternehmen geprüft werden sollte. Während die Anmerkungen zum Kostenbewusstsein und zur privaten Vergabe in der Auswertung wieder aufgenommen werden (vgl. SB: 49-50), fehlt dort der Hinweis zur Informationspraxis. Weiterhin führen die Angaben im Kapitel zum Vollzug zu verschiedenen Fragen, die der Schlussbericht offen lässt:

- Wie sind die doch erheblichen Kostenunterschiede für die Schneeräumung auf den fünf Pässen zu erklären? (vgl. SB: 23-25)
- Warum liegt der Lohnanteil bei den Schneeräumungskosten am Klausen und Susten fast dreissig Prozent höher, als bei den übrigen Pässen (SB: 25)?
- Wie erfolgt die Vergabe von Schneeräumungsarbeiten an private Unternehmen? Weshalb werden am Klausen und Susten teilweise private Geräte eingesetzt, bei den anderen Pässen jedoch nicht (vgl. SB: 22, vgl. zu den Felsräumungsarbeiten auch SB: 18).
- Wie lange dauern durchschnittlich die Schneeräumungsarbeiten? Bestehen diesbezüglich grosse Schwankungen?
- Wie werden die Räumungsarbeiten des Kantons mit jenen des Militärs (Furka und Oberalp) koordiniert (vgl. SB: 25)?
- Warum wurde im Jahr 1993 um rund eine Woche später mit den Schneeräumungsarbeiten begonnen (siehe SB: 23, Anm. 6; vgl. dazu Kapitel 5.2.3.4.5. und 5.2.5.4.)?

Aufgrund dieser Fragen wird deutlich, dass sich die Vollzugsanalyse - so wie sie im Schlussbericht dargestellt ist - aufgrund der knappen Evaluationsressourcen auf die wesentlichsten Aspekte beschränken musste. Da die Vollzugsanalyse nur knapp erfolgen konnte, hätte sich unseres Erachtens zur Einschätzung, dass der Vollzug einwandfrei funktioniere, ein Vorbehalt aufgedrängt (vgl. Kapitel 5.2.3.4.10).

Im übrigen enthält der Bericht eine ausgewogene Berichterstattung. Wir betrachten vor allem den ersten (die ernerische Perspektive) der angeführten kritischen Aspekte als erheblich. Wir kommen so zur Ansicht, dass die Evaluation dieses Kriterium nur mit Einschränkungen zu erfüllen vermag.

5.2.3.3.8. Finanzielle Verantwortlichkeit

Die Evaluation musste mit einem sehr engen Zeit- und Finanzbudget auskommen. Die Zuweisung der Mittel ist - soweit uns Informationen vorliegen - klug und verantwortlich erfolgt. Es kam zwar punktuell zu Nachforschungen und Analysebestrebungen, die sich

im Nachhinein als nicht zweckmässig erwiesen. Da dies jedoch nicht vorhersehbar war und zudem zumeist bei Evaluationen geschieht, ist dies der Evaluation in keiner Weise anzulasten. Das Kriterium wird erfüllt.

5.2.3.4. Genauigkeit

5.2.3.4.1. Identifikation des Gegenstandes

Die Evaluatoren grenzen den Untersuchungsbereich klar ab. Neben der Identifikation der fünf zu untersuchenden Urner Pässe (SB: 1), enthält der Bericht ebenso eine Definition des Untersuchungszeitraums (SB: 13), wie auch die thematische Abgrenzung, dass lediglich die Schneeräumungsarbeiten zur Öffnung der Passstrassen im Frühling untersucht werden sollen (SB: 9). Weiterhin halten die Autoren fest, dass die Zweckmässigkeit der Zielsetzungen des Evaluandums nicht in Frage gestellt werden soll (SB: 11). Zudem weisen die Autoren deutlich darauf hin, dass kein Vergleich mit alternativen Massnahmen durchgeführt wird (SB: 13) und dass auf eine eingehende Untersuchung der Nebenwirkungen im Rahmen der Kurzevaluation verzichtet werden muss (SB: 30 und 46). Zusammenfassend können wir festhalten, dass bei dieser Evaluation eine klare Umschreibung des Untersuchungsgegenstandes vorliegt.

Inwiefern die Evaluation den Gegenstand ausreichend untersucht hat, kann nur unter Berücksichtigung der zeitlichen und finanziellen Ressourcen, die für die Untersuchung zur Verfügung standen, erfolgen. Da diese ja, wie bereits erwähnt, sehr knapp bemessen waren, ist der Umstand, dass die Evaluation in bestimmten Bereichen sicher eingehender hätte erfolgen können, nicht allzu schwer zu gewichten. Trotzdem muss an dieser Stelle darauf hingewiesen werden, dass unter anderem in folgenden Aspekten eine weitergehende Analyse sehr wünschbar wäre:

- Ausweitung des Untersuchungsbereichs auf die angrenzenden Gebiete ausserhalb des Kantons Uri (vgl. Kapitel 5.2.3.1.1.),
- Ausweitung des Untersuchungsbereichs auf weitere Ortschaften im Kanton Uri, die ja potentiell auch Einflüssen des Evaluandums ausgesetzt sind (vgl. Kapitel 5.2.3.1.3.),
- Ausweitung der Fragestellung auf die Nebenwirkungen (SB: 30 und 46), die ja gemäss den Aussagen der Autoren zumindest teilweise nicht vernachlässigt werden sollten (siehe SB: 46-47),
- vertiefte Untersuchung der Wirkungszusammenhänge und des Vollzugs (vgl. Kapitel 5.2.3.3.7.),

- Evaluation der Auswirkungen im Tourismusbereich über das eigentliche Gastgewerbe hinaus (vgl. Kapitel 5.2.3.1.3.) und
- detailliertere Analyse des erhobenen Datenmaterials (vgl. insbesondere Kapitel 5.2.3.4.8., siehe auch Kapitel 5.2.5.4.).

Gesamthaft kommen wir zum Ergebnis, dass die Evaluation das Kriterium erfüllt, wobei deutlich darauf hinzuweisen ist, dass die Studie im Vergleich zu einer umfangreicheren Evaluation sicherlich deutlich weniger an analytischer Tiefe aufzuweisen hat.

5.2.3.4.2. Kontextanalyse

Der Bericht enthält einige Informationen zum Kontext des Evaluandums. Zu Beginn der Studie wird das Umfeld der Evaluation erläutert (SB: 1). Dort wird auch darauf hingewiesen, dass um den Zeitpunkt der Schneeräumung und um die in diesem Zusammenhang entstehenden Kosten divergierende Ansichten bestehen. Aus den Hinweisen der Evaluatoren wird klar, dass die Studie den Zweck verfolgt, die Urner Baudirektion beziehungsweise das Bauamt, verantwortlich für die Schneeräumung und gleichzeitig Auftraggeber der Studie, mit Grundlagen auszustatten, um die Entscheidungsprozesse zu verbessern. Weniger offensichtlich ist jedoch, wer die unterschiedlichen Interessen vertritt. Dazu liegen lediglich punktuelle Informationen vor. So etwa der Hinweis, dass Widerstände gegen eine allzu späte Öffnung der Passstrassen aus Tourismuskreisen im Kanton Uri, aber auch im Kanton Bern (Meiringen-Haslital) bestehen. Wer jedoch für eine möglichst späte Öffnung oder für eine Reduktion der Kosten eintritt, erwähnt der Bericht nicht. Das Kapitel des Evaluationsberichts zur Zieldefinition enthält zwar einige Hinweise zur Interessenlage einiger Akteure, eine tiefergehende, systematische Analyse der Interessen fehlt jedoch (vgl. dazu auch Kapitel 5.2.3.1.1.).

Wir haben schon verschiedentlich darauf hingewiesen, dass die Regionen ausserhalb des Kantons Uri, welche im Einzugsgebiet der Pässe liegen, ebenfalls in die Untersuchung hätten miteinbezogen werden müssen. Die Passstrassen bilden Infrastruktureinrichtungen, die eine interkantonale Funktion wahrnehmen (vgl. dazu Kapitel 5.2.3.1.1., 5.2.3.1.3. und 5.2.3.3.7.). Zudem wurden sie mit erheblicher finanzieller Unterstützung des Bundes (über 70-prozentiger Bundesanteil) gebaut bzw. ausgebaut.

Andererseits enthält die Studie einige Passagen, die Aspekte im Kontext des Evaluandums sehr gezielt und zweckmässig aufgreifen. Wir denken dabei an die Ausführungen zu den Nebenwirkungen des Evaluandums, aber auch zu den Angebots- und Nachfragefaktoren, welche das Tourismusaufkommen beeinflussen können (SB: 31-32). Auch

die Ausführungen zu den Auswirkungen auf die Alpwirtschaft sind an dieser Stelle hervorzuheben (SB: 41-44). Sicherlich kann die in sehr kurzer Zeit entstandene und mit nur knappen Mitteln ausgestattete Kurzevaluation nicht jene Informationsdichte aufweisen, wie dies bei einer 'Langevaluation' möglich wäre. Unter Berücksichtigung dieser Rahmenbedingung sind wir aber der Ansicht, dass der Kontext ausreichend untersucht worden ist und dass deshalb das Kriterium erfüllt wird.

5.2.3.4.3. Beschreibung von Zielen und Vorgehen

Der Schlussbericht zur Evaluation enthält eine Beschreibung zur Zielsetzung des Projektes, aufgrund derer auch eine explizite Fragestellung entwickelt wird (SB: 1-2, vgl. dazu auch SB: 9-13).

Bezüglich des in der Evaluation gewählten Vorgehens sind differenziertere Betrachtungen anzustellen. Wir unterscheiden im folgenden die Methodik im Sinne von Untersuchungskonzept oder Ansatz von einer Methodik in einem verfahrenstechnischen Sinn.

Auf der Ebene des methodischen Ansatzes (oder Zuganges) wird das Vorgehen im Schlussbericht sehr deutlich dargelegt. Der Evaluationsansatz richtet sich nach einem Konzept, das an anderer Stelle entwickelt wurde (siehe Neuenschwander/Sommer/Walter 1990: 20-73, zusammengefasst in SB: 3-8) und nun in dieser Evaluation an die vorliegende Thematik, Fragestellung und an die zur Verfügung stehenden Evaluationsmittel adaptiert wird. Die Ausführungen zum Evaluationsansatz reichen aus, um diesen identifizieren und einschätzen zu können.

Bezüglich der verwendeten Verfahren und Techniken finden sich im Schlussbericht einige Angaben, die jedoch in den empirischen Teil des Textes integriert und daher nicht so leicht zugänglich sind, wie die Aussagen zur Konzeption der Studie (SB: 3-8, vgl. oben). Die Hinweise vermögen die verwendeten Verfahren nicht immer ausreichend zu beschreiben. Besonders deutlich zeigt sich dieser Mangel bei der Datenerhebung durch Interviews. Aufgrund der Angaben lässt sich zwar folgern, dass mit verschiedenen Akteuren Gespräche durchgeführt wurden. Es verbleibt jedoch eine Vielzahl von offenen Fragen zu Form und Inhalt der Gespräche, zu den Gesprächspartnern und deren Auswahl, zur Gesprächslänge etc. In den übrigen Untersuchungsteilen reichen die Angaben im Schlussbericht zumeist aus, um das Vorgehen der Evaluatoren für den Leser in groben Zügen nachvollziehbar zu machen. In Detailspekten bleiben jedoch verschiedene Fragen offen. Hierzu einige Beispiele:

- In der Studie werden die Kosten der Schneeräumung aufgeführt. Es werden dazu auch einige Erläuterungen angeführt. Unklar bleibt jedoch die Frage, ob es sich dabei um Vollkosten handelt. (SB: 23-25)
- Bei der Untersuchung der Auswirkungen der Passöffnung auf das Gastgewerbe wird im Falle von Andermatt auf die Monate Mai und Juni Bezug genommen (SB: 34), bei den Gemeinden Hospental und Realp hingegen auf die Monate Mai, Juni und Juli (SB: 36). Erläuterungen dazu fehlen (vgl. auch Kapitel 5.2.3.1.4.).
- Für die Untersuchung der Auswirkungen auf die Gastbetriebe in den betrachteten Ortschaften stützen sich die Evaluatoren erstens auf Angaben aus einer vom KIGA Uri durchgeführten Umfrage bei den Gastbetrieben (SB: 32, Anm. 5), zweitens auf Angaben der Polizeidirektion Uri und der lokalen Verkehrsvereine (SB: 32, Anm. 4) und drittens auf Zahlen aus den vom Bundesamt für Statistik durchgeführten jährlichen Erhebungen bei schweizerischen Hotel- und Kurbetrieben (SB: 32, Anm. 6). Diese Daten beziehen sich auf unterschiedliche Grundgesamtheiten und Stichproben. Dies führt zu Unklarheiten und Missverständnissen. (vgl. dazu auch Kapitel 5.2.3.1.4.)

Da diese Fragen lediglich Detailaspekte betreffen, sind sie zur Gesamtbewertung des Kriteriums als wenig erheblich einzustufen. Die Informationslücken zur Durchführung der Interviews sind jedoch unseres Erachtens wesentlich, so dass die Evaluation das Kriterium nur mit diesem Vorbehalt erfüllt.

5.2.3.4.4. Verlässliche Informationsquellen

Die Studie enthält praktisch immer die erforderlichen Quellenangaben. Dies gilt insbesondere bei Informationen quantitativer Art. Nur in wenigen Ausnahmefällen fehlen hier die entsprechenden Angaben - so etwa für die entstehenden Kosten für eine motorisierte Alpauffahrt via Linthal auf den Urnerboden (SB: 44). Im Bereich des qualitativen Datenmaterials ist die Angabe von Informationsquellen kritischer zu bewerten. Auch bei einer Beurteilung der Verlässlichkeit der Informationsquellen fallen die im vorstehenden Kapitel angeführten Informationslücken zur Durchführung der Interviews negativ ins Gewicht (siehe Kapitel 5.2.3.4.3.). In einem Teilbereich der Studie, in dem die Ziele der zu evaluierenden Massnahme untersucht werden, enthält der Schlussbericht widersprüchliche Angaben. Weil keine Zielsetzung zur Schneeräumung in schriftlicher Form bestand, musste diese "aus den mündlichen Angaben der zuständigen Personen in der kantonalen Verwaltung" (SB: 14) abgeleitet werden. An einer anderen Stelle des Berichts wird darauf verwiesen, dass das Zielsystem "aus den Angaben der verschiedenen Interessengruppen" (SB: 17) entwickelt worden sei. Aufgrund dieser Hinweise ist es dem Leser

nicht möglich die Frage zu beantworten, ob in den entsprechenden Untersuchungsteilen die Ziele der Schneeräumung aus der Sicht der kantonalen Verwaltung dargelegt werden, oder ob dabei auch die Sichtweisen anderer Interessengruppen berücksichtigt wurden. Da die Analyse des Zielsystems für den weiteren Evaluationsprozess von grosser Relevanz ist, wäre hier grössere Sorgfalt erforderlich gewesen.

Wegen der angeführten Mängel erfüllt die Evaluation dieses Kriterium nur knapp.

5.2.3.4.5. Valide Messung

Zum qualitativen Material liegen nicht genügend Informationen vor, um dessen Validität bewerten zu können.

Die für die Evaluation erhobenen quantitativen Daten sind bezüglich ihrer Validität einer differenzierten Bewertung zu unterziehen:

- *Öffnungsdaten der Pässe:* Dazu liegt jährliches Datenmaterial für die Jahre 1961 bis 1993 für alle fünf hier zur Diskussion stehenden Pässe vor. Die Gültigkeit der Informationen ist - soweit wir dies beurteilen können - gewährleistet. Es stellt sich jedoch an dieser Stelle die Frage, ob dieses Datenmaterial nicht mit Angaben zum Beginn der Schneeräumungsarbeiten (bzw. zu deren Dauer) hätten ergänzt werden müssen. Wir möchten dies mit einem Beispiel begründen: In der Studie wird wiederholt darauf hingewiesen, dass im Frühling 1993 der Beginn der Schneeräumungsarbeiten um rund eine Woche hinausgezögert worden sei (SB: K-4, 23 (Anm. 6), 49, 49-50, 53, vgl. dazu auch Kapitel 5.2.5.4.). Andererseits zeigt sich, betrachtet man die Angaben des Bauamtes Uri zu den Passöffnungsdaten, dass in diesem Jahr die Pässe durchwegs deutlich vor dem langjährigen Mittel geöffnet werden konnten (vgl. SB: 34 und Kapitel 5.2.5.4.). Auf den ersten Blick erscheint dies nicht sonderlich problematisch und zudem durch die relativ geringen Schneehöhen in diesem Winter auch leicht erklärbar zu sein. Folgt man jedoch den Begründungen in der Studie, fällt auf, dass das Jahr 1993 für zwei verschiedene Argumentationslinien als Beispiel hinzugezogen wird. Das Jahr 1993 zeige einerseits, dass durch eine Verzögerung des Beginns der Schneeräumungsarbeiten Kosten reduziert werden können (SB: K-4, 23 (Anm. 6), 49, 49-50). Die geringeren Kosten der Schneeräumung für 1993 könnten auch im Zusammenhang stehen mit den relativ geringen Schneemengen in diesem Jahr. Andererseits belege dieses Jahr auch, dass durch eine sehr frühe Öffnung der Passstrassen die Logiernächte in den Gastbetrieben nicht hätten erhöht werden können (SB: 34). Hier stellt sich das Problem, dass davon ausgegangen werden muss, dass sich die Umsatzzahlen je nach Zeitpunkt gegenüber der Passöffnung unterschiedlich sensitiv zeigen und dass die

Zahlen für die Logiernächte von verschiedenen anderen Faktoren mitbeeinflusst werden können.

- *Kosten der Schneeräumung:* Aufgrund der Angaben im Schlussbericht wird nicht ersichtlich, ob die Angaben Vollkosten darstellen (SB: 23-25). Es werden zwar zum Datenmaterial einige Hinweise angebracht, diese machen jedoch zu wenig deutlich, wofür die angegebenen Kosten stehen (vgl. dazu Kapitel 5.2.3.4.3.). In diesem Zusammenhang wird auch darauf hingewiesen, dass die Schneeräumungskosten nicht vollständig erfasst sind, weil Herstellerfirmen von Schneefräsen zu Werbezwecken verschiedentlich grössere Streckenabschnitte gratis geräumt oder da das Militär einen Teil der Räumungsarbeiten und die damit in Zusammenhang stehenden Kosten übernommen haben (SB: 25, vgl. Kapitel 5.2.3.4.6.). Es fehlen jedoch Angaben, in welchem Umfang und in welchen Jahren es zu diesen Fremdleistungen gekommen ist. Zu den Kosten der Schneeräumung auf den verschiedenen Pässen ist weiterhin anzumerken, dass sich diese zwischen den Pässen stark unterscheiden. Erklärungen für diese Unterschiede (wie unterschiedliche Strassenlänge, Geländeverhältnisse, Mechanisierung etc.) führt der Bericht jedoch nicht an. Möglicherweise wären durch eine verfeinerte Analyse mögliche Potentiale zur Vollzugsverbesserung sichtbar geworden.
- *Daten zu Umsatz und Logiernächten im Gastgewerbe:* Die Evaluation versucht, aufgrund dieser Angaben die Auswirkungen der Terminierung der Passöffnung auf den Tourismus zu untersuchen (siehe SB: 33-40). Das für die Analyse verwendete Datenmaterial musste im Rahmen der Evaluation aufbereitet werden. Dies erfolgte mittels Schätzungen und Hochrechnungen. Die dazu verwendeten Annahmen sind im Schlussbericht grösstenteils aufgeführt und damit transparent. Auch die zum Teil recht schmale Basis für die Hochrechnungen wird in der Studie explizit angeführt. Das resultierende Datenmaterial beinhaltet aufgrund der Datenaufbereitung einen relativ grossen Unsicherheitsfaktor (vgl. Kapitel 5.2.3.4.6.). Neben dieser Problematik sind hier zwei weitere Vorbehalte anzubringen: Erstens sind die Daten nicht in der Lage, darüber Auskunft zu geben, wie gross die Umsätze bzw. die Umsatzausfälle im ganzen Tourismusbereich ausfallen. Dafür reichen die Angaben zu den Gastbetrieben alleine nicht aus. Zweitens ist unseres Erachtens die Auswahl der betrachteten Ortschaften zu überdenken. Wie die Auswahl der berücksichtigten Gemeinden erfolgte, lässt der Schlussbericht offen. Es ist a priori nicht von der Hand zu weisen, dass auch in anderen Ortschaften innerhalb des Kantons Uri Auswirkungen des Evaluandums festzustellen sind. Selbstverständlich ist davon das Urserental deutlich stärker betroffen als Göschenen, Gurnellen, Erstfeld, Bürglen oder gar Flüelen. Eine Begründung dafür, dass diese Gemeinden vom Passtourismus nicht tangiert werden, fehlt jedoch. Bei der Beurteilung der Validität dieses Datenmaterials ist mitzuberücksichtigen, dass sich die

Datenlage in diesem Fall als äusserst schwierig erwies und dass es im Rahmen dieser Evaluation aufgrund der knappen zeitlichen, wie auch finanziellen Ressourcen kaum möglich war, eine bessere Datenqualität zu erreichen. Aufgrund dieser Überlegungen sind wir der Ansicht, dass das hier zur Diskussion stehende Datenmaterial nicht ausreichend valide ist, dass unter Berücksichtigung der Rahmenbedingungen (Datenlage, Ressourcen) jedoch auch kaum mehr möglich war.

- *Daten zur Alpwirtschaft (Urnerboden):* Das Datenmaterial zu den Terminen der Alpauffahrt ist, soweit uns Informationen vorliegen, genügend valide (SB: 41-45). Die Angaben zu den Kosten einer Verschiebung der Passöffnungsdaten basieren wiederum auf Schätzungen und Hochrechnungen. Die durchgeführten Berechnungen und die benutzten Annahmen sind in der Studie bis auf eine Ausnahme ausführlich und verständlich dargelegt. Unklar bleibt, ob für Kälber, Schafe oder Ziegen auch ein Verlust an Sömmerungsbeiträgen des Bundes zu berücksichtigen ist (vgl. SB: 41 und 42). Gemäss der 'Verordnung Bewirtschaftungsbeiträge' des Bundes (SR 910.21, bes. Art. 13) werden auch für diese Sömmerungsbeiträge ausgerichtet, wobei diese deutlich kleiner sind, als jene für Kühe. Diese Ausfälle fallen kaum ins Gewicht, da sie die genannten Beträge nur geringfügig erhöhen würden.
- *Datenmaterial zu den Unterhaltsarbeiten am Gotthard-Strassentunnel:* Soweit uns Informationen dazu vorliegen, ist das Zahlenmaterial valide (SB: 24-25). Fragen wirft hier einzig die Angabe zu den Unterhaltswochen auf. Gemäss den Hinweisen im Öffnungs- und Schliessungskonzept der Urner Alpenpässe' (Bauamt 1994: 4) fallen die Unterhaltsarbeiten in die Wochen 23, 24 und 25. Im Gegensatz dazu werden in der Evaluation nur die Wochen 24 und 25 genannt (SB: 24 und 25).
- *Datenmaterial zu potentiellen Nebenwirkungen:* Die dazu angeführten Angaben sind deutlich als erste grobe Anhaltspunkte oder als grobe Schätzungen gekennzeichnet (SB: 46-47). Wir verzichten deshalb auf eine nähere Untersuchung der Gültigkeit dieser Angaben.

Zusammenfassend betrachtet, fallen vor allem die Lücken bei der Berichterstattung zum qualitativen Datenmaterial ins Gewicht. Eine Bewertung der Validität ist in diesem Bereich nicht möglich.

Bei den quantitativen Daten wiederum, bestehen in einigen Punkten Validitätsprobleme. Diese sind vor allem durch die teilweise schwierige Datenlage und die beschränkten Ressourcen der Evaluation entstanden. Positiv ist hervorzuheben, dass die Evaluatoren grösstenteils den Entstehungsprozess des quantitativen Datenmaterials transparent machen. Es wird zudem verschiedentlich auf die beschränkte Aussagekraft des in der Evaluation ge-

wonnenen Datenmaterials verwiesen. Deshalb kommen wir zum Schluss, dass die Evaluation dieses Kriterium bezüglich den quantitativen Angaben trotz den angeführten Schwächen mehrheitlich erfüllt.

5.2.3.4.6. Reliable Messung

Wie schon mehrfach erwähnt, reichen die Angaben zur qualitativen Datenerhebung nicht aus, um eine fundierte Bewertung abzugeben.

Im quantitativen Bereich fällt auf, dass sich die Evaluation - sofern das entsprechende Datenmaterial zur Verfügung stand - auf eine möglichst breite Datenbasis abstützt. Bei den Jahresdaten arbeiten die Evaluatoren mit Angaben, die bis 1961 (Termine Passöffnungen bzw. Alpauffahrt am Klausenpass, siehe SB: 19-21 bzw. 43) oder 1975 (Kosten der Schneeräumung) zurückreichen. Dies erhöht die Zuverlässigkeit der Messungen.

In der Analyse ergeben sich jedoch hinsichtlich der Reliabilität auch einige Probleme:

- *Verspätete Öffnung des Sustenpasses im Jahr 1989:* Im Jahr 1989 wurde die Öffnung des Sustenpasses durch einen Felssturz massgeblich verzögert. Die Evaluatoren gehen nun davon aus, dass die dadurch eingetretene Verzögerung vergleichbar ist mit der Situation, wenn auf eine Schneeräumung völlig verzichtet werden würde. Im Jahr 1989 wurde der Sustenpass gemäss Angaben des Bauamts Uri erst am 12. August geöffnet. Im Vergleich zum durchschnittlichen Öffnungstermin (15. Juni, SB: 20) ist im Jahr 1989 somit eine Verzögerung um 58 Tage, also um gut acht Wochen (SB: 38) eingetreten (vgl. dazu die widersprüchlichen Angaben in SB: 35, Anm. 11 ("um gut 7 Wochen"), 38 ("57 Tage") und 40 ("um 7 Wochen")). Die Evaluatoren vergleichen nun die Anzahl der Logiernächte in Wassen und Meien für das Jahr 1989 mit dem langjährigen Durchschnitt dieses Wertes und stellen dabei unter anderem eine Differenz von 9% (für die Monate Juni bis August) fest. Dieser Ausfall an Logiernächten in den beiden Ortschaften Meien und Wassen wird sodann auf die anderen untersuchten Orte übertragen. Diese Vorgehensweise ist in mehrfacher Hinsicht wenig zuverlässig. Erstens geht man dabei von der Annahme aus, dass das Jahr 1989 bezüglich Drittfaktoren (Schneemenge, Tourismusentwicklung, Meteorologie etc.) ein Durchschnittsjahr darstellt, was zumindest in Frage zu stellen ist. Beispielsweise fällt auf, dass es im Jahre 1989 bei allen anderen betrachteten Pässen zu einer vergleichsweise frühen Passöffnung gekommen ist. Zweitens wird implizit angenommen, dass die Abhängigkeit des Umsatzes im Tourismusbereich vom Passverkehr in allen untersuchten Orten gleich stark ausgeprägt ist, was aufgrund der doch sehr unterschiedlichen Rahmenbedingungen zu bezweifeln ist. Drittens verzögert sich die Passöffnung bei den fünf untersuch-

ten Pässen bei einem Verzicht auf jegliche Schneeräumungsarbeiten nicht überall um den gleichen Zeitraum (SB: 40), was die Abschätzung der potentiellen Ausfälle im Tourismusbereich weiter erschwert. Aufgrund dieser grossen Zahl an Unsicherheiten sind wir der Ansicht, dass hier die Zuverlässigkeit nicht ausreichend gegeben ist. Probleme ähnlicher Art bestehen auch bezüglich der Angaben zu den verzögerten Passöffnungen im Jahr 1993 (vgl. dazu Kapitel 5.2.3.4.5.).

- *Umsatzzahlen im Gastgewerbe:* Die hochgerechneten Umsatzzahlen der Gastbetriebe in den betrachteten Orten basieren auf Angaben aus den Jahren 1988 bis 1991, "wobei im Normalfall die Umsatz- und Übernachtungszahlen pro Gastbetrieb nur für ein einziges Jahr vorliegen" (SB: 31, Anm. 5). Die Zuverlässigkeit der Hochrechnungen, welche von dieser (schmalen) Basis ausgehen, ist nicht gegeben. Andererseits bestanden jedoch aufgrund der Datenlage und der knappen finanziellen und zeitlichen Ressourcen kaum praktikable Alternativen. Hinzu kommt, dass der Schlussbericht mehrfach Stellen enthält, die auf diese Problematik hinweisen (so SB: 35, 50).
- *Kosten der Schneeräumung:* Wir haben bereits darauf hingewiesen (in Kapitel 5.2.3.4.5.), dass die Werte zu den Schneeräumungskosten in bestimmten Jahren und für bestimmte Pässe aufgrund von Fremdleistungen (Militär, Schneefräsenhersteller) nicht die ganzen Kosten darstellen. Dies reduziert die Zuverlässigkeit des Datenmaterials.
- *Einsatz von Maschinen und Personal:* Die Angaben zu diesem Bereich sind ohne Angabe des Bezugsjahres aufgeführt (SB: 22). Es ist anzunehmen, dass der Einsatz sowohl von Personal wie auch von Maschinen von Jahr zu Jahr unterschiedlich ausfällt.

Aufgrund der vorangehenden Bemerkungen, und zwar insbesondere gestützt auf die mangelnde Zuverlässigkeit der Abschätzungen aufgrund der Verzögerung der Passöffnung am Susten im Jahre 1989, sind wir der Ansicht, dass die Evaluation im quantitativen Bereich - trotz einiger positiver Aspekte - das Kriterium der Reliabilität nicht erfüllt. Eine Bewertung der Reliabilität des qualitativen Materials ist aufgrund der lückenhaften Berichterstattung nicht möglich.

5.2.3.4.7. Systematische Datenüberprüfung

Es liegen uns keine expliziten Hinweise zu einer systematischen Datenüberprüfung vor. Es ist lediglich bekannt, dass der Schlussbericht in seiner Entwurfsform dem Auftraggeber vorgelegen hatte. Wir sind deshalb darauf angewiesen, auf indirekte Hinweise zu achten. Aufgrund des Schlussberichts kommen wir zur Feststellung, dass kaum fehler-

hafte Angaben zu beobachten sind. Zu den wenigen Ausnahmen gehören die folgenden Beispiele:

- Im Zusammenhang mit der verspäteten Öffnung des Sustenpasses im Jahre 1989 enthält der Schlussbericht unterschiedliche Angaben über den Zeitraum der Verzögerung (SB: 35, Anm. 11; 38 und 40, vgl. Kapitel 5.2.3.4.6.).
- Im Bericht finden sich widersprüchliche Angaben zur Anzahl der Gastbetriebe in Andermatt (SB: 32 und 35, Anm. 12).
- Im Kapitel mit den Untersuchungsergebnissen für den Gotthardpass wird eine falsche Aussage zur Relation zwischen dem potentiellen Wertschöpfungsausfall (126'000.- sFr.) und den Kosten für die Schneeräumung (217'000.- sFr.) gemacht. Es heisst dort: "Eine Deckung des Wertschöpfungsausfalls" ... "käme also den Staat etwa gleich teuer zu stehen wie die Aufwendungen für die Schneeräumung." (SB: 50, Auslassung TW)

Daneben finden sich lediglich einige wenige redaktionelle Mängel, wie etwa der Verweis auf ein inexistentes Kapitel (SB: 44) oder die Angabe einer falschen Jahreszahl bei einem Literaturverweis (SB: 6, Anm. 2).

Obwohl keine direkten Angaben dazu vorliegen, kommen wir zum Schluss, dass das Kriterium erfüllt ist. Die festgestellten Mängel sind rar und zudem wenig bedeutend.

5.2.3.4.8. Analyse quantitativer Informationen

Die Evaluatoren arbeiten in der Studie mit verschiedenen quantitativen Auswertungstechniken. Häufig werden die quantitativen Daten in Graphiken visualisiert. Die Forscher führen Mittelwerte und Extremwerte an, um das Datenmaterial überschaubar und vergleichbar zu machen. Zur Untersuchung des Zusammenhangs zwischen der Terminierung der Passöffnung und den Übernachtungszahlen beim Gastgewerbe setzen die Autoren bivariate Regressionsanalysen ein (siehe SB: 34-40). Nur gerade einmal verwenden die Evaluatoren einen Rangordnungsvergleich zur Analyse des Datenmaterials (SB: 34). Aufgrund dieser methodischen Vielfalt und der zumeist adäquaten Umsetzung der diversen Verfahren entsteht ein positives Bild. Es drängen sich aber trotzdem einige kritische Bemerkungen auf:

- Verschiedentlich mangelt es den quantitativen Analysen etwas an *analytischer Tiefe*. Gemeint ist damit etwa der Umstand, dass mehrfach Datenmaterial erhoben, aufbereitet und dargestellt wird, dass jedoch auf die weitergehende analytische und interpretative Durchdringung verzichtet wird. Beispiele dazu sind etwa die Bereiche Personal-

und Maschineneinsatz (SB: 22) oder die Schneeräumungskosten bei den fünf Pässen (SB: 25). Dort wird zwar festgestellt, dass die Kosten von Pass zu Pass sehr unterschiedlich ausfallen, Erklärungen dafür werden jedoch nicht gesucht. Auch wenn man die Differenzen in den Lohnkostenanteilen für die untersuchten Pässe betrachtet, sucht man vergebens nach Erklärungen. Wir vermuten, dass diese weitergehenden Analysen aufgrund des zeitlichen Drucks bei der Erarbeitung der Studie nicht erfolgt sind. (vgl. dazu auch Kapitel 5.2.3.4.1. und 5.2.5.4.)

- Die durchgeführten *Regressionsanalysen* erweisen sich aus verschiedenen Gründen als wenig robust. Die zur Schätzung verwendeten Fallzahlen sind sehr klein ($n=14$). Beim Datenmaterial handelt es sich um Zeitreihendaten. Wegen möglicher Verzerrungen (Autokorrelation der Residuen, Heteroskedastizität) ist hier der Einsatz der Methode der kleinsten Quadrate ('ordinary least squares', OLS) als Schätzverfahren nicht angebracht. Die Regressionsgleichung enthält zudem - darauf verweisen auch die Autoren (SB: 35) - nicht alle wichtigen exogenen Faktoren. Die Schätzungen ergeben nicht-signifikante Parameterschätzwerte, was jedoch durchaus auch auf die angeführten Probleme zurückgeführt werden könnte. Dieser Umstand scheint auch den Autoren bewusst gewesen zu sein, weshalb sie auf eine weitergehende Interpretation der Ergebnisse der Regressionsanalysen zumeist verzichten. (vgl. Kapitel 5.2.3.2.1. und 5.2.3.4.10)
- Die Evaluatoren vergleichen in jenem Teil der Studie, welcher sich mit den *Auswirkungen des Passöffnungstermins* auf das Tourismusaufkommen befasst, die Kosten der Schneeräumung mit dem Wertschöpfungsausfall beim Gastgewerbe im Falle eines völligen Verzichts auf die Schneeräumung. Wir haben bereits an anderer Stelle auf die Schwierigkeiten im Zusammenhang mit der Validität und Reliabilität der Daten zum Wertschöpfungsausfall hingewiesen (vgl. Kapitel 5.2.3.4.5 und 5.2.3.4.6.). Abgesehen davon stellt sich bei diesem Vergleich jedoch ein weiteres Problem. Da unseres Wissens ein vollständiger Verzicht auf die Schneeräumung nie zur Diskussion stand, sondern nur die Frage, wann die Schneeräumung durchgeführt werden soll, scheint es zweckmäßiger, eine andere (realistischere) Bezugsbasis für den Vergleich zu wählen. Gemeint ist damit beispielsweise die Verschiebung der Passöffnung um zwei Wochen und den damit entstehenden Wertschöpfungsausfall beim Gastgewerbe. Darauf gehen die Evaluatoren jedoch nicht ein. Sie diskutieren auch das mögliche Einsparungspotential bei den Schneeräumungskosten bei einer derartigen Verschiebung kaum (vgl. dazu Kapitel 5.2.3.4.10. und Kapitel 5.2.5.4.), obwohl dies ja eine Empfehlung der Evaluatoren selbst darstellt (siehe SB: 49-54).

Wegen der knapp bemessenen zeitlichen und finanziellen Ressourcen, welche zur Durchführung der Evaluation eingesetzt werden konnten, waren die Auftragnehmer gezwungen, bei der Datenauswertung einen pragmatischen Ansatz zu wählen. Trotz diesen schwierigen Rahmenbedingungen zeigt die Studie doch einen reichhaltigen Methodeneinsatz. Zudem sind die Analysen mehrheitlich technisch korrekt und sowohl dem Datenmaterial wie auch der Fragestellung angepasst erfolgt. Deshalb sind wir der Ansicht, dass die Studie trotz der aufgeführten Probleme das Kriterium erfüllt.

5.2.3.4.9. Analyse qualitativer Informationen

Aufgrund der Angaben im Schlussbericht ist nicht eruierbar, auf welche Art und Weise die qualitativen Datenbestände ausgewertet wurden. In dem Kapitel zur Zieldefinition (SB: 14-17), in dem dies noch am ehesten möglich wäre, wird der Leser durch die widersprüchlichen Angaben über die kontaktierten Akteure (siehe dazu Kapitel 5.2.3.4.4.) verunsichert.

Wegen der mangelhaften Berichterstattung zur Auswertung des qualitativen Datenmaterials im Schlussbericht ist es uns nicht möglich, eine Bewertung abzugeben.

5.2.3.4.10. Gerechtfertigte Folgerungen

Wir folgen bei der Diskussion dieses Kriteriums dem Abschnitt 'Auswertung und Umsetzungsvorschläge' im Evaluationsbericht (Kapitel 3.5., SB: 48-54).

- *Ungenügendes Zielsystem:* Die hier angeführten Schlussfolgerungen sind explizit begründet und aufgrund der erfolgten Evaluation gerechtfertigt.
- *Unproblematischer Vollzug:* Die Vollzugevaluation musste sich wegen der knappen Mittel, welche für die Untersuchung zur Verfügung standen, auf die wesentlichsten Gesichtspunkte beschränken. Deshalb sind wir der Ansicht, dass die Folgerungen, die einen einwandfreien Vollzug attestieren, mit einem Hinweis auf die knappe Analyse hätten versehen werden sollen (siehe dazu auch Kapitel 5.2.3.3.7). So kommen wir zum Schluss, dass die Evaluation in diesem Bereich das Kriterium nicht gänzlich erfüllt.
- *Ausbaufähiges Kostenbewusstsein:* Die Erkenntnis, dass bei der Festsetzung der Öffnungsdaten kaum Kostenüberlegungen berücksichtigt wurden, ist aufgrund der Evaluation berechtigt. Andererseits führt diese Feststellung nicht direkt zu der von den Evaluatoren formulierten Empfehlung, dass die Passöffnung vermehrt auf ein späteres Datum festzulegen sei. Eine mögliche Alternative, welche wir im Anhang begründet

haben (siehe Kapitel 5.2.5.4.), würde die Festlegung von Kostenvorgaben für die Schneeräumung beinhalten. Es erscheint aufgrund der starken Schwankungen, welche durch die unterschiedlichen Witterungs- und Schneeverhältnisse entstehen, wenig zweckmässig, die Festlegung der Passöffnungstermine vollständig von diesen jeweils spezifischen Rahmenbedingungen zu entkoppeln. Eine Steuerung via Kostenvorgaben erlaubt ein flexibles und dennoch kostensparendes Reagieren auf die Verhältnisse. Zur Sicherstellung der Passöffnung für unumgängliche Nutzungen (wie etwa die Zurverfügungstellung einer Ersatzroute zur Ermöglichung der Unterhaltsarbeiten am Gotthardstrassentunnel beim Gotthardpass) müssten die Vorgaben mit entsprechenden Ausnahmeregelungen ergänzt werden. Die Aussagen bezüglich der Kosteneinsparungsmöglichkeiten aufgrund einer verzögerten Inangriffnahme der Räumungsarbeiten anhand des Jahres 1993 sind mit Vorsicht zu interpretieren, da unklar bleibt, welcher Teil der finanziellen Einsparungen auf die Verzögerung zurückzuführen ist. Eine Möglichkeit zur Quantifizierung des Sparpotentials bei einer zeitverschobenen Passöffnung versuchen wir im Anhang zu entwickeln (vgl. Kapitel 5.2.5.4.). Für eine vertiefte Untersuchung dieses Aspekts sind jedoch weitergehende Abklärungen unumgänglich. Die Frage, welche eine Vergabe der Schneeräumungsarbeiten an private Unternehmen thematisiert, ist unseres Erachtens gerechtfertigt.

- *Ergebnisse und Empfehlungen zu den einzelnen Pässen:* Neben den bereits in den vorangehenden Abschnitten vorgebrachten Einwänden werden die Folgerungen zu den einzelnen Pässen explizit begründet. Die Evaluatoren verzichten jedoch auch hier auf die Quantifizierung der Kosteneinsparungen, die zu erwarten wären, wenn den Empfehlungen aus der Evaluation gefolgt werden würde. Dies würde es dem Leser ermöglichen, die Folgerungen einzuschätzen (vgl. Kapitel 5.2.3.4.8.). Der Hinweis, dass der Kanton Uri für die Schneeräumung am Gotthardpass - es handelt sich dabei um eine Nationalstrasse 3. Klasse - dank Bundesbeiträgen weniger als 5'000.- sFr. zu tragen habe (SB: 50), erweckt einen falschen Eindruck. Der Kanton Uri trägt lediglich fünf Prozent dieser Kosten selbst. Dies entspricht im Durchschnitt der letzten Jahre einem Betrag von weniger als 2'500.- sFr. Ebenfalls im Abschnitt zum Gotthardpass haben die Evaluatoren fälschlicherweise die Schneeräumungskosten mit dem Umsatzausfall (statt dem Wertschöpfungsausfall) im Gastgewerbe in Relation gesetzt (SB: 50, vgl. oben Kapitel 5.2.3.4.7.). Im Abschnitt zum Sustenpass wird der geschätzte Wertschöpfungsausfall bei einem Verzicht auf die Schneeräumung mit "rund 66'000 Fr." angegeben, was der Angabe im Hauptteil der Studie ("rund 30'000 bis 66'000 Fr.", SB: 39) nicht ganz entspricht. Im Abschnitt zum Furka- und Oberalppass werden die Betrachtungen zum Wertschöpfungsausfall im Urserental ohne Einbezug der Schnee-

räumungskosten am Gotthard gemacht. Dies entspricht nicht der Vorgehensweise in der detaillierten Analyse (vgl. Tabelle 3-23, SB: 40).

- *Zusammenfassung*: Diese gibt die Folgerungen soweit korrekt wieder und ist gut auf die vorherigen Aussagen abgestützt. Drei Hinweise sind anzuführen. Erstens findet sich zu Beginn der Zusammenfassung folgender Satz: "Die Kurzevaluation hat aufgedeckt, dass die Ziele der Schneeräumung bis heute nicht schriftlich fixiert sind." (SB: 53) Es ist zu vermuten, dass dies den beteiligten Akteuren schon vor der Evaluation bekannt war. Zweitens ist ein unpräzise formulierter Satz enthalten, welcher zu Missverständnissen Anlass geben könnte: "Wir empfehlen die Passöffnung verstärkt von Mitte Juni auf Ende Juni zu verschieben und auf Räumungsarbeiten vor diesem Zeitpunkt zu verzichten." (SB: 54) Für die Räumungsarbeiten werden durchschnittlich rund zwei Wochen benötigt. Der zweite Teil des Zitats würde somit bedeuten, dass eine Passöffnung erst Mitte Juli möglich wäre. Der dritte Hinweis, aber von der Aussage her sicherlich der wichtigste der drei, bezieht sich auf folgenden Satz: "Ebenfalls wurde aufgezeigt, dass eine geringfügige Verschiebung der Passöffnung keine oder nur" ... "sehr unbedeutende Auswirkungen auf die Umsätze im Gastgewerbe hat." (SB: 53, Auslassung TW) Neben den Schwierigkeiten im Zusammenhang mit der Abschätzung der Umsatzzahlen im Gastgewerbe, welche wir schon an anderer Stelle ausführlich diskutiert haben (vgl. Kapitel 5.2.3.4.5. und 5.2.3.4.6.), erfolgt hier eine Überinterpretation einer Regressionsanalyse. Wir haben uns diesbezüglich bereits an anderer Stelle geäußert (vgl. Kapitel 5.2.3.4.8.). Aufgrund unserer Analyse, wie auch aufgrund der durch die Evaluatoren selbst geäußerten Vorbehalte (siehe SB: 35), ist diese Aussage durch die durchgeführten empirischen Analysen nicht zu stützen.

Die Evaluation enthält neben diesen inhaltlichen Folgerungen auch einen kurzen Abschnitt zur methodischen Auswertung (Kapitel 4, SB: 55). Die Ausführungen beziehen sich schwergewichtig auf die Nützlichkeit und Zweckmässigkeit von Kurzevaluationen - also Evaluationen mit beschränkten Ressourcen. Auch wenn diese Einschätzungen vielleicht etwas zu euphorisch ausgefallen sind, sind sie dennoch ausreichend begründet und soweit auch einschätzbar, so dass sich dazu keine weiteren Bemerkungen aufdrängen.

Wir sind der Ansicht, dass die Folgerungen in einigen Punkten nicht den Vorgaben des Kriteriums entsprechen. Besonders bedeutend erscheinen dabei die Empfehlungen zur Festlegung von Terminen für die Passöffnung, welche zum Teil nicht ausreichend begründet werden, die Hinweise zum unproblematischen Vollzug, welcher für die gemachten Feststellungen nicht ausreichend untersucht worden ist, sowie die Aussagen zu den Auswirkungen auf die Umsätze im Gastgewerbe, welche aufgrund der empirischen Analysen nicht haltbar erscheinen. Es finden sich zudem verhältnismässig häufig - es

handelt sich dabei ja um einen vielgelesenen Teil einer Evaluationsstudie - Ungenauigkeiten und Fehlerhaftigkeiten. In vielen Fällen werden die Folgerungen aber auch ausreichend begründet, so dass deren Einschätzung möglich ist. Wir kommen aus diesen Gründen zum Schluss, dass die Evaluation das Kriterium teilweise erfüllt.

5.2.3.4.11. Objektive Berichterstattung

Die Evaluation ist gegenüber Verzerrungen auf der methodischen Ebene aufgrund der Vielfalt der Verfahren recht gut geschützt. Problematisch erscheint jedoch vor allem die Wahl des Untersuchungsgebietes zu sein. Aufgrund der grenzüberschreitenden Charakteristik der untersuchten Pässe ist die Beschränkung des Untersuchungsgebietes auf den Kanton Uri wenig sachgerecht. Die rein ernerische Perspektive führt zu einer einseitigen Betrachtung des Gegenstandes. Ein zweiter kritischer Punkt stellt die Reduktion der Tourismusbranche auf das Gastgewerbe dar. Neben diesem sind ja noch weitere Gewerbebezüge am Tourismusaufkommen beteiligt. Wegen dieses Vorgehens besteht die Gefahr, dass die Evaluation die tangierten Umsätze als zu tief einschätzt, was zu einer Verzerrung der Untersuchungsergebnisse führen könnte. Die Würdigung des Vollzugs im Kapitel zur Auswertung (SB: 49) erscheint - auf dem Hintergrund der eher knappen Vollzugsanalyse (SB: 18-25) und den daraus gewonnen Erkenntnissen (SB: 26) - etwas unkritisch erfolgt zu sein. Aus dem Satz: "Der administrative und technische Vollzug der Schneeräumung bietet praktisch keine Probleme" (SB: 26) im Analyseteil der Studie, wird in der Auswertung: "Der administrative Vollzug der Schneeräumung funktioniert einwandfrei." (SB: 49).

Aufgrund dieser Feststellungen scheint die Evaluation nur teilweise vor Verzerrungen geschützt zu sein. Die Evaluation ist zwar methodisch breit abgestützt. Aufgrund des Untersuchungsdesigns (Abgrenzung Untersuchungsgebiet) und der restriktiven Rahmenbedingungen für die Durchführung ergeben sich jedoch einige wesentliche Einschränkungen. Deshalb kann das Kriterium nur teilweise als erfüllt betrachtet werden.

5.2.4. Gesamteindruck

Die Bewertung zur Evaluation 'Schneeräumung von Urner Passstrassen' fällt äusserst positiv aus. Dies gilt im Prinzip für alle vier Kriteriengruppen. Nur gerade im Bereich der Genauigkeit sind wenige Abstriche zu machen. Die Studie ist finanziell wie zeitlich nur mit kleinen Ressourcen ausgestattet. Sie bearbeitet eine Thematik, die sehr stark eingegrenzt ist und als relativ technisch bezeichnet werden kann. Sicherlich weist die Studie nicht jene Informationsdichte und analytische Tiefe einer grösseren Evaluation auf. Es ist jedoch gelungen - im gegebenen Rahmen - eine gute Evaluation zu erarbeiten. Besonders positiv fällt ins Gewicht, dass die (enge) Fragestellung konsequent verfolgt wurde, dass ein pragmatischer Methodeneinsatz angepasst auf die gegebenen Möglichkeiten erfolgte und dass die Berichterstattung in Aufbau, Sprache und Argumentation überzeugt. Schwächen sind vor allem zwei auszumachen. Erstens bleibt durch die enge Abgrenzung der Fragestellung und des Untersuchungsbereichs ein wichtiger Bereich (die ausserkantonalen Fragen) aus der Evaluation ausgeklammert. Dies ist unseres Erachtens nicht adäquat, da sich das Evaluandum durch seine grenzüberschreitende Charakteristik auszeichnet. Zweitens ergeben sich in der Datenanalyse einige Probleme, die jedoch - bei den gegebenen finanziellen und zeitlichen Möglichkeiten - kaum befriedigend hätten gelöst werden können.

Zusammenfassend ist die Studie sowohl im Konzeptionellen wie auch in der Umsetzung als sehr gelungen zu bezeichnen, auch wenn sich in einigen Aspekten die (durch die knappen finanziellen und zeitlichen Ressourcen bedingten) Restriktionen deutlich manifestieren.

5.2.5. Anhang

5.2.5.1. Unterlagen zur Evaluationsstudie

Die Untersuchung dieser Evaluationsstudie beruht auf der Analyse von schriftlichen Materialien und auf Gesprächen mit einigen relevanten Akteuren (vgl. Kapitel 5.2.5.2.). Zur Erarbeitung dieser Fallstudie sind die folgenden schriftlichen Unterlagen verwendet worden:

- Forschungsgesuch vom 10/2/1991
- Bewilligungsschreiben Abt. IV des SNF vom 21/1/1992
- Ausschreibungsunterlagen 'Beiträge des NFP 27 an Kurzevaluationen der Kantone und Städte' vom 18/6/1992
- Gesuch des Bauamtes des Kantons Uri an NFP 27 vom 30/10/1992 (GES/OF)
- Schreiben PL NFP 27 an Bauamt Kanton Uri vom 3/11/1992
- Schreiben Bauamt Kanton Uri an PL NFP 27 vom 17/11/1992
- Vertrag zwischen dem NFP 27 und dem Kanton Uri vom 6/1/1993 bzw. 19/1/1993
- Offerte von Ecoplan, Februar 1993 (EK)
- Arbeitsgrundlage 'Datenbedarf', undatiert [Februar 1993]
- Fragebogen 'Gastgewerbe Erhebung Saisoncharakter', Erhebungsjahr 1992, des KIGA Uri, undatiert
- Entwurf Schlussbericht vom [3]/9/1993 (ESB)
- Schlussbericht vom [10]/9/1993 (SB)
- Schreiben (inkl. Journal/Logbuch) Bauamt Kanton Uri an PL NFP 27 vom 12/11/1993
- Schreiben (inkl. Schlussrechnung Ecoplan) Bauamt Uri an PL NFP 27 vom 1/12/1993

5.2.5.2. Liste der kontaktierten Personen

- Werner Bussmann, Programmleitung NFP 27, Bern
- Jean W. Flückiger, Sektion Tourismus, Abteilung Raumwirtschaft, Bundesamt für Statistik, Bern
- Edoardo Frei, Hauptabteilung Strassen, Bauamt, Baudirektion Kanton Uri, Altdorf
- Heini Sommer, Altdorf
- Stefan Suter, Bern
- Verena Schüpbach, Abteilung Projektierung, Umwelt, Verkehr, Bundesamt für Strassenbau, Bern

5.2.5.3. Literaturverzeichnis

- BAUAMT, KANTON URI (Hrsg.) (1994): Öffnungs- und Schliessungskonzept der Urner Alpenpässe. Grundlagen - Fixdaten - Betriebliche Massnahmen. Stand: März 1994. Altdorf: Bauamt (mimeo).
- DÜRSTELER, Urs (1984): Kantonale Einkommenswirkungen der Bautätigkeit des Bundes. Eine empirische Analyse. Diss. Hochschule St. Gallen. Bern: Haupt.
- ECOPLAN (Hrsg.) (1992): Bestelliste für Ecoplan-Publikationen. Bern und Altdorf: Ecoplan.
- ECOPLAN (Hrsg.) ([1994]): Publikationen und Projekte. Broschüre. Bern/Altdorf: Ecoplan.
- ECOPLAN (Hrsg.) (o.J.): Ecoplan. Wirtschafts- und Umweltstudien. Informationsbroschüre. Bern und Altdorf: Ecoplan.
- HÄUSEL, Urs (1985): Die regionale Inzidenz von drei Infrastrukturobjekten. Regionalökonomische Auswirkungen der N2-Osttangente Basel, des Biozentrums Basel und des Kantonssspitals Basel. Diss. Universität Basel.
- HARVEY, Andrew C. (1990): The econometric analysis of time series. Second edition. New York: Philip Allan.
- ITEN, Karl (1990): Adieu - Altes Uri. Aspekte des Wandels eines Kantons vom 19. ins 20. Jahrhundert. Zürich: Neue Zürcher Zeitung.
- MASUHR, K.P. (1988): Volkswirtschaftliche Auswirkungen verschiedener Energieszenarien. Studie im Auftrag des Bundesamtes für Energiewirtschaft. Expertengruppe Energieszenarien, Schriftenreihe Nr. 25. Prognos (Basel), 25. Januar 1988. Bern: EDMZ.
- MEIER, Ruedi/WALTER, Felix (1991): Umweltabgaben für die Schweiz. Ein Beitrag zur Ökologisierung von Wirtschaft und Gesellschaft. Chur/Zürich: Rüegger.
- NEUENSCHWANDER, René/SOMMER, Heini/WALTER, Felix (1990): Grundlagen für die Erfolgskontrolle von Subventionen des Kantons Bern. Studie im Auftrag der Finanzdirektion des Kantons Bern. Schlussbericht. Bern: Ecoplan (mimeo).
- NEUENSCHWANDER, René/SOMMER, Heini/WALTER, Felix (1991): Soziale Kosten von Verkehrsunfällen in der Schweiz. Studie im Auftrag des Dienstes für Gesamtverkehrsfragen des Eidgenössischen Verkehrs- und Energiewirtschaftsdepartements und des Bundesamtes für Bildung und Wissenschaft des Eidgenössischen Departements des Innern. Schlussbericht, April 1991. Bern/Altdorf: Ecoplan.
- NEUENSCHWANDER, René/SOMMER, Heini/SUTER, Stefan/WALTER, Felix (1992): "Externe Kosten des Agglomerationsverkehrs und Internalisierung." In: Schweizerische Zeitschrift für Volkswirtschaft und Statistik, Vol. 128, No. 3: 437-451.
- NEUENSCHWANDER, René/SUTER, Stefan/WALTER, Felix/SOMMER; Heini (1992a): Internalisierung externer Kosten im Agglomerationsverkehr. Fallbeispiel Region Bern. Bericht 15A des NFP 'Stadt und Verkehr'. Zürich: NFP 25.
- NEUENSCHWANDER, René/SUTER, Stefan/WALTER, Felix/SOMMER; Heini (1992b): Externe Kosten im Agglomerationsverkehr. Fallbeispiel Region Bern. Bericht 15B des NFP 'Stadt und Verkehr'. Zürich: NFP 25.
- RÜTTER, Heinz (1990): "Wirtschaftsfaktor Tourismus." In: Die Volkswirtschaft, No. 12, 1990: 15-22.
- SAYRS, Lois W. (1989): Pooled time series analysis. Newbury Park: Sage.
- SOMMER, Heini (1990): Kantonale Nationalstrassen-Rechnung. Die Kosten und Erträge des Kantons Uri, ihr Vergleich mit anderen Kantonen und eine ökonomische Analyse der bundesstaatlichen Regelungen im Nationalstrassenbereich. Diss. Universität Bern, 1989. Bern: Haupt.

- SOMMER, Heini/NEUENSCHWANDER, René/WALTER, Felix (1993): Externe Nutzen des Verkehrs. Wissenschaftliche Grundlagen. Bericht 39 des NFP 'Stadt und Verkehr'. Zürich: NFP 25.
- SOMMER, Heini/NEUENSCHWANDER, René/WALTER, Felix (1993): Externe Nutzen des Verkehrs. Kurzfassung. Studie im Auftrag des Dienstes für Gesamtverkehrsfragen und des NFP 25 'Stadt und Verkehr, September 1993. GVF-Bericht 1/93. Bern: Generalsekretariat EVED, Dienst für Gesamtverkehrsfragen.
- SUTER, Stefan/WALTER, Felix (1993): Gestaltungsmöglichkeiten und Auswirkungen verschiedener Submissionsordnungen. Evaluation und Bericht im Auftrag der Bau-, Verkehrs- und Energiedirektion des Kantons Bern, mitfinanziert durch NFP 27 und Eidg. Finanzverwaltung. Schlussbericht, Dezember 1993. Bern/Altdorf: Ecoplan.
- WALTER, Felix/SUTER, Stefan/NIEUWKOOP, Renger van (1993): Umweltabgaben in Europa. Überblick und Vorstudie für die Tagung 'Umweltabgaben in Europa - Konsequenzen für die Schweiz' im Auftrag des Bundesamtes für Umwelt, Wald und Landschaft (BUWAL) und des Bundesamtes für Energiewirtschaft (BEW). Schriftenreihe Umwelt, Nr 198. Bern: BUWAL.

5.2.5.4. Ergänzungen zur quantitativen Auswertung

Die Autoren verzichten auf eine eingehendere Analyse des Zusammenhangs zwischen dem Zeitpunkt der Passöffnung und den Kosten der Schneeräumung. Wir möchten dies an dieser Stelle nachholen.

Wir stützen uns für die Analyse auf folgende Datengrundlagen. Die Schneeräumungskosten für die unterschiedlichen Jahre und Pässe entnahmen wir dem Schlussbericht der Evaluation (SB: 24). Die Termine für die Passöffnungen entstammen einer Zusammenstellung, welche uns das Bauamt Uri freundlicherweise zur Verfügung stellte. Der Evaluationsbericht enthält dazu lediglich die graphische Repräsentation des Datenmaterials (siehe SB: 19-21).

Zur Untersuchung des Zusammenhangs wird ein 'pooled time series'-Design (siehe Sayrs 1989) eingesetzt. Wir verfügen über fünf Zeitreihen mit Jahresdaten, welche sich auf die fünf Pässe und die Jahre 1981 bis 1993 beziehen. Die Werte für die beiden Indikatoren mussten folgendermassen aufbereitet werden:

- *Kosten der Schneeräumung*: Zentrierung und Prozentuierung; das heisst, die analysierten Werte stellen die prozentualen Abweichungen der für diesen Pass durchschnittlichen Schneeräumungskosten eines bestimmten Jahres dar.
- *Zeitpunkt der Passöffnung*: Zentrierung; das heisst die analysierten Werte stellen die Abweichungen vom durchschnittlichen Öffnungstermin des jeweiligen Passes in einem bestimmten Jahr in Tagen dar.

Die Beobachtung 'Susten 1989' wurde in der Analyse nicht berücksichtigt, da sich damals die Passöffnung aufgrund eines Felssturzes bis in den August hinein verzögerte (vgl. dazu die Hinweise in Kapitel 3.4.6.). Wir verfügen somit über 64 Beobachtungen zur weiteren Analyse (5 Pässe, 13 Jahre, ohne 'Susten 1989').

Wir verwenden für die Analyse ein bivariates Regressionsmodell in der Form:

$$y = a + bx + e$$

mit: $y = \text{Kosten der Schneeräumung}$

$x = \text{Tag der Passöffnung}$

Wir gehen damit von der Hypothese aus, dass ein linearer Zusammenhang zwischen den beiden Variablen besteht, was aufgrund einer vorgängig durchgeführten graphischen Analyse der Daten mittels eines Streudiagramms zweckmässig scheint.

Da mit zentrierten Daten gearbeitet wird, entfällt das Absolutglied a (korrekterweise $a=0$). Die Regressionsgleichung lautet demzufolge nun:

$$y = bx + e$$

Zur Schätzung verwendeten wir das integrierte Statistiksoftwarepaket SAS (Statistical Analysis System) und setzten dabei eine OLS-Schätzung ein.

Die Schätzung ergab folgende Resultate (T-Werte in Klammern):

$$b = 2.71 (5.33)$$

$$R^2 = 0.31$$

Aufgrund der Tatsache, dass die Daten der fünf Pässe gemeinsam für die Analyse eingesetzt wurden, wäre denkbar, dass dies die Schätzung verzerren könnte. Aufgrund von Regressions-schätzungen, bei denen die Pässe in der Form von Dummy-Variablen integriert wurden, kann dies jedoch verneint werden. Die Parameterschätzwerte zu diesen Variablen waren in keinem Fall signifikant. Eine ebensolche Überprüfung wurde auch bezüglich der Jahre durchgeführt. Mit einer Ausnahme (1993, siehe unten) ergeben sich keine signifikanten Schätzwerte, womit auch keine Jahresabhängigkeit vorliegt.

Aufgrund dieser Ergebnisse kommen wir zur Folgerung, dass die Schneeräumungskosten ansteigen, je später der Pass geöffnet wird. Dies ist jedoch keineswegs als linearer Zusammenhang zu interpretieren. Beide Variablen werden von einer oder mehreren vorgelagerten Variablen (Witterungs- und Schneeverhältnisse) beeinflusst. Die dazugehörigen Hypothesen lassen sich wie folgt formulieren:

- Je besser die Witterungs- und Schneeverhältnisse für die Räumung sind, umso tiefer fallen die Räumungskosten aus.
- Je besser die Witterungs- und Schneeverhältnisse für die Räumung sind, umso früher wird der Pass geöffnet.

Verbindet man diese beiden Hypothesen mit den Schätzergebnissen, lässt sich folgern, dass die verantwortlichen Entscheidungsträger, je später umso eher steigende Räumungskosten in Kauf zu nehmen scheinen.

Die Evaluatoren argumentieren, dass im Jahre 1993 massive Kosteneinsparungen möglich wurden, weil die Schneeräumung (aus ungeklärten Gründen) um rund eine Woche verzögert wurde (SB: 53). Sie verzichteten jedoch darauf, die möglichen Einsparungen zu quantifizieren.

Die vorangehende Regressionsanalyse kann nun um eine weitere Variable ergänzt werden. Diese ist wiederum als Dummy konstruiert und steht für das Jahr 1993. Das heisst, alle Beobachtungen, die im Jahr 1993 erfolgt sind, erhalten für diese dichotome Variable den Wert '1', die übrigen Beobachtungen den Wert '0'. Die bivariate Regressionsgleichung wird um diese Variable ergänzt:

$$y = b_1x_1 + b_2x_2 + e$$

mit: y = Kosten der Schneeräumung

x_1 = Tag der Passöffnung

x_2 = Jahr 1993 (Dummy)

Da die Passöffnung im Jahre 1993 um rund eine Woche später erfolgte, als dies nach den bisherigen Modalitäten der Terminierung der Fall gewesen wäre, müssen die entsprechenden Werte zur Variable 'Tag der Passöffnung' für die Beobachtungen im Jahr 1993 um sieben Tage reduziert werden.

In einem ersten Schritt haben wir eine bivariate Korrelationsanalyse mit folgenden Ergebnissen durchgeführt (Die Werte entsprechen den Produkt-Moment-Korrelationskoeffizienten nach Pearson (r). Alle Koeffizienten sind bei einem Niveau von $p \leq 0.01$ signifikant):

	y	x_1	x_2
y	1	/	/
x_1	.60	1	/
x_2	-.45	-.51	1

Dies zeigt einerseits die bereits anhand der bivariaten Regressionsanalyse konstatierte Korrelation zwischen x_1 und y ($r = .60$). Andererseits ist aus obiger Matrix auch erkennbar, dass zwischen x_2 und y ein negativer Zusammenhang festzustellen ist ($r = -.45$). Die negative Korrelation zwischen x_1 und x_2 wiederum zeigt an, dass im Jahre 1993 die Pässe früher als im Durchschnitt üblich geöffnet worden sind ($r = -.51$). Dieser Koeffizient wurde durch die oben erwähnte Verschiebung um sieben Tage zusätzlich erhöht.

Die trivariate Regressionsschätzung anhand obenstehender Gleichung ergibt folgende Werte (T-Werte in Klammern):

$$b_1 = 2.26 (4.30)$$

$$b_2 = -23.90 (-1.30)$$

$$R^2 = 0.37$$

Unter der Annahme, dass im Jahr 1993 effektiv eine Woche später, als dies die Verhältnisse suggeriert haben, mit der Schneeräumung begonnen wurde und dass dies nur gerade in diesem Jahr geschehen ist, kommen wir zum Schluss, dass die Kosten der Schneeräumung bei einer Verschiebung um lediglich eine Woche um rund 24 Prozent reduziert werden würden.

Dieser Wert ist jedoch in mehrfacher Hinsicht zu problematisieren. Erstens ist aufgrund der Korrelationsmatrix ersichtlich, dass auch zwischen den beiden 'unabhängigen' Variablen eine Korrelation auftritt, also möglicherweise aufgrund von Multikollinearität verzerrte Schätzwerte resultieren. Zweitens ist unser Modell sicherlich fehlspezifiziert, da nicht alle relevanten Variablen darin enthalten sind und weil nicht anzunehmen ist, dass der Zusammenhang zwischen x_2 und y linear verläuft. Drittens ist der Parameterschätzwert zu b_2 nicht signifikant. Die Schätzung ist somit zu wenig verlässlich, um das Einsparungspotential alleine darauf basierend quantifizieren zu wollen.

Aufgrund der durchgeführten Analysen erscheint die Empfehlung der Evaluatoren, die Passöffnung an fixen Terminen auszurichten, in Frage gestellt zu werden. Dadurch würde nämlich die terminliche Festsetzung der Passöffnung von den Witterungs- und Schneebedingungen entkoppelt. Da diesbezüglich sehr grosse und auch kostenrelevante Schwankungen zu beobachten sind, scheint es uns zweckmässiger zu sein, die Terminierung der Schneeräumung nicht über Termine zu steuern, sondern dafür Kostenvorgaben vorzusehen. Je nach Gewichtung einer frühen Passöffnung kann damit:

1. ebenfalls ein Einsparungspotential erschlossen werden und
2. auf die jährlichen und lokalen Schnee- und Witterungsverhältnisse besser Rücksicht genommen werden.

Die multivariate Regressionsanalyse versuchte das Kosteneinsparungspotential einer Verschiebung des Passöffnungstermins zu quantifizieren. Die so geschätzten Einsparungsmöglichkeiten sind recht bedeutend. Allerdings ist diese Schätzung mit grossen

Vorbehalten zu versehen. Zur präziseren Erfassung sind weitere Abklärungen dringend erforderlich.

5.2.6. Stellungnahme der Evaluatoren

Eine der wichtigsten Voraussetzung für das Funktionieren von Märkten ist die Markttransparenz: Märkte funktionieren nur dann optimal, wenn deren Teilnehmer über Angebot und Nachfrage informiert sind. Als Ökonomen und Evaluatoren sind wir interessiert, dass auch auf "unserem" Evaluationsmarkt Transparenz herrscht. Wir begrüßen daher das vorliegende Projekt der Meta-Evaluation, weil mit dem gewählten Vorgehen ein wesentlicher Beitrag zur vermehrten Transparenz im Schweizerischen Evaluationsmarkt geleistet wird und in Zukunft noch geleistet werden kann.

Wir stellen uns auch der Qualitätsbeurteilung, die mit dem gewählten Vorgehen des Meta-Evaluationsteams verbundenen ist. Aufgrund der insgesamt sehr positiven Beurteilung unserer Studie in Kapitel 5.2.4.2 und 6.2 fällt uns dies natürlich auch entsprechend einfacher. Trotzdem möchten wir auf *zwei allgemeine Aspekte* hinweisen, die bei der Beurteilung der Einzelkriterien u.E. problematisch sind:

- Aufgrund des gewählten Ansatzes werden die Evaluatoren der einzelnen Studien auch für Mängel "belangt", die nicht in ihrem Entscheid oder Einflussbereich lagen. Aus Sicht der Autoren ist dies selbstverständlich störend. Wir würden daher empfehlen, grundsätzlich zwischen Kriterien zu unterscheiden, die im Entscheidbereich der Autoren liegen und solchen, die von den Rahmenbedingungen oder den Vorstellungen und Wünschen des Auftraggebers vorgeben werden. Besonders bei Kurzevaluationen ist diese Differenzierung besonders wichtig: Der beschränkte Mittel- und Zeiteinsatz führt notwendigerweise zu Rahmenbedingungen, die einer idealtypischen Vorstellung nicht entsprechen, weil bewusst auf die Klärung und/oder vertiefte Behandlung gewisser Fragestellungen verzichtet wird.
- Bei der Beurteilung verschiedener Einzelkriterien sind wir der Ansicht, dass eine Ausgangsbasis bzw. eine "Messlatte" gewählt wird, die zu relativ "harten" Urteilen führt. Bei den teilweise ausserordentlich umfassenden Ansprüchen, die der Beurteilung des Meta-Evaluationsteams zugrunde liegen, wird vernachlässigt, dass im Normalfall kein unbeschränktes Zeit- und Kostenbudget zur Verfügung steht. Die Evaluatoren sind deshalb oft gezwungen, bei Fragen von zweitrangiger Bedeutung Vereinfachungen und Kürzungen vorzunehmen, um innerhalb des gegebenen Budgetrahmens zu bleiben. So kann u.E. z.B. ohne Konsequenzen für die Qualität einer Untersuchung darauf verzichtet werden, im Schlussbericht einer Evaluation auch noch alle angefragten Informationsquellen anzuführen, die sich nachträglich als wenig oder nicht zweckmässig erwiesen haben. Diesbezüglich haben wir im Rahmen der vorliegenden Meta-Evalua-

tion verschiedentlich die Sicht von privatwirtschaftlich tätigen Evaluatoren vermisst, für welche die Einhaltung der Budgetvorgaben unmittelbar gewinnrelevant ist.

Die folgenden *Erläuterungen* beziehen sich auf *spezifische Aussagen der Meta-Evaluation* zu unserer Studie "Schneeräumung von Urner Passstrassen". Aufgrund der vielfältigen Interdependenzen zwischen den Kriterien müssten die nachstehenden Erläuterungen bei weiteren Kriterien berücksichtigt werden. Wir verzichten jedoch darauf und beschränken uns auf eine einmalige Darstellung:

- *Verzicht auf Begleitgruppe* (Kapitel 5.2.3.1.1): Es war der ausdrückliche Wunsch des Auftraggebers, auf eine Begleitgruppe entgegen der ursprünglichen Absicht zu verzichten. Vorerst sollten Entscheidungsgrundlagen beschafft werden und erst anschliessend sollte über die Frage der Passöffnungsdaten eine öffentliche Diskussion geführt werden.
- *Verbreitung des Berichts* (Kapitel 5.2.3.1.6): Wie bereits erwähnt, wollte der Auftraggeber vorerst eine Analyse des Problems, ohne bereits in einem politisch sensiblen Bereich eine öffentliche Diskussion auszulösen. Auf Wunsch des Auftraggebers haben wir daher zur Verbreitung des Berichts keine eigenen Anstrengungen unternommen.
- *Politische Überlebensfähigkeit* (Kapitel 5.2.3.2.2): Die Beschränkung des Untersuchungsraums auf den Kanton Uri war wiederum ein ausdrücklicher Wunsch des Auftraggebers. Dabei waren sich sowohl Auftraggeber als auch Auftragnehmer bewusst, dass für eine umfassende Beurteilung der Passöffnungsfrage grundsätzlich sämtliche Kantone bzw. die Auswirkungen in diesen Kantonen mitzubehalten sind.
- *Beschreibung von Zielen und Vorgehen* (Kapitel 5.2.3.4.3.) / *Verlässliche Informationsquellen* (Kapitel 5.2.3.4.4.): Die Kritik bezüglich den Informationslücken zur Durchführung der Interviews und der Darstellung der Interviewergebnisse erachten wir als gerechtfertigt. In diesem Sinne nehmen wir die Anregungen gerne entgegen.
- *Reliable Messung*: Problematisch sind u.E. die Ausführungen zu den (Umsatz-) Ausfällen im Tourismus, weil diese nicht ausreichend auf die enormen Datenschwierigkeiten Rücksicht nehmen. Grundsätzlich ist jede Evaluatorin und jeder Evaluator aus eigenem Interesse bemüht, mit möglichst aussagekräftigen und konsistenten Daten zu arbeiten. Immer wieder sind jedoch die Evaluatoren mit der Situation konfrontiert, dass die vorhandenen bzw. die mit vernünftigem Aufwand beschaffbaren Daten nicht diesen Idealvorstellungen entsprechen: Kompromisse bezüglich der Datenqualität sind daher oft nicht zu vermeiden. In solchen Fällen ist es u.E. für die Glaubwürdigkeit der

Arbeit entscheidend, dass auf das Datenproblem und die beschränkte Aussagekraft ausreichend hingewiesen wird.

Von besonderer Bedeutung ist die "Kompromissbereitschaft" bei Kurzevaluationen: Es ist gerade das Ziel von Kurzevaluationen, mit einem stark beschränkten Mitteleinsatz zu ersten Ergebnissen bzw. Hinweisen auf einen evtl. Handlungsbedarf zu kommen. Dies bedeutet aber auch, dass Kompromisse sowohl in bezug auf die Datenbeschaffung, die theoretische Wirkungsanalyse als auch andere Vorgehensschritte notwendig bzw. unvermeidbar sind.

Abgesehen von diesen grundsätzlichen Erwägungen muss bezüglich unseres Fallbeispiels "Schneeräumung von Urner Passstrassen" zusätzlich auf die *ausserordentlichen Schwierigkeiten* bei der Datenbeschaffung im Tourismusbereich des Kantons Uri hingewiesen werden. Erst nach erheblichem Aufwand war es überhaupt möglich, an gewisse Umsatzzahlen zu gelangen. Aufgrund fehlender statistischer Unterlagen und auch aus Datenschutzgründen war es dem Kanton Uri beispielsweise nicht möglich, uns irgendwelche Angaben zu Umsatz- oder Übernachtungszahlen im Gastgewerbe zur Verfügung zu stellen. Letztlich standen wir vor der Entscheidung, entweder auf jegliche Aussagen zum möglichen Umsatzausfall in der Tourismusbranche zu verzichten, oder wenigstens anhand der hochgerechneten - und ebenfalls nur sehr schwierig zu beschaffenden - Umsatzzahlen im Gastgewerbe eine Abschätzung der *Grössenordnung* vorzunehmen. Auf die Schwierigkeiten der Hochrechnung sowie die Unsicherheiten bezüglich der absoluten Zahlenangaben wird an mehreren Stellen in unserem Text ausdrücklich hingewiesen (vgl. z.B. S. 35 u. 50 unserer Studie), immer wieder sprechen wir auch explizit von einer "Abschätzung einer Grössenordnung" (vgl. z.B. S. 35, 36, 40).

Die Beurteilung der Meta-Evaluation, dass das Kriterium der "Reliabilität der Messung" nicht erfüllt ist, erachten wir angesichts der oben erwähnten Ausgangslage als zu kategorisch.

Mit der Darstellung und Beurteilung der übrigen Aspekte bzw. Kriterien können wir uns - im Rahmen eines Ermessensspielraums, der dem Meta-Evaluationsteam zusteht - einverstanden erklären. Jedenfalls drängen sich weitere wichtige Präzisierungen oder Ergänzungen aus unserer Sicht nicht auf.

Heini Sommer und Stefan Suter

5.3. FALLSTUDIE 'EVALUATION DES BUWAL' (STUDIE NR. 9)

5.3.1. Die Evaluation

5.3.1.1. Das Thema

Die Evaluation (Knoepfel/Zimmermann/Sailer/Matafora 1991; Schlussbericht (SB)) befasst sich mit dem Bundesamt für Umwelt, Wald und Landschaft (BUWAL) im Eidgenössischen Departement des Innern (EDI). Das BUWAL entstand anfangs 1989 aus der Fusion des Bundesamtes für Umweltschutz (BUS) und des Bundesamtes für Forstwesen und Landschaftsschutz (BFL). Im Auftrag der Geschäftsprüfungskommission (GPK) des Nationalrates (NR) untersuchte die Evaluation die Politikleistungen *und* die Organisationsstrukturen. Zwischen diesen beiden Aspekten bestünden, so die Grundthese der Untersuchung, Zusammenhänge, so dass eine gemeinsame Betrachtung - "im Sinne einer politikorientierten Verwaltungsanalyse" (SB: 2) - als zweckmässig erscheine. Die Evaluation befasste sich vor allem mit den Bereichen Luftreinhaltung, ländlicher Gewässerschutz und der Umweltverträglichkeitsprüfung (UVP). Sie beschäftigte sich sowohl mit der Schnittstelle Kantone-Bund, wie auch mit der Bundesebene. Auf kantonaler Ebene erfolgte eine Breitenuntersuchung, welche die Gesamtheit der Kantone abdeckte, und eine Vertiefung in vier ausgewählten Kantonen (Basel-Stadt, Thurgau, Schwyz und Waadt). Auf Bundesebene wurden sowohl die amtsinternen, wie auch die externen Beziehungsnetze und Strukturen einer Untersuchung unterzogen. Aus den Untersuchungen resultierten Empfehlungen, die sowohl die Politikleistungen als auch die Verwaltungsorganisation betreffen.

5.3.1.2. Die Fragestellung

"Die Expertise soll auf folgende Fragen Antwort geben:

- Welche Schwierigkeiten haben die kantonalen Behörden, die Massnahmen zur Luftreinhaltung zu vollziehen?
- Hat im Bereich der Luftreinhaltung das BUWAL die Kompetenzen überschritten, die ihm vom USG eingeräumt werden?

Zusatzfrage: Gibt es Widersprüche

- im USG
- zwischen USG und den Vollzugsverordnungen

- zwischen den verschiedenen Vollzugsverordnungen
- zwischen der Strategie des Bundesrates gegen die Luftverschmutzung und dem USG und seinen Vollzugsverordnungen?
- Ist die Struktur des BUWAL geeignet, Einheit und Durchschlagskraft der Umweltpolitik sicherzustellen?
- Funktioniert das BUWAL, das wie ein Bundesamt organisiert ist, in Wirklichkeit wie eine Gruppe?" (EK: 1-2; vgl. auch SB: 1-2 und Bericht GPK: 4)

Die Fragestellung wurde sodann in der Diskussion auf den ländlichen Gewässerschutz und auf die Umweltverträglichkeitsprüfung ausgeweitet (GPK 1992: 4).

5.3.1.3. Das Vorgehen

Neben dem Studium einschlägiger Dokumente und Unterlagen - die Evaluatoren hatten, da sie im Auftrag der GPK des Nationalrates tätig waren, ein Akteneinsichtsrecht erhalten - und den Erfahrungen aus vorgängig oder parallel durchgeführten Untersuchungen (vgl. SB: 15-17; Knoepfel/Zimmermann 1987a, Knoepfel/Zimmermann 1993, Knoepfel/Imhof/Zimmermann 1994, Zimmermann/Balthasar/Neuenschwander 1991, Freiburghaus et al. 1990 und 1991 und Zimmermann/Aeschlimann 1991) basiert die Evaluation vor allem auf Befragungen.

Auf der kantonalen Ebene führten die Evaluatoren in einem ersten Schritt eine quantitative, schriftliche Befragung bei den für den Umweltschutz zuständigen Stellen aller Kantone durch. Die nachfolgende vertiefte Untersuchung mit problemorientierten Interviews erfolgte in vier ausgewählten Kantonen (Basel-Stadt, Thurgau, Schwyz und Waadt). Diese Gespräche dienten der Vertiefung und Verfestigung der Resultate aus der quantitativen Befragung. Diese Abklärungen wurden ergänzt durch einzelne Fallstudien zu vier Subventionsverfahren für Abwasserreinigungsanlagen und zu drei UVP-Verfahren mit Anhörungsverfahren.

Auf der Bundesebene erfolgte ebenfalls zuerst eine quantitative Befragung mit einem Fragebogen bei allen Mitarbeitern des BUWAL. Aufgrund der Erkenntnisse aus der quantitativen Befragung wurden zur Vertiefung Gespräche mit verschiedenen Mitarbeitern des BUWAL, aber auch mit dem Departementsvorsteher und dem Generalsekretär des EDI sowie mit Vertretern anderer Bundesämter geführt. Weiter wurden der 'Verwaltungsausstoss' des BUWAL und die Beschlüsse des Bundesrates zur Luftreinhaltepolitik einer Analyse unterzogen. Zudem wurde ein Rechtsgutachten zu den zentralen Erlassen erstellt.

5.3.1.4. Der Auftraggeber

Die Evaluation wurde im Auftrag der Geschäftsprüfungskommission (GPK) des Nationalrates (NR) vorgenommen. Die Sektion EDI der GPK NR begleitete den Evaluationsprozess verhältnismässig eng. Im Verlauf des Evaluationsprozesses hat sich die Zusammensetzung dieser Sektion mehrmals verändert:

- 1989: - Houmard, Marc-André (FDP/BE) (Präsident)
 - Brügger, Cyrill (SPS/FR)
 - Darbellay, Vital (CVP/VS)
 - Hösli, Fritz (SVP/GL)
 - Meier, Hans (GP/ZH)
 - Wanner, Christian (FDP/SO)
- 1990/91: - Hösli, Fritz (SVP/GL) (Präsident)
 - Brügger, Cyrill (SPS/FR)
 - Etique, Pierre (FDP/JU)
 - Hess, Peter (CVP/ZG)
 - Meier, Hans (GP/ZH)
 - Wanner, Christian (FDP/SO)
- 1991/92: - Brügger, Cyrill (SPS/FR) (Präsident)
 - Friderici, Charles (LPS/VD)
 - Meier, Hans (GP/ZH)
 - Seiler, Rolf (CVP/ZH)
 - Tschopp, Peter (FDP/GE)
 - Wanner, Christian (FDP/SO)

Die Sektion EDI wurde 1992 neu in 'Vertikale Sektion 3, Leistungsstaatliche Aufgaben' umbenannt. Diese Änderung der Bezeichnung stand im Zusammenhang mit dem Leitbild, das sich die GPK versuchsweise gegeben hat und zentral eine neue Struktur der Sektionen (Vertikale und Horizontale Sektionen) vorsieht (vgl. Amtl. Bulletin, Nationalrat 1992: 239-244, Ständerat 1992: 484-485). Der Einfachheit halber wird diese Sektion im folgenden für die ganze Periode mit 'Sektion EDI' bezeichnet.

Das Sekretariat der GPK hat die Projektarbeiten begleitet. Zuständig war für diese Evaluation die stellvertretende Sekretärin, Gabriella Bardin-Arigoni, die heute als stellvertretende Generalsekretärin im EDI tätig ist und der Sekretär der GPK, Philippe Mastronardi, der heute an der Hochschule St. Gallen öffentliches Recht lehrt.

5.3.1.5. Die Forscher

Peter Knoepfel (Jahrgang 1949) studierte in Bern Rechtswissenschaften und promovierte mit einer Arbeit zur 'Demokratisierung der Raumplanung' (Dr. iur., siehe Knoepfel 1977).

Von 1977 bis 1982 war er wissenschaftlicher Mitarbeiter am Internationalen Institut für Umwelt und Gesellschaft am Wissenschaftszentrum Berlin. Dort arbeitete er an einem Forschungsprojekt zu Vollzugsproblemen bei Umweltqualitätsnormen im Rahmen des Forschungsverbundes 'Implementation politischer Programme' (siehe Knoepfel/Weidner 1980c, Weidner/Knoepfel 1983, siehe auch Knoepfel/Weidner 1980a und b, Knoepfel 1981, Knoepfel/Weidner 1985, Knoepfel/Lewanski 1983, Knoepfel/Lundqvist/Prud'homme/Wagner 1987, Weidner/Knoepfel 1985). Anschliessend wurde er Professor am 'Institut des hautes études en administration publique' (IDHEAP) in Lausanne (heute Chavannes-près-Renens). Neben seiner Lehrtätigkeit leitete und bearbeitete er seither eine Vielzahl von verschiedenen Forschungsprojekten mit Schwergewicht in der Umweltpolitik (siehe die zahlreichen Hinweise im Literaturverzeichnis). Weiterhin ist er Vorstandsmitglied der Schweizerischen Vereinigung für Politische Wissenschaft, Mitherausgeber der Zeitschrift 'GAIA - Ökologische Perspektiven in Natur-, Geistes- und Wirtschaftswissenschaften', Präsident der 'Schweizerischen Kommission für Umweltbeobachtung' und Mitglied der Expertengruppe des Nationalen Forschungsprogramms 'Klimaänderungen und Naturkatastrophen' (NFP 31). Seit 1994 ist er zudem Direktor des IDHEAP. Knoepfel hat die hier zur Diskussion stehende Evaluation geleitet. Daneben trug er vor allem als 'Sprecher' des Projekts an Sitzungen mit dem Auftraggeber, mit konzeptionellen und inhaltlichen Beiträgen, mit seiner Teilnahme an verschiedenen Interviews und mit einem massgeblichen Anteil am Hauptbericht zum Projekt bei.

Willi Zimmermann (Jahrgang 1945) hat vor seinem Studium eine Berufslehre als Elektronikgerätemechaniker absolviert. Nach Aufenthalt in London und Paris entschloss sich Zimmermann, die Matura nachzuholen. Nach Abschluss der Eidg. Maturitätsprüfungen besuchte er an der Hochschule St. Gallen das Grundstudium in Volks- und Betriebswirtschaftslehre, Recht und Statistik (1969-1971). Anschliessend wechselte er an die Ludwig-Maximilians-Universität in München, wo er Sozialwissenschaften und Philosophie studierte. Er schloss dort das Studium im Jahre 1979 mit der Promotion ab (Dr. phil.). Seine Dissertation trägt den Titel 'Vom Bewusstsein zum Diskurs. Eine Untersuchung zum Verhältnis von Erkenntnistheorie und politischer Theorie am Beispiel der empirisch-analytischen Theorien von John Locke und David Hume.' In den Jahren 1981/1982 absolvierte Zimmermann sodann den Nachdiplomstudiengang Entwicklungsländer (NADEL) an der ETH in Zürich. 1982 gründete er die 'Forschungsstelle Wissenschaft und Politik' (FWP) in Bern, ein privates Forschungsinstitut, wo er bis heute tätig ist. In diesem Rahmen hat er an einer Vielzahl unterschiedlicher Forschungsprojekte mitgewirkt. Neben den Arbeiten zur Forschungs und Technologiepolitik (Freiburghaus/Zimmermann 1985, 1987, Zimmermann/Kolly/Mahon 1989, Freiburghaus/Zimmermann/Balthasar 1990, Freiburghaus/Balthasar/Zimmermann 1991) bildeten

dabei die Arbeiten im Bereich des Umweltschutzes (Knoepfel/Zimmermann 1987b, Zimmermann 1988, Zimmermann/Aeschlimann 1991, Zimmermann/Balthasar/Neuenschwander 1991, Knoepfel/Imhof/Zimmermann 1994) und speziell des Umweltschutzes (vor allem Gewässerschutz) in der Landwirtschaft Schwerpunkte (Knoepfel/Zimmermann 1987a 1990, 1993, Zimmermann/Knoepfel 1987, Knoepfel/Larrue/Zimmermann 1988, Knoepfel/Zimmermann/Descloux 1989). Seit 1992 ist Zimmermann - neben seiner Tätigkeit an der FWP auch 'Maître d'enseignement et de recherche' am 'Département de science politique' der Universität Genf. Zimmermann übernahm im hier zur Diskussion stehenden Evaluationsprojekt folgende Aufgaben: Teilnahme an den Sitzungen mit dem Auftraggeber, Koordination, konzeptionelle und inhaltliche Beiträge, und massgeblicher Anteil am Hauptbericht.

Giorgio Sailer (Jahrgang 1962) hat an der Universität Bern Volkswirtschaft und Recht studiert und im Jahre 1989 mit der Arbeit 'Impatto e dialogo delle strade nazionali con l'ambiente. Evoluzione lungo i trent'anni di realizzazione nel territorio ticinese' an der Universität Bern (Prof. Dr. Bruno Fritzsche) promoviert (Dr. rer. pol., Sailer 1990). Er bearbeitete verschiedene Forschungsprojekte (touristische Transportanlagen, Luftreinhaltung und Verkehr im Kanton Tessin). Neben seiner forscherschen Tätigkeit war er unter anderem als Journalist, Übersetzer, Mitarbeiter des statistischen Amtes des Kantons Tessin und ist heute auch als Lehrer an einer kantonalen Handels- und Maturitätsschule tätig. Er hat an verschiedenen Teilen der Studie mitgewirkt. Das Hauptgewicht seiner Tätigkeiten lag bei der quantitativen Befragung der Mitarbeiter des BUWAL (SB, Anhänge 7 und 8). Weiter hat er ein Kurzportrait des BUWAL verfasst (SB, Anhang 0).

Enzo Matafora hat an der Universität Genf Politikwissenschaft studiert (lic. ès sc. pol.). Nach Abschluss der hier zur Diskussion stehenden Evaluation hat er - ebenfalls in Genf - das 'Diplôme d'études supérieures en science politique' erworben (dipl. ès sc. pol., 1994). Die Diplomarbeit trägt den Titel 'Les réactions institutionnelles face à la problématique de l'asile. Analyse des interventions au Conseil national de 1973 à 1991.' Bis auf einen Hinweis auf die Mitwirkung an einem Artikel (Knoepfel/Imhof/Matafora 1991) haben wir keine weiteren Publikationen von Enzo Matafora ausfindig machen können. Bei der hier betrachteten Evaluationsstudie arbeitete Matafora in erster Linie an der kantonalen Umfrage (SB, Anhänge 3 und 4). Er verfasste zudem die Fallstudien zu den UVP-Verfahren (SB, Anhang 6).

Jörg Leimbacher hat im Rahmen des Projektes den Teilauftrag Recht bearbeitet, der ein Rechtsgutachten zur Luftreinhalte-Verordnung (LRV) und zur Verordnung über die Umweltverträglichkeitsprüfung (UVPV) beinhaltet (SB, Anhang 10). Leimbacher hat an der Universität Bern mit der Arbeit 'Die Rechte der Natur' in Rechtswissenschaften promo-

viert (Prof. Peter Saladin; Leimbacher 1988, siehe auch Leimbacher/Saladin 1988). Zusammen mit Peter Saladin hat er auch ein Gutachten im Auftrag des Bundesamtes für Umwelt, Wald und Landschaft (BUWAL) zum Katastrophenschutz (Art. 10, Umweltschutzgesetz) erstellt (Leimbacher/Saladin 1990).

Peter Knoepfel nahm bei diesem Projekt mehr die strategischen Aspekte der Projektleitung wahr, während sich Willi Zimmermann mit der Projektleitung auf der operativen Ebene befasste. Wie bereits aus den obigen Hinweisen hervorgeht, weisen Peter Knoepfel und Willi Zimmermann bereits eine langjährige gemeinsame Projekterfahrung auf.

5.3.2. Beschreibung des Evaluationsprozesses

Vorbemerkung: Um der Bedingung zur Wahrung der Anonymität durch die GPK NR nachzukommen, die wir zur Akteneinsicht eingegangen sind, verzichten wir im folgenden verschiedentlich auf die namentliche Erwähnung von Akteuren. Weiterhin musste - aus demselben Grund - auf die Schilderung bestimmter, stark personenorientierter Vorgänge verzichtet werden, so dass die nachfolgende Darstellung gewisse Lücken aufweist.

Im Juni 1988 hatte der Bundesrat den Entscheid getroffen, dass das Bundesamt für Umweltschutz (BUS) und das Bundesamt für Forstwesen und Landschaftsschutz (BFL) auf den 1. Januar 1989 zum Bundesamt für Umwelt, Wald und Landschaft (BUWAL) zusammengelegt werden soll. Bei der Behandlung des Geschäftsberichts des Bundesrates für das Jahr 1988 hatte die Sektion EDI der Geschäftsprüfungskommission des Nationalrates (GPK NR) an ihrer Sitzung vom 3. und 4. April 1989 diese Änderung der Organisationsstrukturen diskutiert. Anlässlich der Plenarsitzung der GPK NR vom 23. August 1989 wurde der Wunsch geäußert, das BUWAL einer Inspektion zu unterziehen. Es wurde entschieden, einen Beschluss erst zu treffen, wenn die neue Zusammensetzung der Sektionen bestimmt sein wird.

An der folgenden Plenarsitzung vom 20. und 21. November 1989 entschied sich die GPK dafür, dass eine Überprüfung des BUWAL vorgenommen werden soll. Die Sektion EDI wurde damit beauftragt, einen Antrag für eine Organisationsüberprüfung auszuarbeiten zu lassen. An einer Sitzung der Sektion EDI der GPK NR vom 5. Februar 1990 wurde über das weitere Vorgehen hinsichtlich der Organisationsüberprüfung des BUWAL diskutiert und beschlossen, dass an die nächste Sitzung der Sektion zwei Experten eingeladen werden, um Projektvorschläge vorzustellen. Am 28. Februar 1990 fand diese Sitzung der Sektion EDI statt, an der Peter Knoepfel und ein weiterer Experte - eine Person mit langer praktischer Verwaltungserfahrung - teilnahmen. Während Knoepfel eine politik- und leistungsorientierte Evaluation vorschlug, die auch die Kantone einbeziehen soll, präsentierte der andere Experte einen Vorschlag zu einer klassischen Organisationsüberprüfung. Die Sektion entschloss sich aufgrund der Präsentationen und wegen Terminproblemen dazu, das Sekretariat damit zu beauftragen, weitere Experten zu kontaktieren und neue Varianten vorzubereiten. Anfangs März 1990 forderte das Sekretariat verschiedene Personen zur Offertstellung auf, wobei diese Bemühungen ergebnislos verliefen. An ihrer Sitzung vom 9. April 1990 hat die Sektion einstimmig beschlossen, der GPK den Antrag zu stellen, eine politikorientierte Analyse durchführen zu lassen und Knoepfel damit zu beauftragen. Das Sekretariat der GPK erarbeitete ein Untersuchungskonzept, das den Untersuchungsbereich, die Fragestellung, die Vorgehensweise sowie den zeitlichen Ablauf und den finanziellen Rahmen grob definierte. Die GPK stimmte an ihrer Sitzung

vom 10. und 11. Mai 1990 dem Antrag der Sektion und der darin enthaltenen Arbeitsgrundlage des Sekretariates einstimmig zu. Am 18. Mai 1990 wurde der andere Experte über diesen Entscheid informiert.

Aufgrund einer Projektskizze von Knoepfel vom 6. Juni 1990, einer Arbeitsgrundlage des Sekretariates vom 11. Juni 1990 sowie den mündlichen Ausführungen von Knoepfel diskutierte die Sektion EDI an ihrer Sitzung vom 28. Juni 1990 die genauere Ausgestaltung der Evaluation. Zentral war dabei insbesondere die Frage, welche Bereiche aus dem weiten Feld der Umweltpolitik einer vertieften Analyse zu unterziehen seien (Luftreinhaltung, Gewässerschutz, Umweltverträglichkeitsprüfung) und in welcher Form die Kantone einbezogen werden könnten. Weiter wurden organisatorische Fragen (Begleitung durch Sektion EDI, Vertretung BUWAL in den Sitzungen usw.) behandelt. Zudem wies Knoepfel darauf hin, dass er verschiedene andere Projekte im Auftrag des BUWAL bearbeite oder bearbeitet habe, was die Sektion jedoch bezüglich Interessenkonflikten als unproblematisch einstufte. Knoepfel wurde damit beauftragt, eine bereinigte Version der Projektskizze zu verfassen und ein Mitglied der Sektion wurde ermächtigt, den Vertrag nach Einsichtnahme in die überarbeitete Projektskizze zu unterzeichnen. Am 10. Juli 1990 lag die überarbeitete Projektskizze (OF) vor. Die Projektarbeiten sollten danach am 1. Oktober 1990 begonnen und Ende September 1991 abgeschlossen werden. Es wurde ein Kreditrahmen von 150'000.- sFr. vereinbart. Neben dem Projektleiter sollte Willi Zimmermann, ein Jurist und zwei Politologen an der Evaluation mitwirken. Die drei letztgenannten Mitarbeiter waren personell zu diesem Zeitpunkt noch nicht bestimmt. Am 10. September 1990 wurde die Auftragsvereinbarung (EK) unterzeichnet.

An den nun folgenden Sitzungen der Sektion nahmen neben den Mitgliedern der Sektion und des Sekretariates jeweils die Evaluatoren und Vertreter des BUWAL teil. Die Sektion pflegte teilweise zu Beginn oder am Schluss der Sitzungen eine sektionsinterne Aussprache (ohne Evaluatoren und ohne Vertreter der Verwaltung) zu führen. An ihrer Sitzung vom 29. Oktober 1990 behandelte die Sektion vorerst allgemeine Fragen (Information der Kantonsregierungen und der Mitarbeiter des BUWAL zum Evaluationsprojekt, Enthebung der BUWAL-Mitarbeiter vom Amtsgeheimnis), um sodann den durch die Evaluatoren vorgelegten Fragebogen an die Kantone zu diskutieren. In einem Schreiben der GPK an die Mitarbeiter des BUWAL und an die Kantonsregierungen vom 13. resp. 14. November 1990 wurde über das Evaluationsprojekt informiert. An der Sitzung der Sektion vom 7. Januar 1991 wurde, nachdem die Evaluatoren einen Bericht über den Stand der Arbeiten abgegeben hatten, schwergewichtig der Fragebogen zuhanden der BUWAL-Mitarbeiter kritisch diskutiert. Am 28. Januar 1991 unterbreiteten die Evaluatoren schriftlich einen Vorschlag zur Auswahl der Kantone für die vertiefende Analyse zu-

handen der Sektion EDI der GPK NR. Diese diskutierte die Kantonsauswahl an ihrer Sitzung vom 29. Januar 1991 und bestimmte die Kantone Waadt, Thurgau, Basel-Stadt und Schwyz als Untersuchungskantone. An der Sitzung der Sektion vom 22. Mai 1991 stellten die Evaluatoren erste Resultate vor und berichteten über den Stand der Arbeiten sowie über erste Ergebnisse. Weiterhin wurde das Rechtsgutachten präsentiert und diskutiert. An der Sitzung vom 8. Juli 1991 der Sektion EDI wurden weitere Resultate der Analysen, einige Empfehlungen und die von den Evaluatoren erarbeiteten Zukunftsszenarien diskutiert. Anfangs September 1991 lag ein erster Entwurf des Schlussberichts (der sogenannte 'Rohbericht') vor. Am 10. bzw. 12. September 1990 gaben Departement bzw. BUWAL eine erste, provisorische Stellungnahme zum Bericht ab. An ihrer Sitzung vom 12. September 1991 diskutierte die Sektion den 'Rohbericht', wobei auch die vorliegenden Stellungnahmen berücksichtigt wurden. Dabei wurde klar, dass einige Teile des Berichtes noch stark zu revidieren und insbesondere eine Entpersonifizierung des Berichtes nötig seien. Es wurde der Entscheid getroffen, dass die Diskussion aufgrund einer überarbeiteten Version des Berichts fortgeführt werden sollte. Am 8. Oktober 1991 legten die Experten die revidierten Berichtsteile vor, die als Grundlage für die Sitzung der Sektion vom 14. Oktober 1991 dienten. An dieser Sitzung wurde entschieden, dass die zwanzig Exemplare des Rohberichts eingezogen und vernichtet werden sollen. Weiterhin wurde durch die Evaluatoren ein Antrag über eine zehnprozentige Erhöhung des Honorars eingebracht. Dieser Antrag wurde am 7. November 1991 gutgeheissen. Am 18. November 1991 lag der Schlussbericht in seiner definitiven Fassung vor.

Das BUWAL gab am 9. Dezember 1991 eine ausführliche Stellungnahme zum Schlussbericht ab. Der Schlussbericht und die Stellungnahme des BUWAL bildeten die Arbeitsgrundlagen für die nächste Sitzung der Sektion der GPK vom 16. Dezember 1991. Weiterhin wurden an diesem Tag auch erste Diskussionen zu den sektionseigenen Empfehlungen geführt, da sich die Sektion dazu entschlossen hatte, dass die GPK einen eigenen Bericht verfassen sollte. Am 9. Januar 1992, an der nächsten Sitzung der Sektion, wurde diese Diskussion fortgesetzt. Ein weiterer Diskussionsgegenstand bildete die Methodik der Evaluation, die kritisch und ausführlich diskutiert wurde. Man hatte zu dieser Frage auch ein Gespräch mit einem nicht-involvierten Experten geführt. Auch an der Sitzung der Sektion vom 3. Februar 1992 wurde schweremotiv der Schlussbericht sowie die sich daraus ergebenden Empfehlungen der GPK diskutiert. An der Plenarsitzung der GPK NR vom 4. Februar 1992 wurde über das Voranschreiten der Arbeiten in der Sektion und über das weitere Vorgehen Bericht erstattet. An der Sitzung vom 23. März 1992 nahm die Sektion weitere Bereinigungen am sektionseigenen Berichtsentwurf vor und debattierte das weitere Vorgehen. Ende März 1992 lag eine bereinigte Fassung des Berichtsentwurfs der Sektion vor. Dieser wurde, zusammen mit dem Schlussbericht der

Evaluatoren und einem Entwurf für eine Zusammenfassung an der Plenarsitzung der GPK NR vom 18. Mai 1992 diskutiert und verabschiedet. Am darauffolgenden Tag wurden die Berichte und die Zusammenfassung an einer Pressekonferenz präsentiert.

5.3.3. Bewertung entlang der Kriterien

Der Entscheid, ob die 'Evaluation des BUWAL' anhand der traditionellen oder der alternativen Kriterien zu bewerten sei, ist schwierig zu treffen und bedarf einer ausführlicheren Diskussion.

Der Schlussbericht enthält einen Hinweis darauf, dass die eingesetzten Leitsätze aufgrund der 'grounded theory'-Methodik erarbeitet wurden (SB: 57). Der Ansatz der 'grounded theory' geht grundsätzlich von einem alternativen Paradigma aus (vgl. Strauss/Corbin 1990: 249-258 und Corbin/Strauss 1990: 424-426). Andererseits ist diese Methodik der kombiniert induktiv-deduktiv angelegten Theoriekonstruktion ('gegenstandsnahe Theoriebildung', siehe Wiedemann 1991) durchaus auch im traditionellen Paradigma einsetzbar (vgl. dazu auch Kapitel 5.3.3.1.4). Dasselbe gilt - mit umgekehrten Vorzeichen - auch für die in der Studie eingesetzten standardisierten Befragungen mit grösstenteils geschlossenen Fragen, die quantitativ ausgewertet werden.

Wichtiger erscheint uns an dieser Stelle die Zielsetzung der qualitativen Interviews zu sein. Diese wurden als Kreuzinterviews konzipiert, um "stark erhärtete Tatsachenaussagen" (SB: 13) generieren zu können. Diese Aussage deutet stark auf eine traditionelle erkenntnistheoretische Position hin. Das gleiche gilt für die Absicht, mit dieser Methode Aussagen zu verifizieren (SB: 13). Eine andere Passage im Schlussbericht weist jedoch darauf hin, dass keine - im engeren Sinne - traditionelle Sichtweise vorliegt: "Eine weitere Grenze besteht darin, dass der empirisch arbeitende Sozialwissenschaftler auch 'objektiv Unwahres' dokumentieren muss. Aus der Sicht eines Spezialisten mögen einige der von uns registrierten Antworten 'an sich' falsch sein; für den Sozialwissenschaftler stellen sie nichtsdestoweniger 'objektive' Sachverhalte dar, weil sie für eine ernstzunehmende Gruppe von Personen für 'wahr' gehalten werden, weshalb sie deren Handeln vermutlich in relevanter Weise mitbestimmen." (SB: 14, vgl. dazu ähnlich auch Knoepfel/Zimmermann 1993: 10)

Wie aus den dargelegten Textpassagen hervorgeht, ist eine schlüssige Zuordnung der Evaluation zu einer der beiden Positionen nicht möglich. In diesem Fall zeigen sich deutlich die Grenzen der Dichotomie der Paradigmen, von der wir ausgegangen sind (siehe dazu unten Kapitel 7.2.). Aufgrund dieser Situation haben wir uns dazu entschlossen, bei den von dieser Zuordnung betroffenen Kriterien auf eine Bewertung zu verzichten (siehe unten die Kapitel 5.3.3.4.5., 5.3.3.4.6. und 5.3.3.4.11).

5.3.3.1. Nützlichkeit

5.3.3.1.1. Identifikation der beteiligten Gruppen

Die Evaluation enthält in ihrer Vorgehensweise bereits die Identifikation einiger 'stakeholders'. Im Rahmen der schriftlichen Umfragen werden alle Mitarbeiter des BUWAL sowie die zuständigen kantonalen Amtsstellen, als beteiligte Gruppen identifiziert. Die standardisierten schriftlichen Befragungen werden durch qualitative Gespräche in vier Kantonen (Waadt, Thurgau, Basel-Stadt und Schwyz) und bei 44 Personen aller Hierarchiestufen aus der Exekutive vertieft. Bei diesen 44 Personen handelt es sich nicht nur um Personen aus dem BUWAL, sondern auch um Vertreter des Departements. Weiter wurden mit Vertretern verschiedener Bundesämter (für Polizeiwesen (BAP), für Verkehr (BAV), für Landwirtschaft (BLW), für Strassenbau (ASB) und für Aussenwirtschaft (BAWI)), die im BUWAL eher als 'schwierige' Partner eingestuft wurden, Gespräche geführt.

Die Evaluation ist sehr eng von der Sektion EDI der GPK NR begleitet worden. Die Sektion ist, wie das den Gepflogenheiten bei parlamentarischen Kommissionen entspricht, entlang dem Parteien- und Sprachenproporz zusammengesetzt (vgl. Kapitel 5.3.1.4.). An der Sitzung der Sektion haben zudem - neben den Evaluatoren selbst - regelmässig oder punktuell weitere Personen teilgenommen. Zu diesem Personenkreis gehören Vertreter des evaluierten Bundesamtes, des Eidg. Departements des Innern, der Parlamentarischen Verwaltungskontrollstelle (PVK) und selbstverständlich waren immer auch Personen aus dem Sekretariat der GPK zugegen.

Daraus wird ersichtlich, dass die beteiligten oder betroffenen öffentlichen Stellen ausreichend identifiziert wurden. Demgegenüber sind private Akteure (Wirtschaftsorganisationen, Umweltverbände, Unternehmen, Privatpersonen (UVP) etc.) nur gerade punktuell - etwa in den Fallstudien - berücksichtigt worden (vgl. SB, Anhänge 5 und 6). Ein weiterer Aspekt, bei dem die Interessen weiterer Akteure indirekt angesprochen wurden, stellt die von den Evaluatoren als Organisationsprinzip für die Sektoralabteilungen vorgeschlagene 'emittentenorientierte Struktur' dar (siehe etwa SB: 132). Die Bedeutung dieser weniger berücksichtigten Akteure sind jedoch für die Evaluation nicht als ausserordentlich zentral einzustufen. Die Evaluation richtete sich ja primär auf das BUWAL selbst, sowie auf dessen Aussenbeziehungen zu anderen Bundesstellen und zu den Kantonen.

Aus diesem Grund kommen wir zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt. Den bestehenden Lücken ist kein allzu grosses Gewicht beizumessen.

5.3.3.1.2. Glaubwürdigkeit des Evaluators

Das Evaluatorenteam, vor allem der Projektleiter Peter Knoepfel, aber auch Willi Zimmermann verfügen über eine sehr grosse Erfahrung in der Bearbeitung von sozialwissenschaftlichen Forschungsprojekten. Beide haben einen ihrer Schwerpunkte in der schweizerischen Umweltpolitik (vgl. dazu die Hinweise in Kapitel 5.3.1.5). Beide können als ausgewiesene Experten für Politikstudien im schweizerischen Umweltbereich bezeichnet werden. Weiterhin haben sie bereits eine grössere Zahl von Projekten gemeinsam bearbeitet. Die beiden Mitarbeiter Giorgio Sailer und Enzo Matafora sind zwar nicht so erfahren, wie die beiden erstgenannten. Aber auch sie verfügen über eine fundierte Ausbildung, welche für die Mitarbeit an diesem Projekt bei weitem ausreicht. Jörg Leimbacher, der das Rechtsgutachten erarbeitet hat, ist ein Experte in Rechtsfragen im ökologischen Bereich. Aufgrund dieser Zusammensetzung kann dem Evaluatorenteam sicherlich eine sehr hohe Kompetenz zugesprochen werden.

Direkt im Zusammenhang mit der erwähnten langen Projekterfahrung von Peter Knoepfel und Willi Zimmermann im Umweltbereich ist an dieser Stelle aber auch auf einen weiteren Umstand hinzuweisen. Beide haben bereits einige Projekte im Auftrag des BUWAL bearbeitet (vgl. Zimmermann/Knoepfel 1987, Zimmermann 1988, Knoepfel/Zimmermann 1990, 1993, Zimmermann/Aeschlimann 1991, BUWAL 1993). Dieser Umstand könnte sich möglicherweise bezüglich der externen Akzeptanz der Evaluatoren in negativer Weise auswirken. Andererseits ist es in der Schweiz bei der Evaluation eines Bundesamtes kaum zu vermeiden, dass lediglich Evaluatorenteam über einen fundierten Erfahrungshorizont im relevanten Politikbereich verfügen, die auch in irgendwelcher Form Kontakte zu dieser Bundesstelle pflegen. Unseres Wissens hat dieser Umstand zu keinen negativen Auswirkungen geführt.

Die Glaubwürdigkeit der Evaluatoren wird dadurch zusätzlich gestärkt, dass sie im Auftrag der GPK NR handelten.

Aus diesen Überlegungen heraus kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt.

5.3.3.1.3. Informationsrahmen und -auswahl

Zur Evaluation wurden die verschiedensten Informationen beschafft (vgl. dazu auch Kapitel 5.3.3.4.4.):

- Die Evaluation kann auf verschiedene, von den Evaluatoren durchgeführte oder parallel zur Evaluation bearbeitete *Projekte* zurückgreifen. Es sind dies namentlich die folgenden Untersuchungen (nach SB: 15-17):
 - Studien zum Umweltschutz (spez. Gewässerschutz) in der Landwirtschaft (Knoepfel/Zimmermann 1987 und 1993),
 - Studie zu den kantonalen Massnahmenplänen zur Luftreinhaltung (Knoepfel/Imhof/Zimmermann 1994),
 - Evaluation der UVP im Kanton Bern (Zimmermann/Balthasar/Neuenschwander 1991, vgl. auch Ledergerber 1992),
 - Evaluation der Kommission zur Förderung der wissenschaftlichen Forschung (KWF), insbes. Umwelttechnologie (Freiburghaus/Zimmermann/Balthasar 1990),
 - Projekte zu den Grundlagen der Umweltforschung (Zimmermann 1988, Zimmermann et al. 1991, Zimmermann/Aeschlimann 1991, BUWAL 1993)
 - Diverse Diplom- und Studienarbeiten am IDHEAP unter Leitung von Peter Knoepfel zur Umweltpolitik
- In der Evaluation wurde eine grosse Zahl von *Dokumenten und Akten* des BUWAL beigezogen. Im Anhang zur Studie finden sich Hinweise auf einen Teil dieser Unterlagen (SB, Anhänge 2 und 9, vgl. SB: 5).
- Die zuständigen *Amtsstellen aller Kantone* wurden in schriftlicher Form befragt (SB, Anhänge 3 und 4). Weiterhin wurden mit Vertretern vier ausgewählter Kantone (Thurgau, Schwyz, Basel-Stadt und Waadt) insgesamt fünfzehn zweistündige Gespräche geführt. Zudem sind auf der kantonalen Ebene Fallstudien zum Gewässerschutz in der Landwirtschaft (Subventionierung von Abwasserreinigungsanlagen (ARA), siehe Anhang 5) und zu drei UVP-Verfahren (siehe Anhang 6) erstellt worden.
- Bei allen *Mitarbeitern des BUWAL* wurde ebenfalls eine schriftliche Befragung durchgeführt. Vertieft wurden die Erkenntnisse aus der quantitativen Analyse durch rund 90 Minuten dauernde Interviews mit 44 BUWAL-Mitarbeitern. Ergänzend wurden Gespräche mit Vertretern verschiedener anderer Bundesämter, mit dem Generalsekretär EDI sowie mit dem Departementsvorsteher geführt. Aufgrund von Informationen aus unseren Gesprächen (siehe Kapitel 5.3.5.2.) sind Leitfadenterviews durchgeführt worden; die Gespräche sind vollständig transkribiert und systematisch ausgewertet worden.
- Als weitere Informationen flossen die Ergebnisse des Rechtsgutachtens zur Luftreinhalteverordnung (LRV) und zur Verordnung über die Umweltverträglichkeitsprüfung (UVPV) in die Evaluation mit ein (SB, Anhang 19).

Zusammenfassend lässt sich festhalten, dass der Evaluation eine sehr breite Informationsbasis zur Verfügung stand. Mit einem kleinen Vorbehalt (vgl. oben Kapitel 5.3.3.1.1.) werden dabei auch die Interessen der Beteiligten und Betroffenen ausreichend miteinbezogen. Aus diesen Gründen kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt.

5.3.3.1.4. Bewertende Interpretationen

Der Bericht enthält eine Darlegung zur Perspektive der Evaluation. In der Einleitung wird der Auftrag, wie auch der politikorientierte Ansatz der Studie angeführt (SB: 1-4). Weiterhin enthält der Bericht die Leitsätze, die zur Analyse und Bewertung der BUWAL-Strukturen beigezogen worden sind (SB: 57-60). Diese sind in diesem Zusammenhang von zentraler Bedeutung, da sie die normativen Orientierungslinien für die Organisationsanalyse bilden.

Das Vorgehen im Rahmen der Evaluation wird im allgemeinen aus den im Schlussbericht (bes. SB: 1-18) bzw. den Anhängen enthaltenen Informationen ausreichend deutlich. In gewissen Teilaspekten bestehen in der Beschreibung Lücken. Dies betrifft insbesondere die oben genannten Leitsätze (siehe SB: 57-60). Der Schlussbericht enthält zwar einige Hinweise, wie diese Leitsätze entstanden sind, diese Darstellung erlaubt es dem Leser jedoch nicht, den Entwicklungsprozess nachvollziehen zu können. Der Hinweis auf die Methodik der 'grounded theory' (SB: 57) vermag nicht zu überzeugen, da diese Vorgehensweise auf die Auffindung von sog. Schlüsselkategorien und weiteren Kategorien grossen Wert legt. Weiterhin sollte bei diesem Ansatz nicht nur der Entstehungsprozess der genannten Kategorien, sondern ebenso deren Verknüpfungen sowie die dazugehörigen empirischen Indikatoren transparent dargelegt werden (vgl. dazu Strauss/Corbin 1990 und Strauss 1994), was hier nicht geschieht. Die deduktiv angelegten Entwicklungsschritte werden ohne konkrete Bezüge angeführt. Der Hinweis, dass sich die Evaluation dabei auf "Erkenntnisse der öffentlichen Managementlehre" (SB: 57) abstützt, reicht in dieser pauschalen Form und ohne jegliche Literaturhinweise nicht aus, um dem Adressaten die theoretischen Grundlagen zugänglich zu machen. Zur Induktion heisst es im Schlussbericht: "Beim induktiven Verfahren werden Daten und Aussagen gesammelt und nach ihrer Häufigkeit kompiliert." (SB: 57) Es fehlen Angaben, die erkennen lassen würden, in welcher Form das 'theoretical sampling' oder die Kodierung (offenes, axiales bzw. selektives Kodieren) erfolgte. Beides sind zentrale Komponenten des 'grounded theory'-Ansatzes (Strauss 1994: 94-115 und Strauss/Corbin 1990: 57-193). Unklar bleibt auch wie Häufigkeiten bei induktivem Vorgehen (das heisst ohne bereits über die (theoretischen) Kategorien zu verfügen) erfasst werden sollen. Zusätzlich ist darauf hin-

zuweisen, dass der 'grounded theory'-Ansatz grösseres Gewicht auf die Zentralität als auf die Häufigkeit von Kategorien legt. (Vgl. dazu auch Strauss/Corbin 1994: 274 und den Artikel von Wiedemann 1991, auf den vermutlich im Schlussbericht (SB: 57) hingewiesen wird)

Der Logik des Berichts folgend, in den Anhängen einen wissenschaftlichen Apparat bereitzustellen, wäre ein zusätzlicher Anhang mit genaueren Erläuterungen zum induktiven und deduktiven Vorgehen unabdingbar. Dies besonders deshalb, da den auf dieser Analyse beruhenden Leitsätzen in der Evaluation eine zentrale Stellung zukommt. Sie nehmen viele der später folgenden Empfehlungen bereits vorweg (vgl. etwa Leitsatz 3: Aufhebung Hauptabteilungen (SB: 58) und Empfehlungen 8 (SB: 74), 15 (SB: 88), 24 (SB: 101) oder Leitsatz 4: Emittentenorientierte Struktur der Sektoralabteilungen (SB: 58) und Empfehlung 6 (SB: 72)).

Im einleitenden Kapitel des Hauptberichts findet sich zu den Leitsätzen ein wichtiger Vorbehalt. Es heisst dort: "Diese [die Leitsätze, TW] lassen sich nur beschränkt aus unseren empirischen Befunden herleiten; soweit darin (erfahrungsgestützte) Wertungen der Evaluatoren eingeflossen sind, sollen diese auf ihre Wünschbarkeit hin überprüft werden, bevor deren konkrete Anwendung im Rahmen des vorgetragenen Reorganisationsvorschlags diskutiert wird." (SB: 18) Wie wir oben gesehen haben, implizieren die Leitsätze weitgehend die nachfolgenden Empfehlungen. Welche Wertungen der Evaluatoren in die Leitsätze eingeflossen sind, ist aus dem Bericht nicht ersichtlich (vgl. auch Kapitel 5.3.3.4.10.). Aufgrund unserer Gespräche (vgl. Kapitel 5.3.5.2.) ist uns bekannt, dass die Leitsätze in einem iterativen Prozess entwickelt worden sind, in dem Informationen aus den Gesprächen mit den Mitarbeitern des BUWAL eine wesentliche Rolle gespielt haben.

Aufgrund der Problematik, die sich im Umfeld dieser Leitsätze ergeben, kommen wir zum Schluss, dass die Evaluation das Kriterium nicht vollständig erfüllt.

5.3.3.1.5. Klarheit des Berichts

Der Schlussbericht ist folgendermassen strukturiert:

BAND 1: HAUPTBERICHT

1. Einleitung (SB: 1-21)

Teil I: Politikleistungen (SB: 22-56)

2. Politikleistungen des kantonalen Vollzugs - Rolle des BUWAL und der bundesrätlichen

Entscheidungen in den Bereichen Luftreinhaltung, ländlicher Gewässerschutz und UVP (SB: 22-41)

3. Politikleistungen des BUWAL und des Bundesrates (SB: 42-56)

Teil II: Verwaltungsprobleme (SB: 57-148)

4. Für die drei Untersuchungsbereiche relevante Struktur-, Führungs- und Personalprobleme innerhalb und ausserhalb des BUWAL (SB: 57-93)
5. Struktur-, Führungs- und Personalprobleme auf der Ebene des gesamten BUWAL bzw. auf der Ebene seiner Aussenbeziehungen (SB: 94-148)

Teil III: Zukunftsperspektiven (SB: 149-164)

6. Szenarien für die längerfristige verwaltungsstrukturelle Konzeption der schweiz. Umweltpolitik auf der Ebene des Bundes (SB: 149-164)

BAND 2: ANHÄNGE

- 'Ein Kurzportrait des BUWAL' (Anhang 0)
- Projektskizze (Anhang 1)
- Zusammenstellung des Verwaltungsausstosses des BUWAL im Bereich der Luftreinhaltung seit 1986 (Anhang 2)
- Umfrage bei kantonalen Fachstellen und Direktionen zur Umweltpolitik des Bundes: Fragebogen (Anhang 3)
- 'Das Buwal aus der Sicht der Kantone'. Kantonale Umfrage im Rahmen der BUWAL-Evaluation (Anhang 4)
- Darstellung von vier Subventionsverfahren für ARA in den GPK-Testkantonen (Anhang 5)
- Darstellung von drei UVP-Verfahren mit Anhörungsverfahren (Kantone SZ, BS und VD) (Anhang 6)
- Umfrage bei den Mitarbeiterinnen und Mitarbeitern des BUWAL: Fragebogen (Anhang 7)
- Umfrage bei den Mitarbeiterinnen und Mitarbeitern des BUWAL: Statistische Auswertung der Ergebnisse (Anhang 8)
- Zusammenstellung der Beschlüsse des Bundesrates zur Luftreinhaltungspolitik (Anhang 9)
- Evaluation des BUWAL. Teilauftrag Recht (Anhang 10)

Wie aus dieser Inhaltsübersicht ersichtlich wird, enthält der Hauptbericht (Band 1) schwergewichtig die Befunde und Empfehlungen der Evaluation, während die Anhänge (Band 2) über die verschiedenen Untersuchungsteile informieren. Im Schlussbericht heisst es dazu: "Die vorliegende Schrift [der Hauptbericht, TW] ist kein wissenschaftliches Dokument, sondern ein Bericht über die Ergebnisse einer mit den Mitteln der Verwaltungswissenschaft durchgeführten Untersuchung. Die wissenschaftlichen Grundlagen und Materialien befinden sich in den Anhängen." (SB: 7) Bei der Strukturierung des Berichts hat man sich somit dazu entschlossen, über die empirischen Analysen im Anhang Bericht zu erstatten, um den Hauptbericht davon zu entlasten. Die Reichhaltigkeit der verschiedenen Untersuchungsteile der Evaluation hat dabei sicher eine Rolle gespielt, umfasst doch der Anhang über 200 Seiten. Diese Trennung hat aber auch ihre Nachteile. So werden teilweise Wiederholungen notwendig und vor allem ist es teilweise mit Schwierigkeiten verbunden, die Bezüge zwischen den Aussagen im Hauptbericht und den Angaben im Anhang herzustellen (vgl. dazu auch Kapitel 5.3.3.4.8.). Trotzdem ist der gewählte Aufbau des Berichts in Anbetracht des Umfangs des Anhangs zweckmässig.

Die Struktur des Hauptberichts entspricht der Vorgehensweise einer politikorientierten Verwaltungsanalyse. Die Kapitel zu den Politikleistungen (Kapitel 2 und 3, SB: 22-41

und 42-56) und eines der Kapitel zu den Verwaltungsproblemen (Kapitel 4, SB: 57-93) sind in die drei Untersuchungsbereiche (also Luftreinhaltung, ländlicher Gewässerschutz und Umweltverträglichkeitsprüfung (UVP)) aufgeteilt. Das führt einerseits dazu, dass die oben angesprochene analytische Vorgehensweise in den Vordergrund gestellt wird. Dies ist aber auch mit dem Nachteil verbunden, dass die thematischen Blöcke (also etwa jener zur Luftreinhaltung) aufgetrennt werden müssen. Bei der Gestaltung des Schlussberichts wurde somit dem analytischen Muster gegenüber der inhaltlichen Ebene Priorität eingeräumt. Dieser Entscheidung ist unseres Erachtens durchaus vertretbar. Es kommt hinzu, dass sich thematisch spezifisch interessierte Leser anhand des Inhaltsverzeichnisses problemlos orientieren können.

Der Hauptbericht verfügt über eine sinnvolle Feinstrukturierung und ist - im Gegensatz zu gewissen Passagen in den Anhängen - sprachlich sorgfältig verfasst und sauber redigiert. Der Bericht verfügt über ein Inhaltsverzeichnis (SB: I-III) und ein Verzeichnis der Abkürzungen (SB: IV-VI). Im Bericht selbst ist keine Zusammenfassung enthalten, wobei der Bericht der GPK, der aufgrund der Evaluation von der GPK erstellt wurde, eine solche enthält (GPK 1992: 4-16). Der Bericht enthält kein Literaturverzeichnis. Im Text finden sich einzelne Literaturangaben, nämlich die Hinweise auf die 'Vorarbeiten' der Evaluatoren (SB: 15-17), auf einige Dokumente des BUS/BUWAL und gesetzgeberische Erlasse (SB, Anhänge 2 und 9; aber auch SB: 36), auf die 'grounded theory' (SB: 57, Kardorff bzw. Flick et al. 1991), auf Beiträge in der Zeitschrift 'Umweltrecht in der Praxis' im Rechtsgutachten (SB, Anhang 10: 7) und auf weitere Texte (wie Baroni et al. 1991, siehe SB: 46 oder Knoepfel 1991a, SB: 34-35). Ein Literaturverzeichnis hätte die Übersicht über die beigezogenen Quellen und die konsultierte Literatur verbessern können.

Die methodische Berichterstattung verteilt sich auf verschiedene Berichtsteile. Neben den Ausführungen in der Einleitung (SB: 1-18) und einzelnen Hinweisen in anderen Textteilen (SB: 57), enthalten vor allem auch verschiedene Anhänge (vor allem Anhänge 3, 4, 7 und 8) methodische Angaben. Für methodisch interessierte Leser ist diese Verzettelung nicht von Vorteil. Sie birgt auch die Gefahr in sich, dass es in der methodischen Berichterstattung zu Redundanzen oder Lücken kommen kann. Eine konzentrierte Darstellung würde den Überblick über die methodischen Aspekte erleichtern. Möglicherweise hätte sich dies in einem Anhang realisieren lassen.

Graphische Hilfsmittel werden im Hauptbericht vor allem in drei Bereichen eingesetzt: Erstens dienen sie zur Veranschaulichung von organisatorischen Strukturen bzw. von Vorschlägen zu deren Änderung oder der Interaktionen darin (für Beispiele vgl. SB: 19, 62, 68, 77, 82, 131). Zweitens symbolisieren die Graphiken die in der Evaluation präsen-

tierten Zukunftsszenarien (vgl. SB: 151-163). Drittens werden in Säulen- und Kuchen-
diagrammen die Ergebnisse der quantitativen Umfragen dargestellt (so SB: 26, 63-65, 70,
101, 107-110, 114, 117, 124). Der Bericht enthält auch eine Übersichtskarte zur Aus-
scheidung der Gewässerschutzzonen (SB: 33). Der Einsatz graphischer Mittel im Haupt-
bericht ist in geeigneter Weise erfolgt.

Der Hauptbericht enthält nur verhältnismässig selten Tabellen (vgl. SB: 67, 73 oder 137).
Gerade an einer Stelle des Berichts hätte der Text mit Hilfe einer Tabelle von der (etwas
schwerfällig wirkenden) Wiedergabe quantitativer Umfrageergebnisse entlastet werden
können (Ziff. 30., SB: 38).

Zusammenfassend lässt sich festhalten, dass es gelungen ist, die verhältnismässig um-
fangreichen Informationen in einer übersichtlichen und klar strukturierten Form darzu-
stellen. Der Bericht ist wegen seines Umfangs, seiner inhaltlichen Dichte und der recht
komplexen Thematik (mit vielfachen Interdependenzen) keine 'leichte Kost'. Wir sind
aber der Ansicht, dass er sich durch eine hohe Klarheit auszeichnet. Die seltenen (und re-
lativ unbedeutenden) kritischen Aspekte betreffen vor allem die methodische Berichter-
stattung und das fehlende Literaturverzeichnis. Wir kommen aufgrund dieser Feststellun-
gen zum Schluss, dass die Evaluation das Kriterium erfüllt.

5.3.3.1.6. Verbreitung des Berichts

Die zwei Teile des Schlussberichts, der Hauptbericht und der Anhang sind in
unterschiedlicher Weise verbreitet worden. Der Hauptbericht ist bei der Eidgenössischen
Drucksachen- und Materialzentrale (EDMZ) publiziert worden (Art.-Nr. 104.850.d) und
kann dort für einen Preis von 20.- sFr. erworben werden. Er ist ausschliesslich in deut-
scher Sprache erhältlich. Der Anhang ist nicht publiziert worden, kann aber über die
Parlamentsdienste oder beim IDHEAP eingesehen oder bezogen werden, wobei der
Hauptbericht jedoch keine expliziten Hinweise auf diese Zugangsmöglichkeiten enthält.

Die Resultate der Evaluation sind in den Bericht der GPK eingeflossen und werden dort
summarisch referiert (GPK 1992: 4-16). Der Bericht der GPK ist an einer Pressekonfe-
renz vorgestellt worden, an der auch eine Zusammenfassung des GPK-Berichtes abgege-
ben wurde. Der GPK-Bericht ist im Bundesblatt (BBl 1992 III: 1502-1537) veröffentlicht
worden und enthält ausführliche Hinweise auf die Evaluation und eine explizite Angabe
der Bezugsquelle (GPK 1992: 36).

Über die Evaluation wurde in der Presse verhältnismässig ausführlich berichtet (vgl. die
Tagespresse vom 20. Mai 1992).

Aufgrund der Charakteristik des Auftraggebers ist für eine publikumswirksame Verbreitung des Berichts gesorgt. Der Anhang ist zwar nicht publiziert worden, der Zugang ist jedoch für Interessierte auch dort sichergestellt. Die Verbreitung des Anhangs hätte durch einen Hinweis auf die Bezugsquelle im Hauptbericht verbessert werden können. Dieser Punkt wird jedoch nicht als bedeutend eingestuft, weshalb wir zum Schluss kommen, dass die Evaluation das Kriterium erfüllt.

5.3.3.1.7. Rechtzeitigkeit des Berichts

Die Auftragsvereinbarung enthält zum Abschlusstermin folgende Festlegung: "Die Übergabe des Berichts hat spätestens im September 1991 zu erfolgen." (EK: 3) Die Offerte enthält den Hinweis, dass die Evaluation Ende September 1991 abgeschlossen werden soll (OF: 14).

Die Evaluatoren haben den sogenannten 'Rohbericht' am 2. September 1991 fertiggestellt. Dieser musste jedoch noch stark überarbeitet werden. Die ersten vier Kapitel lagen am 8. Oktober 1991 in überarbeiteter Fassung vor. Die übrigen Kapitel (mit den Anhängen) sind bis zum 18. November 1991 bereinigt worden. Aufgrund von Hinweisen aus der Diskussion mit der Sektion EDI der GPK NR sind nach diesem Termin noch weitere kleinere Anpassungen erfolgt. Die vom Auftraggeber geforderten Überarbeitungen haben eine Erweiterung des finanziellen Rahmens um 15'000.- auf gesamthaft 165'000.- sFr. ausgelöst. Eine zeitliche Verlängerung des Auftragsverhältnisses ist in den uns zugänglichen Unterlagen nicht erwähnt. Aus diesen Überlegungen heraus kommen wir zur Feststellung, dass der Bericht nicht ganz zeitgerecht fertiggestellt werden konnte. Da die nach dem 18. November 1991 vorgenommenen Änderungen minimaler Art waren, hat die Verzögerung *maximal* eineinhalb Monate betragen. Verglichen mit der vorgesehenen Gesamtdauer des Projekts von zwölf Monaten, ist dies keine allzu grosse Verzögerung.

Betrachten wir die im Anschluss an die Fertigstellung des Berichts erfolgten Aktivitäten, lässt sich folgendes sagen: Die Sektion hat die bereits vor dem Abschlusstermin des Evaluatoren-Berichts aufgenommenen Arbeiten am GPK-Bericht weitergeführt und Ende März 1992 abgeschlossen. Im Anschluss daran hat die Gesamt-GPK den Sektionsbericht an ihrer Sitzung vom 18. Mai 1992 diskutiert und verabschiedet. Im Bericht der GPK heisst es zur weiteren Vorgehensweise: "Die Geschäftsprüfungskommission ersucht den Bundesrat, ihr bis Ende 1992 mitzuteilen, welche Konsequenzen er aus diesem Bericht zu ziehen gedenkt." (GPK 1992: 35) Der Bundesrat hat sich nach verschiedenen Verzögerungen am 7. März 1994 - also rund fünfzehn Monate später als erwartet - zum GPK-Bericht geäußert. In dieser Beziehung ist die in der Evaluation zustandegekommene Verzö-

gerung von eineinhalb Monaten als unwesentlich zu beurteilen. Aus diesem Grund und weil der Rohbericht zeitgerecht vorlag, sind wir der Ansicht, dass die Evaluation das Kriterium - trotz kleiner Verspätung - knapp erfüllt.

5.3.3.1.8. Wirkung der Evaluation

Es ist an dieser Stelle nicht möglich, eine abschliessende Bewertung der Wirkungen der Evaluation zu leisten. Wir müssen uns darauf beschränken, einige Wirkungsaspekte zu thematisieren.

Die Evaluation ist im Auftrag der GPK NR bearbeitet worden. Die Sektion EDI dieser Kommission hat sich im Verlauf der Evaluation dazu entschlossen, den Bericht der Evaluatoren nicht als den eigenen zu übernehmen, sondern aufgrund des Evaluationsberichts selber einen Bericht zu verfassen. Vergleichen wir diese beiden Berichte auf der Ebene der gemachten Empfehlungen, kommen wir zu folgenden Feststellungen (siehe GPK 1992: 33-35):

- Verschiedene Empfehlungen der Evaluatoren werden von der GPK übernommen, so jene zu einem departementsübergreifenden Organ im Bereich der Umweltpolitik (Empfehlung 6.1.1. der GPK, vgl. Empfehlung 10, SB: 76), zur Aufgabenumverteilung in umweltpolitischen Verkehrsfragen (Empfehlung 6.1.2. der GPK, vgl. Empfehlung 9, SB: 75-76), zur Einführung einer Matrixorganisation mit Querschnittsabteilungen und Sektoralabteilungen (Empfehlung 6.2.1. der GPK, vgl. Empfehlungen 17, 18 und 25, SB: 91, 92, 102), zur emittentenorientierten Strukturierung der Sektoralabteilungen (Empfehlung 6.2.2. der GPK, vgl. Empfehlung 6, SB: 72), zur klaren Abgrenzung strategischer und operativer Kompetenzen mit sich daraus ergebenden Teilempfehlungen (Empfehlung 6.2.3. der GPK, vgl. Empfehlungen 8, 15, 24, 31, SB: 74, 88, 101, 111-112), zur Ausgliederung der Landeshydrologie und -geologie und zur Bildung einer neuen Institution zur Umweltbeobachtung (Empfehlung 6.3 der GPK, vgl. Empfehlungen 12, 13, SB: 86, 87), zur Zukunft des Vollzugs der Umweltpolitik (Empfehlung 6.4 der GPK, vgl. SB: 149-164) und zur Bereinigung von Widersprüchen und Unklarheiten in der LRV und der UVPV (Empfehlung 6.5 der GPK, vgl. Empfehlungen 2, 4, SB: 45, 55).
- Bei einer Empfehlung ist die GPK der Evaluation nur teilweise gefolgt. In Empfehlung 22 der Evaluation (SB: 98) wird vorgeschlagen, die auf die Nutzung des Waldes konzentrierten Sektionen der Forstdirektion aus dem BUWAL auszugliedern und die mit dem Schutz des Waldes befassten Sektionen in die Abteilungen 'Naturschutz' bzw. 'Landschaftsschutz' einzugliedern. Die Evaluation fügt diesem Vorschlag eine Even-

tualempfehlung an, für den Fall, dass der erstgenannte Vorschlag aus "politischen Opportunitätsgründen" (SB: 98) abgelehnt werden sollte. Dieser Vorschlag beinhaltet die Bildung zweier neuer Abteilungen 'Waldschutz' und 'Waldnutzung' im BUWAL, anstelle der bestehenden Hauptabteilung. Diesem Eventualvorschlag folgt die GPK (Empfehlung 6.1.3.).

- Die GPK hat in ihrem Bericht eine zusätzliche Empfehlung eingefügt, die im Evaluationsbericht nicht enthalten war. Dieser Vorschlag betrifft die Einrichtung einer ausserparlamentarischen Umweltkommission (Empfehlung 6.2.4.). Die Idee zu einer derartigen Kommission ist erst verhältnismässig spät - nach Abschluss der Evaluation - von einem Mitglied der Sektion zur Sprache gebracht und anschliessend von der Sektion aufgenommen worden.

Aufgrund dieses Vergleichs lässt sich festhalten, dass die Empfehlungen der Evaluatoren bei der GPK NR auf grosse Akzeptanz stiessen. Die zentralen Postulate der Evaluation wurden von der Kommission übernommen.

Die Evaluation stiess jedoch nicht überall auf ein derart positives Echo. Die Vertreter der Exekutive und der Verwaltung waren den vorgebrachten Empfehlungen gegenüber deutlich kritischer eingestellt. Sie waren der Ansicht, dass das BUWAL sicher auch gewisse Schwächen aufweise, dass diese jedoch erstens geringfügiger Art waren und dass sie zweitens keinen so grossen Reformbedarf implizieren, sondern mit kleineren Modifikationen zu beheben wären. Die Ursachen für die Schwächen sahen die Verwaltungsvertreter primär in der wachsenden Zahl von Aufgaben, mit der die Ressourcenentwicklung in keiner Weise Schritt halten konnte. Wegen der wachsenden Diskrepanz zwischen Aufgaben und Mitteln war es dem Amt nicht möglich, alle Forderungen zu erfüllen, die an es herangetragen wurden. Ein weiterer Grund für gewisse Schwachstellen lagen gemäss der Ansicht der Verwaltung in der räumlichen Verzettlung des Amtes. Bei Abschluss der Evaluation war das BUWAL insgesamt an vier verschiedenen Orten untergebracht. Versucht man die Gründe für die eher ablehnenden Reaktionen zu eruieren, ist auf verschiedene Aspekte hinzuweisen:

- Die Evaluatoren haben anfangs September 1991 den sogenannten *Rohbericht* fertiggestellt. Dieser Bericht ist an die Sektion EDI der GPK NR, an das Sekretariat der GPK und - ohne dass die Evaluatoren dies gewusst hätten - an die Verwaltung (Departement und BUWAL) gegangen. Dieser Bericht war wirklich relativ 'roh', er enthielt einige sehr kantige Formulierungen, wie auch äusserst deutliche Hinweise zu personellen Belangen. Dieser Bericht hat in der Verwaltung einiges an Staub aufgewirbelt und ver-

schiedene Personen stark beunruhigt. Dies hat die Bereitschaft zu einer positiven Aufnahme der Studie wohl kaum erhöht.

- Auch der definitive Schlussbericht enthält einige *unbequeme Forderungen* für Personen, die sich mit der bisherigen Arbeit des Amtes persönlich identifizierten. Die Befolgung der Gesamtheit der Empfehlungen hätte zu einer fundamentalen Reorganisation geführt. Verschiedene Personen, insbesondere die Leiter der Hauptabteilungen würden ihre Position verlieren, wenn die Hauptabteilungen aufgelöst werden würden. Die Verkleinerung der Direktion auf drei Mitglieder hätte denselben Personenkreis betroffen, da die Hauptabteilungsleiter bis anhin auch Mitglieder der Direktion waren. Der Bericht sieht zudem auch Änderungen vor, die auch weitere Personen und deren Interessen im BUWAL (aber auch ausserhalb des Amtes und des Departements) betreffen. Die Evaluation trägt an das BUWAL Empfehlungen zu einer *grundlegenden Organisationsreform* heran. Diese Forderungen nach einer radikalen Veränderung lösten bei den Betroffenen Verunsicherung aus.
- Die von den Evaluationsempfehlungen *betroffenen Personen* wurden zwar im Rahmen der quantitativen und teilweise auch durch die qualitativen Interviews am Evaluationsprozess beteiligt. Vertreter des Amtes, wie auch des Departements konnten den Fortgang der Evaluation auch als Teilnehmer an den Sitzungen der Sektion EDI der GPK NR (bzw. auf dem Korrespondenzweg) mitverfolgen. Die Evaluation kann aber deswegen nicht als partizipativ konzipierte Evaluation bezeichnet werden. Dafür war der Einfluss und der Beteiligungsgrad der Betroffenen zu gering. Die Studie ging aber diesbezüglich deutlich weiter, als dies im Rahmen von Inspektionen der GPK ansonsten üblich ist.
- Die Evaluation ist im *Auftrag der GPK NR* erfolgt. Dadurch ergeben sich hinsichtlich der Wirkungen verschiedene Konsequenzen. Erstens kam der Anstoss zur Evaluation nicht von der untersuchten Organisation selbst, sondern erfolgte von aussen. Die Untersuchung wurde nicht vom BUWAL gewünscht. In der vorliegenden Form war sie wahrscheinlich sogar - zumindest bei einem Teil des Amtes - unerwünscht. Zweitens hatte die Evaluation, weil sie im Auftrag des Parlamentes erfolgte, aber auch ein anderes Gewicht erhalten, als dies für andere Auftragsverhältnisse gelten würde. Der Auftraggeber hatte erstens eine starke Legitimation zur Durchführung einer solchen 'Überprüfung', ist die Oberaufsicht über die Exekutive doch eine zentrale Aufgabe des Parlamentes. Zweitens beinhaltet dies auch, dass eine parteipolitisch breite Abstützung des Auftraggebers bestand, die das Gewicht der Untersuchung erhöhte. Drittens war dadurch die Möglichkeit gegeben, dass das Parlament auch die Umsetzung kritisch beobachten und nötigenfalls Druck ausüben konnte. Diese Möglichkeiten wurden von der horizontalen Sektion 4 (Wirksamkeit) der GPK NR auch genutzt, indem sie sich

bei ihrer Inspektion zur Kohärenz der Luftreinhaltepolitik vom Oktober 1992 bis März 1994 auf die Evaluation des BUWAL abstützte und verschiedene bereits dort gemachte Empfehlungen wiederum aufgriff (GPK 1994, bes. 20-21 (Empfehlungen 1a., 2 und 6b)).

- Die in der Evaluation eingesetzte *Methodik der Befragung* hatte zu Einwänden methodischer Art geführt. So wurde diese Methodik von Bundesrat Flavio Cotti, dem damaligen Vorsteher des Eidgenössischen Departements des Innern (EDI), vor der Presse als "Umfragen nach dem Stil der Boulevardzeitungen" bezeichnet (TA, 20. Mai 1995, vgl. auch Berner Zeitung, 20. Mai 1995). Aus dieser Äusserung wird zwar eine negative Konnotation deutlich, ob Cotti damit die Absicht verfolgte, die Methodik der Befragung im allgemeinen zu disqualifizieren, ob er mit dieser Aussage primär die Art der spezifischen Umsetzung in dieser Evaluation kritisieren wollte oder ob es vielmehr darum ging, die missliebigen Ergebnisse der Evaluation durch eine vorgeschobene Methodenkritik zu diskreditieren (wie dies die Evaluatoren vermuten, siehe ASSP-Bulletin, Nr. 76 (Oktober 1992): 17), kann aufgrund der uns zur Verfügung stehenden Informationen nicht beurteilt werden.
- Für die Evaluation wurden drei *Untersuchungsbereiche* (Luftreinhaltung, ländlicher Gewässerschutz und Umweltverträglichkeitsprüfung (UVP)) ausgewählt. Auch wenn sich die Evaluation in bestimmten Teilen auch mit anderen Bereichen befasste - so etwa bei den quantitativen und qualitativen Mitarbeiterbefragungen beim BUWAL -, konzentrieren sich die vertiefenden Analysen auf die drei oben angeführten Bereiche. Dies wiederum lässt natürlich die Kritik aufkommen, dass aufgrund einer Untersuchung dreier spezifischer Bereiche keine Aussagen über das Amt als Ganzes möglich seien. Diesem Faktor hat die Evaluation in gewisser Weise Rechnung getragen, indem sie die Befunde streng getrennt dargelegt hat (vgl. Kapitel 4 bzw. 5, SB: 57-93 bzw. 94-148) und indem versucht wurde, möglichst unterschiedliche Bereiche in die Evaluation miteinzubeziehen. Trotzdem kann man dem Einwand der unvollständigen empirischen Abstützung nicht jede Berechtigung absprechen (vgl. dazu Kapitel 5.3.3.4.10.).
- Die in der Evaluation empfohlene *Trennung von Schutz und Nutzung* beim Wald (Empfehlung 22, SB: 98, vgl. auch Leitsatz 5, SB: 58) wird aus grundsätzlichen Erwägungen abgelehnt. Es wird die Meinung vertreten, dass Schutz- und Nutzungsfunktionen zu integrieren seien und dass sich die bisherige Lösung bewährt habe, so dass Änderungen weder angebracht noch nötig erschienen. In diesem Beispiel zeigt sich das Problem, dass die in der Evaluation formulierten Leitsätze für die Evaluation zentral

sind. Wird ein Leitsatz völlig abgelehnt, sind die durch diesen implizierten Empfehlungen nicht aufrechtzuerhalten (vgl. Kapitel 5.3.3.4.10.).

Neben diesen Punkten ist aber auch die weitere Entwicklung nach Abschluss der Evaluation zu beachten. Philippe Roch hat Bruno Böhlen im Juni 1992 an der Spitze des BUWAL abgelöst. Im März 1993 ist Ruth Dreifuss in den Bundesrat gewählt worden und hat von Flavio Cotti das EDI übernommen. Das BUWAL hat auf Anfang Juli 1994 eine neue Organisationsstruktur erhalten. Vergleicht man das heute gültige Organigramm mit dem vorhergehenden sowie mit den Evaluationsempfehlungen, stellt man fest, dass einige der Empfehlungen aufgenommen wurden. Dass dabei auch die oben erwähnten personellen Wechsel eine Rolle gespielt haben, ist zu vermuten, ohne weitere Abklärungen aber nicht definitiv zu beurteilen. Die Änderungen in der BUWAL-Struktur betreffen unter anderem folgende Bereiche:

- Die Matrixform ist neues Organisationsprinzip, dem auch bei weiteren Strukturänderungen gefolgt werden soll.
- Die Amtsdirektion umfasst heute nur noch drei Personen, den Direktor, einen Stellvertretenden und einen Vizedirektor.
- Die Hauptabteilungen 'Zentrale Dienste' und 'Immissionsschutz' sind aufgehoben worden. Die betreffenden Abteilungen sind heute der Direktion direkt unterstellt.
- Die Abteilungen 'Gewässerschutz und Fischerei' und 'Abfall' sind nicht mehr Teil der Hauptabteilung 'Wasser und Bodenschutz' bzw. neu der Hauptabteilung 'Stoffe, Boden, Sicherheit', sondern sind der Direktion direkt unterstellt.
- Der Rechtsdienst ist zu einer Abteilung aufgewertet worden. Das Gleiche gilt für die ehemalige 'Stabsstelle Internationales'. Der 'Informationsdienst' heisst neu 'Sektion Kommunikation'. Es ist vorgesehen, diese Sektion in absehbarer Zeit ebenfalls zu einer Abteilung aufzuwerten.

Im Gegensatz zu den Empfehlungen in der Evaluation bleiben die Hauptabteilungen 'Natur- und Landschaftsschutz', 'Wald' (Eidg. Forstdirektion) und 'Landeshydrologie und -geologie' unverändert bestehen. Betrachtet man die personellen Aspekte dieser Reorganisation, stellt man fest, dass möglicherweise auch diese Hauptabteilungen später einmal aufgelöst werden könnten. Der ehemalige Leiter der (aufgelösten) Hauptabteilung 'Zentrale Dienste' ging in Pension. Sein Kollege bei der Hauptabteilung 'Immissionsschutz' wurde neu als Vizedirektor vollamtliches Mitglied der Direktion. Vermutlich sind bei den zurzeit noch bestehenden Hauptabteilungen ebenfalls Pläne zu deren Auflösung vorhanden, wobei die Lösung der damit zusammenhängenden perso-

nellen Fragen abgewartet wird. Gesamthaft betrachtet, stellt man aufgrund des Vergleichs fest, dass doch eine grössere Anzahl von Empfehlungen der Evaluation bei der erfolgten Reorganisation beachtet wurden. Die Frage, ob die Reorganisation jedoch als direkte Folge der Evaluation betrachtet werden kann, muss an dieser Stelle offen bleiben, da uns zu deren Beantwortung zu wenig Informationen vorliegen.

Im wissenschaftlichen und im internationalen Kontext ist weiterhin auf ein Referat bzw. einen Artikel hinzuweisen, die von Peter Knoepfel aufgrund der Erfahrungen in der Evaluation gehalten bzw. verfasst wurden (Knoepfel 1993b). Sie befassen sich vor allem mit den im Rahmen der Evaluation entwickelten Zukunftsszenarien (SB: 149-164).

Wie bereits eingangs erwähnt, müssen wir auf eine abschliessende Bewertung der Wirkungen der Evaluation verzichten.

5.3.3.2. Anwendbarkeit

5.3.3.2.1. Praktische Verfahren

Im Rahmen der Evaluation werden verschiedene Verfahren eingesetzt. Im Vordergrund stehen dabei quantitative und qualitative Befragungen und die Dokumentenanalyse. Zur Auswertung der erhobenen Daten sind im quantitativen Bereich einfache Häufigkeitsauszählungen (mit Subsample-Vergleichen) vorgenommen worden. Im qualitativen Bereich ist gemäss den Angaben im Schlussbericht mit dem 'grounded theory'-Ansatz gearbeitet worden (vgl. dazu Kapitel 5.3.3.4.9.). Zwei Verordnungen des Bundes wurden in einem Rechtsgutachten untersucht (siehe SB, Anhang 10).

Die in den quantitativen Befragungen eingesetzten Fragebogen sind verhältnismässig umfangreich, umfassen sie doch im Falle der kantonalen Befragung bis zu 39 Fragen (vgl. SB, Anhang 3), im Falle der Befragung beim BUWAL bis zu 59 Fragen (vgl. SB, Anhang 7). Diese Befragungen sind flächendeckend, das heisst bei allen Kantonen und bei allen Mitarbeitern des BUWAL, durchgeführt worden. Der Rücklauf ist als zufriedenstellend bis gut zu bezeichnen (Kantone 68%, BUWAL 82%, SB: 10-11).

Die qualitativen Interviews sind zur Vertiefung eingesetzt worden. In vier ausgewählten Kantonen wurden insgesamt fünfzehn, rund zweistündige Gespräche durchgeführt. Beim BUWAL wurden 44 Personen in rund neunzig Minuten dauernden Gesprächen befragt. Weiter wurden ergänzende Gespräche bei anderen Bundesämtern und beim Departement geführt (vgl. Kapitel 5.3.3.1.3.).

Der Umfang des für die Dokumentenanalyse untersuchten Materials kann aufgrund der Angaben im Schlussbericht nicht restlos erschlossen werden (vgl. SB, Anhänge 2 und 9, vgl. SB: 5).

Die Auswertung des quantitativen Datenmaterials (vor allem Häufigkeitsauszählungen) ist verhältnismässig einfach gestaltet worden. Bei der qualitativen Analyse bestehen unsererseits Unsicherheiten bezüglich der verwendeten Methodik (vgl. Kapitel 5.3.3.4.9.).

Insgesamt betrachtet ist in der Evaluation eine grosse Fülle von Informationen beschafft worden, die als empirische Grundlage sicherlich ausreicht. Eigentliche Störungen des Evaluationsprozesses sind keine zu verzeichnen. Es stellt sich an dieser Stelle als Randbemerkung höchstens die Frage, ob in Anbetracht der zur Verfügung stehenden Ressourcen für die Datenauswertung nicht eher eine zu grosse Datenmenge erhoben wurde.

Aufgrund der Breite der eingesetzten Verfahren und der verschiedenen Informationsquellen stehen der Evaluation ausreichende Informationen zur Verfügung. Zudem ist die Evaluation dadurch auch gegen Störungen wenig anfällig. Die Evaluation erfüllt das Kriterium.

5.3.3.2.2. Politische Überlebensfähigkeit

Wir haben bereits oben (Kapitel 5.3.3.1.8.) ausführlich erläutert, dass die Evaluation einerseits bei der GPK NR - dem Auftraggeber - auf ein recht positives Echo gestossen ist und andererseits die Reaktion der Verwaltung (vorerst) eher als skeptisch oder gar ablehnend zu bezeichnen ist. Angesichts der Thematik - die Evaluation befasst sich mit einer Organisation - ist die Akzeptanz gerade bei den Mitgliedern dieser Organisation und den übergeordneten Instanzen von grosser Bedeutung. Andererseits war es die Aufgabe der Evaluation, ein Bundesamt einer Überprüfung zu unterziehen. Dies wiederum verlangt nach einer deutlichen Artikulation der vorgefundenen Schwächen (wie auch Stärken, vgl. Kapitel 5.3.3.3.7.). Die Evaluatoren stehen in einem Dilemma, entweder die Forderungen des Auftraggebers zu erfüllen oder den Konsens mit der zu evaluierenden Organisation zu suchen. Sicher ist es das Ziel, zwischen diesen beiden Positionen einen Mittelweg zu finden. Divergieren die Positionen jedoch zu stark, wird dies äusserst schwierig. Die Evaluation ist - soweit dies von uns beurteilt werden kann - eher den Bedürfnissen des Auftraggebers nachgekommen, wobei dem sogleich anzufügen ist, dass die spätere Entwicklung Hinweise dazu gibt, dass auch die Interessen des BUWAL (und speziell der unteren und mittleren Hierarchieebenen) nicht stark vernachlässigt wurden.

Aufgrund der uns zugänglichen Informationen haben Interventionen seitens der Sektion EDI der GPK NR oder der Exekutive nicht dazu geführt, dass Evaluationsoperationen beschnitten oder Resultate verzerrt worden wären. Gerade ein Fall ist uns bekannt, bei dem ein Interview nicht in der gewünschten Weise durchgeführt werden konnte. Ein Missbrauch von Evaluationsergebnissen ist uns nicht bekannt.

Wir kommen deshalb zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt. Primär aufgrund der schwierigen Konstellation - gemeint ist damit in erster Linie das problematische Dreiecksverhältnis Auftraggeber - Auftragnehmer - Evaluandum - ist das Kriterium nicht vollständig erfüllt worden.

5.3.3.2.3. Kostenwirksamkeit

Die Evaluation hat direkte Kosten im Umfang von 165'000.- sFr. verursacht. Im Vergleich zu den anderen in der Meta-Evaluation untersuchten Evaluationen gehört sie somit in die 'mittlere Preisklasse'. Die zusätzlich entstandenen Kosten, insbesondere der Aufwand der an den Sitzungen der Sektion EDI teilnehmenden Personen ist etwas höher als üblich einzustufen. Hinzu kommt der Aufwand, der beim Sekretariat der GPK verursacht wurde.

Den Wert der beschafften Informationen einzuschätzen, ist äusserst schwierig. Für den Auftraggeber sind Informationen erarbeitet worden, die sich als nützlich erwiesen haben. Für die Exekutive scheinen die beschafften Informationen vorerst eher von geringerem Wert gewesen zu sein. Dem ist jedoch hinzuzufügen, dass sich dies seit Abschluss der Evaluation offensichtlich teilweise geändert hat. Die ersten Reaktionen beurteilten die vorgebrachten Empfehlungen entweder als wenig zweckmässig oder dann als bereits hinreichend bekannt, also ohne grossen Neuigkeitswert. Aufgrund der Mitte 1994 erfolgten Reorganisationsschritte scheint sich diese Einstellung jedoch gewandelt zu haben. Inwiefern sich die von der Evaluation empfohlenen Änderungen bewähren werden, ist aus zwei Gründen heute noch kaum zu beurteilen. Erstens müssen sich die neuen Strukturen zuerst ausreichend einspielen und zweitens scheint der Reformprozess noch nicht abgeschlossen zu sein.

Erfahrungen aus der Evaluation, vor allem die im letzten Teil des Berichts entwickelten Zukunftsszenarien (SB: 149-164), sind durch Peter Knoepfel in Referaten und Artikeln auch im wissenschaftlichen Kontext bekannt gemacht worden (vgl. dazu Dente 1995, Knoepfel 1993b).

Aufgrund dieser Überlegungen ist die Kostenwirksamkeit grundsätzlich gegeben. Es bleibt jedoch abzuwarten und genauer zu prüfen, in welchem Ausmass das Kriterium erfüllt wird.

5.3.3.3. Korrektheit

5.3.3.3.1. Formale Verpflichtung

Die Bewertung anhand dieses Kriteriums stützt sich auf den Vergleich zwischen der Evaluation und den Angaben in der Projektskizze vom 10. Juli 1990 (OF) sowie jenen Verpflichtungen, die in der Auftragsvereinbarung vom 10. September 1990 enthalten sind (EK).

Finanzen: In der Projektskizze wird ein Kostenrahmen von 150'000.- sFr. festgehalten (OF: 15). Dieser Betrag ist auch Bestandteil der Verpflichtungen im Evaluationskontrakt (EK: 2). Gegen Ende des Evaluationsprozesses, am 14. Oktober 1991, wird seitens der Evaluatoren um eine zehnpromtente Honorarerhöhung (um 15'000.- sFr.) ersucht. Diesem Antrag wird am 7. November 1991 entsprochen. Eine detaillierte Schlussabrechnung zur Evaluation liegt uns nicht vor. Soweit wir unterrichtet sind, haben die Evaluatoren den Betrag von 165'000.- sFr. erhalten, wobei damit jedoch nicht der ganze Aufwand gedeckt werden konnte. Die Evaluatoren haben einen Teil ihres Aufwandes selber tragen müssen. Die finanziellen Verpflichtungen sind eingehalten worden.

Zeit: Wie bereits an anderer Stelle ausführlich dargelegt, konnte die Evaluation nicht ganz termingerecht abgeschlossen werden. Die Verzögerung hielt sich jedoch in engen Grenzen (vgl. Kapitel 5.3.3.1.7.).

Personen: Übereinstimmend sind in Projektskizze (OF: 15) und Vertrag (EK: 2) folgende Angaben enthalten: Peter Knoepfel leitet das Projekt und trägt die Gesamtverantwortung. Er bearbeitet das Projekt zusammen mit Willi Zimmermann und mit drei weiteren Personen (zwei Politologen und ein Jurist). Die drei Mitarbeiterstellen sind mit Giorgio Sailer (Ökonom), Enzo Matafora (Politikwissenschaftler) und Jörg Leimbacher (Jurist) besetzt worden. Peter Knoepfel sowie Willi Zimmermann haben - wie vertraglich festgehalten - das Projekt geleitet und bearbeitet. Bis auf die kleine Abweichung, dass eine der Stellen, die für einen Politologen vorgesehen war, durch einen Ökonomen besetzt wurde, werden die personellen Verpflichtungen eingehalten.

Inhalt: Hier sind drei unterschiedliche Aspekte zu beachten:

- Der *Untersuchungsbereich* wird in der Projektskizze sowie im Auftrag folgendermassen festgelegt: 1. Vollzugsprobleme auf der Ebene der Kantone (Massnahmenvollzug, Beziehungen zwischen dem Bund und den Kantonen, Beziehungen zwischen den Kantonen und den Gemeinden); 2. Vollzugsprobleme auf der Ebene des Bundes (Beziehungen unter den verschiedenen beteiligten Bundesämtern, interne Organisation des BUWAL) (OF: 1 und EK: 1; vgl. SB: 1). Dieser Bereich ist in der Evaluation nahezu vollständig abgedeckt worden. Lediglich das Verhältnis zwischen den Kantonen und Gemeinden ist nur rudimentär untersucht worden (vgl. den Hinweis zur vertiefenden Analyse in vier Kantonen, OF: 4). Auf diese Lücke findet sich auch ein Hinweis im Schlussbericht: "Allerdings erlaubten es die verfügbaren Ressourcen nicht, die kantonalen Vollzugspolitiken mit der ursprünglich vorgesehenen Tiefe zu analysieren." (SB: 4)
- Der zu untersuchende *Politikbereich* wird in der Projektskizze (in abnehmender Priorität) wie folgt abgegrenzt: 1. Luftreinhaltepolitik, 2. Umweltverträglichkeitsprüfung und 3. ländlicher Gewässerschutz. Im Gegensatz dazu enthält die Auftragsvereinbarung eine etwas andere Umschreibung: "Analyse des Vollzugs der Umweltschutzpolitik, insbesondere im Bereich der Luftreinhaltung (der betreffenden Verordnungen und der Massnahmen des Berichtes 'Luftreinhaltung-Konzept' des Bundesrates" (EK: 1). Da die Projektskizze (OF) im Vertrag als integrierender Bestandteil bezeichnet wird (EK: 2) ergibt sich hier ein gewisser Widerspruch. Die Evaluation befasst sich mit den genannten Politikbereichen.
- Die zu behandelnde *Fragestellung* ist in Auftrag (EK: 1-2) und Offerte (OF: 2) - bis auf die abweichenden sprachlichen Formulierungen und die Reihenfolge - identisch festgehalten. Die Evaluation hat die recht ausführlich festgelegte Fragestellung (siehe auch SB: 1-2) bearbeitet. Der Evaluationsbericht enthält eine Zusammenfassung der Antworten auf die festgelegte Fragestellung (SB: 134-137).

Zusammenfassend lässt sich festhalten, dass die Evaluation die inhaltlichen Verpflichtungen praktisch vollständig erfüllt hat. Lediglich hinsichtlich der Beziehungen zwischen den Kantonen und den Gemeinden ist eine kleine Lücke festzustellen.

Methodik/Vorgehen: Hinsichtlich der gewählten Vorgehensweise und der eingesetzten Methoden lassen sich im Evaluationskontrakt selbst keine Hinweise finden (vgl. EK). Die Offerte skizziert die beabsichtigte Vorgehensweise relativ detailliert, wobei einige Aspekte noch offen gelassen wurden. Sie hält insgesamt fünf Untersuchungsstufen fest, nämlich erstens die Analyse des 'Verwaltungsausstosses', zweitens die Untersuchung der Kantone, drittens die Untersuchung der Bundesverwaltung, viertens die Analyse und fünftens die Formulierung von Schlussfolgerungen und Empfehlungen (OF: 14). Diesem

Grobplan ist die Untersuchung gefolgt. Betrachtet man die in der Projektskizze enthaltenen, detaillierteren Hinweise, stellt man verschiedentlich Abweichungen vom ursprünglich geplanten Vorgehen fest. Exemplarisch sei dies anhand der kantonalen Untersuchung kurz erläutert. In der Offerte heisst es, dass in der ersten Phase eine Untersuchung aller Kantone (Volluntersuchung) "mittels teilstandardisierten Interviews (ca. 7 Stunden pro Kanton)" (OF: 11) geplant ist. Die Volluntersuchung wurde jedoch nicht mit mündlichen Interviews, sondern mit einem standardisierten, schriftlichen Fragebogen durchgeführt. Ähnliche Abweichungen zeigen sich auch in den anderen Untersuchungsteilen. Die Gründe für diese Abweichungen sind uns nicht bekannt. Der Schlussbericht enthält keine diesbezüglichen Hinweise. Es ist zu vermuten, dass im Laufe der Evaluation neue Aspekte zu Änderungen Anlass gaben und dazu führten, dass die Erreichung der festgelegten Ziele mit anderen Mitteln aussichtsreicher erschienen sind. Um dies jedoch fundiert bewerten zu können, wären weitere Abklärungen erforderlich.

Gesamthaft betrachtet hat die Evaluation die finanziellen Verpflichtungen vollständig eingehalten. Im personellen und inhaltlichen Bereich sind kleinere Divergenzen festzustellen, denen jedoch kein grosses Gewicht zukommt. Bezüglich der Methodik und des Vorgehens ist festzuhalten, dass die Evaluation den übergeordneten Verpflichtungen nachgekommen ist, dass sich jedoch im Detail verschiedene, zum Teil grössere Änderungen ergaben. Punktuell sind zwischen dem Evaluationskontrakt (EK) und der Offerte (OF), die einen integralen Bestandteil des Evaluationskontraktes darstellt, Widersprüche bzw. Abweichungen zu beobachten. Wir kommen zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt.

5.3.3.3.2. Deklaration von Interessenkonflikten

Die Evaluation des BUWAL entstand im Spannungsfeld zwischen den Ansprüchen der Legislative, vertreten durch die GPK NR bzw. deren Sektion EDI, und der Exekutive, also dem BUWAL und dem Departement. Die hier bestehenden Interessenkonflikte haben wir bereits an anderer Stelle diskutiert (vgl. Kapitel 5.3.3.1.8.).

Die Evaluatoren selbst waren einem weiteren Interessenkonflikt ausgesetzt. Sowohl Peter Knoepfel wie auch Willi Zimmermann haben vor und während der hier untersuchten Evaluation verschiedentlich Forschungsaufträge des BUWAL bearbeitet (so Zimmermann/Knoepfel 1987, Knoepfel/Zimmermann 1990). Daraus entsteht die Gefahr, dass die Evaluatoren durch das persönliche Interesse, auch weiterhin derartige Aufträge zu erhalten, bei der Durchführung der Evaluation beeinflusst werden könnten. Es liegen uns aber keinerlei Hinweise vor, dass sich die Evaluatoren durch derartige Überlegungen hätten

leiten lassen. Weiterhin vertritt die Evaluation gegenüber dem Evaluandum eine betont kritische Haltung. Andererseits haben die Evaluatoren unseres Wissens auch nach Abschluss der Evaluation Forschungsaufträge vom BUWAL zugesprochen erhalten (vgl. etwa BUWAL 1993). Zudem war dem Auftraggeber - auch auf explizite Hinweise der Evaluatoren hin - bekannt, dass die Evaluatoren schon mehrfach im Auftrag des BUWAL tätig waren.

Peter Knoepfel war in seiner Eigenschaft als Präsident der Schweizerischen Kommission für Umweltbeobachtung (SKUB) einem weiteren Interessenkonflikt ausgesetzt. Die Kommission vertritt die Ansicht, dass eine sinnvolle Ausgestaltung der Umweltbeobachtung ausserhalb des BUWAL erfolgen sollte. Der Bericht enthält auch Aussagen, die in diese Richtung weisen (vgl. dazu Empfehlung 13, SB: 87). Der Schlussbericht enthält keinen Hinweis auf diese Interessenkonstellation (vgl. dazu auch Knoepfel 1992b).

Bis auf diese Ausnahme sind unseres Erachtens Interessenkonflikte offen und aufrichtig behandelt worden. Wir kommen deshalb zum Schluss, dass die Evaluation das Kriterium grösstenteils erfüllt.

5.3.3.3.3. Vollständige und aufrichtige Offenlegung

Der Evaluationsbericht legt sowohl die Ergebnisse wie auch die Grenzen der Evaluation offen und vollständig dar. Der Schlussbericht enthält ausführliche Angaben zu den Befunden und zu den daraus abgeleiteten Empfehlungen. Die Schlüsse sind klar dargelegt und werden begründet. Neben den Befunden und den Empfehlungen liegt auch eine klare Beantwortung der vom Auftraggeber definierten Fragestellung vor (SB: 134-137). Auch die Darlegung der Grenzen der Evaluation sind explizit im Schlussbericht festgehalten. So weisen die Evaluatoren etwa deutlich darauf hin, dass natur- und ingenieurwissenschaftliche Produkte des BUWAL nicht bewertet werden konnten (SB: 5 und 42). Ebenso wird klar gemacht, dass der Hauptbericht (ohne den Anhang) lediglich einen Bericht über die Ergebnisse der Evaluation darstellt und dass für die wissenschaftlichen Grundlagen und Materialien der Anhang konsultiert werden sollte (SB: 7). Der Bericht thematisiert auch die Problematik, dass sich die Evaluation auch auf Vorarbeiten der Evaluatoren abstützt und dass diese im Bericht nicht ausführlich dargelegt werden können (SB: 7, vgl. dazu SB: 15-17). Auch Probleme in der Umsetzung der Interviewtechnik werden offen erläutert (SB: 13). Weiter finden sich auch Hinweise auf die Zielorientierung der Untersuchung (Konzentration auf Schwachstellen) oder darauf, dass nicht alle, sondern lediglich ausgewählte Segmente der BUWAL-Tätigkeit vertieft untersucht werden konnten (SB: 17).

Aufgrund dieser Feststellungen erfüllt die Evaluation das Kriterium.

5.3.3.3.4. Herstellung von Öffentlichkeit

Während des Evaluationsprozesses wurde sorgsam darauf geachtet, dass keine Informationen zum Projekt an aussenstehende Personen weitergegeben wurden. Die direkt oder indirekt an der Evaluation beteiligten Akteure, also die Sektion EDI der GPK NR, das Sekretariat der GPK, die Vertreter des BUWAL und jene des Departements wurden durch die Evaluatoren regelmässig über den Projektverlauf und die Teilergebnisse in Kenntnis gesetzt. Die Sektion EDI informierte die Gesamt-GPK mehrmals über den Fortgang des Projektes, wobei sich die Berichterstatter auf allgemeine Hinweise beschränkten. Der sogenannte Rohbericht, also der erste Entwurf zum Schlussbericht vom 2. September 1991, der sehr kantig und direkt formuliert war, löste im Kreis der involvierten Akteure grosse Hektik aus und wurde wieder zurückgezogen und vernichtet, weil zu befürchten war, dass er ansonsten eine weitere Verbreitung finden könnte. Dies sollte aber unter allen Umständen vermieden werden, da der Text unter anderem nicht ausreichend anonymisiert war. Dieser Bericht hätte, wäre er veröffentlicht worden, die Privatsphäre verschiedener Akteure gefährden können. Die Evaluatoren gingen bei der Erarbeitung des Rohberichts davon aus, dass dieser lediglich für die Mitglieder der Sektion EDI der GPK NR und das Sekretariat GPK zugänglich sein würde.

Nach Abschluss der Evaluation und nachdem die GPK NR ihren Bericht verabschiedet hatte, wurden die beiden Berichte (GPK 1992 und SB) am 19. Mai 1992 an einer Pressekonferenz präsentiert. Es sind uns zwei Presseberichte bekannt geworden, die bereits vor diesem Zeitpunkt Hinweise auf Ergebnisse der Evaluation enthielten (Der Bund, 18. 10. 1991 und Solothurner Zeitung, 11. 4. 1992). Wie die beiden Journalisten zu den Informationen gekommen sind, ist uns nicht bekannt.

Aufgrund des Inhalts war es sicherlich gerechtfertigt, die Teilergebnisse und insbesondere den Rohbericht unter Verschluss zu halten. Eine vorsichtiger Formulierung des Rohberichts (insbesondere eine Entpersonifizierung) hätte die Problematik deutlich entschärfen können. Nach Bereinigung des Berichts ist die Öffentlichkeit auf die Evaluation und ihre Ergebnisse aufmerksam gemacht worden (vgl. die Tagespresse vom 20. 5. 1992).

Aufgrund der Präsentation des Berichts, die das Recht der Öffentlichkeit auf Information gewährleistet, wie auch aufgrund der ergriffenen Vorsichtsmassnahmen zur Wahrung des Rechts auf Privatsphäre, wird das Kriterium erfüllt.

5.3.3.3.5. Schutz individueller Menschenrechte

Der im vorangehenden Kapitel angesprochene Rohbericht ist auch für die Bewertung dieses Kriteriums relevant. Dieser Bericht enthielt Aussagen, die in der Lage waren, die Rechte und das Wohlergehen von Einzelpersonen zu tangieren. Insbesondere stellte sich dabei das Problem, dass angeführte Aussagen aus Interviews Rückschlüsse auf die Gesprächspartner zuliessen und dass dadurch einzelne Personen unnötigen Schwierigkeiten ausgesetzt wurden. Da auch die Direktion des BUWAL und das EDI Zugang zu diesem Bericht hatten, ist dies problematisch. Die Evaluatoren gingen davon aus, dass der Rohbericht lediglich an einen eng definierten Personenkreis (Sektion EDI und Sekretariat GPK) gehen würde. Der Text wurde aber auch weiteren Personen zugestellt. Im Nachgang zu dieser Panne haben sich sowohl die Evaluatoren wie auch der Auftraggeber darum bemüht, die Verbreitung des Berichts so weit wie möglich zu begrenzen.

Da die uns vorliegenden Informationen zur Bewertung dieses Kriteriums nicht ausreichen, können wir hier keine Einschätzung vornehmen.

5.3.3.3.6. Menschliche Interaktionen

Soweit uns bekannt, wurde die Würde und der Wert der Menschen, mit denen es im Verlauf der Evaluation zu Interaktionen kam, im allgemeinen respektiert. Davon sind jedoch zwei Aspekte auszunehmen:

- Erstens stellt sich auch hier die bereits im vorangehenden Kapitel angeführte Problematik, dass bei der Abfassung des Rohberichts zuwenig darauf geachtet wurde, dass die Anonymität der Gesprächspartner gewahrt wird (vgl. dazu auch SB, Anhang 7: [2], Ziffer 1).
- Zweitens findet sich im Begleitbrief zum Fragebogen an die Kantone folgende Aussage: "Da die meisten Fragen recht holzschnittartig gestellt sind, beabsichtigen wir nicht, die Ergebnisse der Umfrage als solche zu veröffentlichen." (SB, Anhang 3: [3]) Auch wenn die Wendung 'als solche' einen gewissen Ermessensspielraum zulässt, sind wir der Ansicht, dass es aufgrund dieser Zusage nicht zulässig ist, die Umfrageergebnisse in der vorliegenden Form zu verbreiten (vgl. SB, Anhang 4, aber auch SB: 38).

Ein weiterer Punkt ist an dieser Stelle zu erwähnen:

- Aufgrund von Hinweisen aus den schriftlichen Unterlagen ist uns bekannt, dass ein Gesprächspartner das von den Evaluatoren geführte Interview als 'Verhör' empfunden

haben. Da es sich dabei um die Äusserung einer einzelnen Person handelt, müsste dieser Punkt weiter abgeklärt werden, um hierzu eine Bewertung zu ermöglichen.

Wegen der ersten zwei Punkte kommen wir zum Schluss, dass die Evaluation das Kriterium nur teilweise erfüllt.

5.3.3.3.7. Ausgewogene Berichterstattung

Vom Auftrag her war es offensichtlich, dass sich die Evaluation speziell mit Schwächen zu befassen hatte und dass Stärken des Evaluandums eher am Rande von Interesse waren. Diese Absicht wurde von Mitgliedern der Sektion EDI der GPK NR gegenüber den Evaluatoren mündlich bestärkt. Dadurch wird natürlich die Forderung nach einer ausgewogenen Berichterstattung zum Problem, steht sie doch in direktem Widerspruch zum Auftrag.

Die Evaluation zeigt sodann - auftragsgemäss - verschiedene Mängel und Schwachstellen des Evaluandums auf. Der Bericht enthält aber durchaus positive Feststellungen, auch wenn diese im Schlussbericht deutlich weniger Raum einnehmen. So wird das BUWAL beispielsweise in den Antworten auf die Fragestellung der GPK NR von verschiedenen Vorwürfen entlastet. Dies betrifft etwa die Aussagen, dass die kantonalen Vollzugsprobleme nicht auf "Politikfehleistungen auf der Ebene des BUWAL" (SB: 135) zurückzuführen seien oder dass das BUWAL im Bereich der Luftreinhaltung seine Kompetenzen nicht überschritten habe (SB: 135). Es finden sich im Bericht weitere Stellen (insbesondere im Bereich der Luftreinhaltung), die durchaus auf Stärken des Evaluandums eingehen (vgl. beispielsweise SB: 27, Ziff. 17 oder SB: 28, Ziff. 19).

In der Einleitung enthält der Bericht ebenfalls Aussagen, welche die Stossrichtung der Untersuchung verdeutlichen sollen: "Entsprechend dem expliziten Auftrag der GPK legt dieser Bericht die Schwerpunkte auf die in unseren Untersuchungen zutage getretenen Schwachstellen. Positive Befunde - solche liegen in der Tat in grosser Zahl ebenfalls vor - werden dementsprechend nicht referiert oder besonders vertieft." (SB: 7) Und an einer anderen Stelle heisst es: "Die getroffene Auswahl der Bereiche hat nach unseren beschränkten Befunden aus den nicht genauer untersuchten übrigen Bereichen der BUWAL-Aktivitäten und aufgrund unserer generellen Kenntnisse der Tätigkeiten dieses Amtes mit Sicherheit eine überproportionale Häufung problematischer Gesichtspunkte zur Amtstätigkeit zu Tage gefördert. Daraus kann leicht der - falsche - Eindruck entstehen, das BUWAL sei insgesamt ein recht problematisches Bundesamt." (SB: 17, vgl. auch SB, Anhang 10: 17)

Trotz dieser Hinweise hinterlässt der Bericht als Ganzes kein positives Bild des Evaluandums. Dieser Eindruck kommt unseres Erachtens vor allem deshalb auf, weil eine weitgehende Reorganisation des Amtes vorgeschlagen wird. Werden grosse Umstrukturierungen empfohlen, ist zu vermuten, dass die bestehende Struktur als wenig tauglich eingeschätzt wird. In die gleiche Richtung weist ein Hinweis in der Zusammenfassung der Antworten auf die Fragen der GPK im Schlussbericht: "Die Struktur des BUWAL ist nach den vorgetragenen Befunden nicht geeignet 'die Einheit und Durchschlagskraft der Umweltpolitik sicherzustellen' (Frage 3)". (SB: 136)

Aufgrund der angeführten Aspekte sind wir der Ansicht, dass sich die Evaluation um eine ausgewogene Berichterstattung bemüht hat, dass aber aufgrund des Auftrages keine vollständige Erfüllung dieses Kriteriums möglich war.

5.3.3.3.8. Finanzielle Verantwortlichkeit

Zu dieser Evaluation liegt uns keine detaillierte finanzielle Abrechnung vor. Wir können uns deshalb nur auf folgende Einzelinformationen abstützen:

- Die Projektmitarbeiter Enzo Matafora und Giorgio Sailer wurden für die Arbeiten an der Evaluation am IDHEAP im Umfang von 50 Stellenprozenten beschäftigt.
- Zur Erstellung des Rechtsgutachten wurde ein Teilauftrag an Jörg Leimbacher vergeben. Die uns vorliegenden Informationen über den finanziellen Umfang dieses Unterauftrages sind uneinheitlich und bewegen sich zwischen 10'000.- und 20'000.- sFr.
- Die schriftlichen Umfragen (inkl. Feld) sind vom IDHEAP selbst durchgeführt worden.
- Willi Zimmermann hat an der Forschungsstelle für Wissenschaft und Politik in Bern an der Evaluation gearbeitet.
- Die mündlichen Befragungen sind mehrheitlich durch Peter Knoepfel und Willi Zimmermann durchgeführt worden.
- An den Sitzungen der GPK haben in der Regel Peter Knoepfel und Willi Zimmermann sowie fallweise Jörg Leimbacher, Enzo Matafora und Giorgio Sailer teilgenommen.
- Die Evaluation hat direkte Kosten von 165'000.- sFr. verursacht. Darin enthalten ist die Honorarerhöhung um 15'000.- sFr., die kurz vor Abschluss der Evaluation bewilligt wurde.

Aufgrund der nur bruchstückhaft vorliegenden Informationen ist eine fundierte Bewertung nicht zu bewerkstelligen. Hinweise auf Verstösse gegen dieses Kriterium liegen uns

keine vor. Da Willi Zimmermann die operative Leitung der Evaluation übernahm und Peter Knoepfel mehr für die strategischen Aspekte zuständig war, stellt sich an dieser Stelle höchstens die Frage, ob es sinnvoll war, dass die beiden Projektmitarbeiter Enzo Matafora und Giorgio Sailer von Willi Zimmermann räumlich getrennt arbeiteten. Durch eine Zusammenlegung der Arbeitsorte hätte möglicherweise der Koordinationsaufwand etwas reduziert werden können. Aber auch die Klärung dieser Frage verlangt die Beschaffung weiterer Informationen.

5.3.3.4. Genauigkeit

5.3.3.4.1. Identifikation des Gegenstandes

Die Evaluation befasst sich mit einer Organisation, dem Bundesamt für Umwelt, Wald und Landschaft (BUWAL), mit seinen Politikleistungen und seinem institutionellen Kontext (Eidg. Departement des Innern (EDI), weitere relevante Bundesämter und die Kantone). Das BUWAL beschäftigt 318 Mitarbeiter (SB: 11), verteilt auf 279 Etatstellen (Stand 1991, SB, Anhang 0: 9), ist für den Vollzug von acht Gesetzen und einer Vielzahl verschiedener Verordnungen zuständig und wirkt an einer grösseren Zahl von weiteren Geschäften mit, bei denen die Federführung durch eine andere Stelle wahrgenommen wird. Daraus ergibt sich ein äusserst komplexer Untersuchungsgegenstand, dessen vollständige Erfassung den Rahmen der vorliegenden Evaluation bei weitem gesprengt hätte.

Aufgrund dieser Konstellation hat man sich dafür entschieden, in einer ersten Phase eine Grobanalyse mit einer quantitativen Vollerhebung durchzuführen und in einem zweiten Schritt die Evaluation mit qualitativen Befragungen stellenweise zu vertiefen. Dieser Vorgehensweise wurde sowohl auf der Bundes- wie auch auf der Kantonsebene gefolgt. Für die Vertiefung wurden einerseits drei Politikteilbereiche (Luftreinhaltung, ländlicher Gewässerschutz und Umweltverträglichkeitsprüfung) bestimmt, andererseits konzentrierten sich die weitergehenden Abklärungen auf kantonaler Ebene auf vier ausgewählte Kantone (Thurgau, Schwyz, Basel-Stadt und Waadt). Die Auswahl der Vertiefungsbereiche bzw. der sogenannten GPK-Testkantone wurde gemeinsam durch die Sektion EDI und die Evaluatoren getroffen, wobei die Evaluatoren vornehmlich Vorschläge ausarbeiteten, während die Sektion die Entscheide dazu gefällt hat.

Durch diese Vorgehensweise und die ergänzenden Untersuchungsteile (Dokumentenanalyse, Fallstudien und Rechtsgutachten) erreicht die Evaluation - im Verhältnis zu den zur Verfügung stehenden Ressourcen - sowohl eine gute Abdeckung des Untersuchungsge-

genstandes als auch vertiefte Erkenntnisse in den ausgewählten Bereichen. Das Kriterium wird erfüllt.

5.3.3.4.2. Kontextanalyse

Der Kontext des Evaluandums kann wie folgt umschrieben werden:

- *Institutioneller Kontext auf Bundesebene:* Hierzu zählen nicht nur das EDI, sondern auch andere Bundesämter inner- und ausserhalb des Departements, die für die Aktivitäten des BUWAL relevant sind. Weiterhin ist auf das Parlament und den Gesamtbundesrat hinzuweisen.
- *Institutioneller Kontext auf kantonaler Ebene:* Hier sind vor allem die kantonalen Umweltschutzfachstellen, die zuständigen Direktionen wie auch die kantonalen Regierungen angesprochen.
- *Policy-Kontext:* Die Tätigkeit des BUWAL im Bereich der Umweltpolitik verfügt über vielfältige Interdependenzen zu anderen Politikbereichen. Die Evaluation geht keineswegs auf alle Politikfelder ein, die hier relevant wären. Sie erreicht aber durch den Einbezug der relevantesten der angrenzenden Politikbereiche (Verkehr, Landwirtschaft, Energie etc.) eine gute Abdeckung des Policy-Kontextes.

Diese Bereiche sind in der Evaluation gründlich untersucht worden. Der Kontext des Evaluandums umfasst jedoch noch weitere Aspekte:

- *Internationaler Kontext:* Die Tätigkeit des BUWAL hat verhältnismässig viele internationale Bezüge. Die internationalen Entwicklungen haben zudem weitgehende Einflüsse auf die Politikgestaltung wie auch auf die Vollzugstätigkeiten.
- *'Öffentlichkeit':* Unter diesem Begriff verstehen wir das weitere Umfeld des Bundesamtes, also beispielsweise die Presse, die diversen Interessengruppen (Wirtschaftsverbände, Umweltschutzorganisationen, Automobilverbände etc.) aber auch die Parteien und die Öffentlichkeit als Ganzes. Diese Akteure bestimmen zu einem hohen Masse die Konjunktur, welche die Umwelt - oder spezielle Teilbereiche davon - als Themen in der öffentlichen Diskussion erhalten. Sie nehmen Einfluss auf Politikausgestaltung und Vollzug.

Diese beiden Aspekte kann die Evaluation nicht in allen Teilen abdecken. Gewisse Punkte - so etwa die Hinweise auf die Positionen des Touring Clubs der Schweiz (TCS) (SB: 43) oder jener zu internationalen Fragen (SB: 65) - sind in der Evaluation durchaus

angesprochen, eine ausreichende Kontextanalyse ist in diesen Bereichen jedoch nicht erfolgt.

Aus diesen Überlegungen heraus kommen wir zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt.

5.3.3.4.3. Beschreibung von Zielen und Vorgehen

Die *Ziele* der Evaluation werden im Schlussbericht deutlich dargelegt und deren Auswirkungen (Stichwort Schwachstellenanalyse, vgl. Kapitel 5.3.3.3.7.) kritisch diskutiert (SB: 1-4 und 7-8).

Hinsichtlich des *Vorgehens* sind die verschiedenen Untersuchungsteile differenziert zu betrachten:

- *Quantitative Befragungen auf kantonaler Ebene und auf Bundesebene:* Die Angaben zum Vorgehen auf kantonaler Ebene finden sich einerseits im Hauptbericht (SB: 5, 9-10) und andererseits in den Anhängen 3 (Fragebogen) und 4 (Auswertung). Auch zur quantitativen Befragung auf Bundesebene enthält der Schlussbericht an verschiedenen Stellen methodische Hinweise (SB: 11-12 und SB, Anhänge 7 (Fragebogen) und 8 (Auswertung). Diese Angaben reichen in beiden Fällen aus, um die Vorgehensweise identifizieren und einschätzen zu können.
- *Qualitative Befragung auf kantonaler Ebene:* Zu diesem Teil liegen lediglich die Hinweise im Hauptbericht vor (SB: 5-6, 10-11 und 13-14). Dort sind zwar einige Angaben zur Vorgehensweise enthalten, diese enthalten aber Lücken und Widersprüche: So heisst es im Bericht, dass "in den vier Testkantonen (BS, TG, SZ, VD) mit je 4 AFU-Beamten" (SB: 10) Gespräche geführt worden sind. Da die Kantone Basel-Stadt und Waadt über kein eigentliches Amt für Umweltschutz (AFU) verfügen (BUWAL 1991: 246, 257 und BUWAL 1994: 335-336, 351-352), lässt diese Aussage Fragen offen. An einer anderen Stelle wird von fünfzehn Interviews gesprochen (SB: 5, siehe auch SB, Anhang 4: 43). Zur inhaltlichen Ausrichtung der Gespräche enthält der Bericht nur sehr grobe Hinweise (SB: 6 und 10, auch SB, Anhang 4: 43). Die Fallstudien im Anhang enthalten einige ergänzende Informationen (so SB, Anhang 5: 2-3). Ein Gesprächsleitfaden wird im Schlussbericht nicht angeführt. Ein weiteres Beispiel bildet die unvollständige Information zur praktischen Durchführung der Gespräche. Daraus ergeben sich weitere Fragen: Wie wurden die Gesprächspartner ausgewählt? Handelt es sich um Einzel- oder Gruppengespräche? Wurden die Interviews mit Tonband festgehalten? Wer führte die Interviews durch? Wurde den Gesprächspartnern

Anonymität zugesichert? Dauerten die Gespräche zwei Stunden (SB: 5 und SB, Anhang 4: 43) oder rund neunzig Minuten (SB: 11)? Aufgrund dieser Unklarheiten ist es nur teilweise möglich, das Vorgehen in diesem Untersuchungsteil zu identifizieren und einzuschätzen.

- *Qualitative Befragung auf Bundesebene:* Die hierzu im Schlussbericht enthaltenen Informationen sind etwas umfangreicher (SB: 6-7 und 12-13), als dies bei der qualitativen Befragung auf der kantonalen Ebene der Fall ist (vgl. oben). So ist die Auswahl der Gesprächspartner diskutiert (SB: 6-7). Es fehlen Widersprüche in den methodischen Hinweisen. Es finden sich lediglich kleinere Informationslücken (Einzel- oder Gruppengespräche, Art der Protokollierung etc.). Das Vorgehen kann hier mehr oder weniger identifiziert und eingeschätzt werden.
- *Fallstudien:* Aufgrund der Hinweise im Schlussbericht sind die vier Fallstudien zu Sanierungen von Abwasserreinigungsanlagen (ARA) und die drei zu UVP-Verfahren aufgrund der BUWAL-Dossiers erstellt worden (SB: 11). Die Einschätzungen der Evaluatoren zu diesen Fallstudien sind in die Gespräche mit den kantonalen Vertretern (siehe oben) eingeflossen. Der Schlussbericht enthält diese Einschätzungen jedoch nur für die Darstellung der vier Subventionsverfahren für ARA (SB, Anhang 5). Die Ausführungen zu den UVP-Verfahren beschränken sich auf eine Deskription (SB, Anhang 6). Inwiefern in diese Fallstudien neben den BUWAL-Dossiers weitere Informationen eingeflossen sind, bleibt offen. Zur Auswahl der sieben Fallstudien heisst es im Schlussbericht: "Die ARA- und UVP-Fälle wurden uns vom BUWAL nach einer eingehenden Diskussion zur Auswahl zur Verfügung gestellt." (SB: 11) Genauere Hinweise zur Auswahl konnten wir bis auf eine Ausnahme im Schlussbericht nicht ausfindig machen. So heisst es zur Auswahl der UVP-Fälle folgendes: "Die von uns analysierten UVP-Verfahren" ... "wurden von unseren [kantonalen] Gesprächspartnern als kaum repräsentativ bewertet. Die vom BUWAL zur Verfügung gestellten Dossiers hätten allzu problemlose Fälle betroffen, die allesamt eine positive Stellungnahme des BUWAL erhalten hätten (teilweise mit neuen Auflagen)." (SB: 40, Ergänzungen und Auslassung, TW) So erhält der Leser zur Fallauswahl insgesamt lediglich die Information, dass eine eingehend diskutierte, positive Auswahl vorliege. Aufgrund der Informationen aus unseren Gesprächen (siehe Kapitel 5.3.5.2.), sind die ausgewählten Fälle durch kantonale Beamte genannt worden. Die Angaben zum Vorgehen bei der Erarbeitung der Fallstudien im Schlussbericht lassen jedoch nur beschränkt eine Identifikation und Einschätzung zu.
- *Rechtsgutachten:* Die zugrundeliegenden Erlasse und die weiter beigezogenen Texte (wie Gerichtsentscheide) sind transparent dargelegt. Eine Identifikation und Einschätzung der Vorgehensweise ist möglich.

- *Quantitative Datenauswertung*: In diesem Bereich ist das Vorgehen im allgemeinen ausreichend detailliert beschrieben, so dass die eingesetzten Verfahren und Techniken identifiziert und eingeschätzt werden können (vgl. Kapitel 5.3.3.4.8.).
- *Qualitative Datenauswertung*: Zur Auswertung des qualitativen Datenmaterials liegen kaum Angaben vor. Einer der spärlichen Hinweise in diesem Bereich ('grounded theory') haben wir bereits an anderer Stelle kritisch diskutiert (siehe dazu Kapitel 5.3.3.1.4.). Die qualitative Datenauswertung kann aufgrund der Angaben im Schlussbericht weder ausreichend identifiziert noch eingeschätzt werden.

Gesamthaft betrachtet liegen die zur Identifikation und Einschätzung der Vorgehensweisen erforderlichen Angaben nicht vollständig vor. Bei den qualitativen Interviews (insbesondere bei jenen mit den kantonalen Vertretern), bei den Fallstudien und vor allem bei der qualitativen Datenauswertung zeigen sich teilweise erhebliche Lücken. Erschwerend kommt hinzu, dass sich die methodischen Angaben über den ganzen Bericht verteilen, was die Übersicht erheblich erschwert (vgl. Kapitel 5.3.3.1.5.). Aus diesen Gründen kommen wir zum Schluss, dass die Evaluation das Kriterium nicht erfüllt.

5.3.3.4.4. Verlässliche Informationsquellen

Die Evaluation bezieht ihre Informationen aus einer ganzen Reihe von Informationsquellen. Es sind dies im einzelnen:

- die im Rahmen der Dokumentenanalyse konsultierten schriftlichen Unterlagen,
- die standardisierten, schriftlichen Befragungen bei den Kantonen und beim BUWAL-Personal,
- die qualitativen Interviews mit kantonalen Vertretern und Mitgliedern der Exekutive (BUWAL, weitere Bundesämter, EDI) und
- eine Reihe von Untersuchungen, welche die Evaluatoren vor oder während dem hier betrachteten Projekt bearbeiteten.

Die Informationsquellen sind im Regelfall ausreichend detailliert beschrieben, so dass die Adäquatheit der Informationen beurteilt werden kann. Bezüglich zweier Informationsquellen sind aber Einschränkungen zu vermerken. Die Unterlagen, die in die Dokumentenanalyse eingeflossen sind, können aufgrund der Angaben im Schlussbericht nicht vollständig erschlossen werden (vgl. immerhin SB, Anhänge 2 und 9, vgl. SB: 5). Der zweite kritische Einwand betrifft die im Rahmen der qualitativen Interviews befragten

Kantonsvertreter. Hier kann aufgrund der Angaben im Schlussbericht die institutionelle Zugehörigkeit der Gesprächspartner nicht eruiert werden (vgl. dazu Kapitel 5.3.3.4.3).

Neben diesen beiden Kritikpunkten ist an dieser Stelle ein weiterer Aspekt anzuführen. Die Evaluation enthält teilweise unpräzise Verweise. Dies betrifft etwa folgende Beispiele:

- Die Angabe "Aufgrund einer anderen Studie zum Kanton Bern" (SB: 40) verweist vermutlich auf die in der Evaluation unter dem Titel 'Evaluation des UVP-Verfahrens im Kanton Bern' erwähnte Studie (SB: 16, Zimmermann/Balthasar/Neuenschwander 1991, vgl. auch Ledergerber 1992).
- Mit dem "Bericht zur Situation der Trinkwasserversorgung des BUWAL von 1990" (SB: 32), wird möglicherweise auf den Zwischenbericht vom Februar 1990 hingewiesen (BUWAL 1990).
- Die Anmerkung "E. von Kardorff (Hrsg.): Handbuch der qualitativen Sozialforschung, München (Psychologie Verlags Union), 1991" (SB: 57) weist vermutlich auf Flick et al. (1991) hin.
- Der Hinweis auf die "Erkenntnisse der öffentlichen Managementlehre" (SB: 57) ist nicht nachvollziehbar.
- Die Angabe "Revision des Berichts 55 zu den Automobilemissionen" (SB: 72, vgl. dazu SB, Anhang 2: 2) soll auf die Nr. 55 der Schriftenreihe Umwelt (BUS 1986) hindeuten.

Erschwerend kommt bei diesen Angaben hinzu, dass in einigen Fällen auf umfangreiche Texte hingewiesen wird, ohne genauere Angaben zu machen. Dies gilt etwa für das oben erwähnte 'Handbuch der qualitativen Sozialforschung', das auf über 500 Seiten nicht ganz hundert Beiträge von jeweils verschiedenen Autoren mit teilweise stark unterschiedlichen Auffassungen umfasst (Flick 1991). Entgegen dieser Beispiele erfolgen jedoch die Literaturverweise im allgemeinen korrekt. In der Regel sind die Literaturstellen präzise und eindeutig ausgewiesen.

Die dargestellte Problematik hätte einmal durch einen detaillierteren Beschrieb der qualitativen Vorgehensweisen und zum anderen durch ein Literaturverzeichnis sowie präzise Quellenangaben im Text bereinigt werden können.

Aufgrund der oben angeführten Feststellungen kommen wir zum Schluss, dass die Adäquatheit der Informationen nicht in allen Fällen eingeschätzt werden kann. Dies ist primär auf die Informationslücken im Bereich der Dokumentenanalyse und der qualitativen

Interviews mit Kantonsvertretern zurückzuführen. Andere Untersuchungsteile sind im Gegensatz dazu durchaus positiv zu bewerten, so dass wir zum Schluss kommen, dass die Evaluation das Kriterium mit Vorbehalt erfüllt.

5.3.3.4.5. Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit

Die Evaluation wird bezüglich dieses Kriteriums nicht bewertet (vgl. die Vorbemerkung in Kapitel 5.3.3.).

5.3.3.4.6. Reliable Messung bzw. Verlässlichkeit

Auch hier wird keine Bewertung vorgenommen (vgl. die Vorbemerkung in Kapitel 5.3.3.).

5.3.3.4.7. Systematische Datenüberprüfung

Wir haben keine expliziten Hinweise ausfindig machen können, die auf eine systematische Datenüberprüfung hinweisen würden. Wir sind deshalb darauf angewiesen, die Angaben im Schlussbericht auf ihre Kohärenz zu überprüfen. Dabei konnten wir feststellen, dass der Schlussbericht im allgemeinen keine Fehlerhaftigkeiten enthält. Von diesem generell positiven Bild weicht jedoch der Schlussbericht an einigen Stellen ab:

- Punktuell finden sich Widersprüche in den Angaben zur Durchführung der qualitativen Interviews (vgl. oben Kapitel 5.3.3.4.3.).
- Es wurden nicht vier, sondern lediglich drei UVP-Fälle untersucht (SB: 6, 11; aber: SB, Anhang 6). Im Kanton Thurgau ist gemäss den Angaben im Evaluationsbericht bisher kein UVP-Verfahren mit Anhörung des Bundes erfolgt (SB: 38-39).
- Die Luftreinhalte-Verordnung ist nicht am 1. Januar 1985 (SB: 22), sondern am 1. März 1986 in Kraft getreten (SR 814.318.142.1, Art. 43, siehe auch SB, Anhang 9: [3]).
- Die Hinweise zur Einschätzung der Grundlagen zum ländlichen Gewässerschutz durch die kantonalen Vertreter im Hauptbericht (SB: 35) entsprechen nicht den Angaben im Anhang (SB, Anhang 4: 31-32). Zur selben Thematik finden sich auch widersprüchliche Prozentangaben im Anhang selbst (SB: 32, Text und Tabelle 19).
- Die Angaben zu Fragen im UVP-Bereich fallen im Hauptbericht (SB: 38, Ziff. 30.) und im Anhang (SB, Anhang 4: 22-26 und 34, Abb. 17) unterschiedlich aus. Zur An-

gabe, dass 40% der Befragten die UVP-Grundlagen schätzen (SB: 38), findet sich im Fragebogen keine zugehörige Frage (SB, Anhang 3).

- Die Abbildung 6 (SB: 64) enthält andere Ergebnisse als die entsprechende Stelle im Anhang (SB, Anhang 8: 13-14, Abbildung 15 und 16).
- Es ergeben sich in der Auswertung der quantitativen Befragungen Probleme mit fehlenden oder nicht beachteten Filterfragen (vgl. unten Kapitel 5.3.3.4.8.).

Der Hauptbericht enthält nur wenige sprachliche Unzulänglichkeiten. Im Gegensatz dazu stehen verschiedene Anhänge (SB, Anhänge 0, 4, 6), die nicht ausreichend genau redigiert worden sind. Möglicherweise hat hier der Zeitdruck bei Projektabschluss eine Rolle gespielt. Zudem ist darauf hinzuweisen, dass ein Teil des Evaluatorenteams nicht Deutsch zur Muttersprache hat.

Da sich die angeführten Schwachstellen auf einige spezifische Bereiche konzentrieren, kommen wir zum Schluss, dass die Evaluation das Kriterium trotz den angeführten Punkten mehrheitlich erfüllt.

5.3.3.4.8. Analyse quantitativer Informationen

Die in der Evaluation analysierten quantitativen Informationen stammen praktisch ausschliesslich aus den zwei standardisierten Befragungen bei den Vertretern der Kantone sowie beim BUWAL-Personal.

Die Datenerhebungsinstrumente, die beiden Fragebogen, finden sich im Anhang zum Schlussbericht (SB, Anhänge 3 und 7). Sie sind verhältnismässig lang ausgefallen. Sie weisen einige problematische Aspekte auf, die wir hier kurz ansprechen möchten:

Kantonale Befragung (vgl. Anhang 3):

- Zwischen Frage 13 "Sind Sie der Meinung, dass das BUWAL in der Luftreinhaltepolitik klare Zielsetzungen verfolgt?" mit den Antwortkategorien "ja", "teilweise" und "nein" sowie der Frage 14 "Beziehen sich diese allfälligen Unklarheiten auf: ..." fehlt ein Filter. Befragte, die der Meinung sind, dass eine klare Zielsetzung verfolgt wird, sollten auf die Frage nach den Unklarheiten nicht antworten müssen.
- Die Frage 23 lautet: "Beurteilen Sie die Ihnen vom BUWAL zur Verfügung gestellten UVP-Grundlagen als ausreichend bzw. zu umfangreich". Die zwei Items können durch die Befragten jeweils als zutreffend oder nicht-zutreffend bezeichnet werden. Diese Frage ist einseitig angelegt. Zudem wird aus der Formulierung nicht klar, ob der Umfang oder die Qualität der UVP-Grundlagen beurteilt werden soll.

- Bei einer weiteren Frage stellen sich Verständnisschwierigkeiten ein. So heisst Frage 33: "Sehen Sie in der UVP ein Verhinderungsinstrument des Umweltschutzes?" Hier bleibt offen, ob die UVP den Umweltschutz oder irgend etwas anderes verhindern soll.

Befragung beim BUWAL-Personal (vgl. Anhang 7):

- Frage 3i) lautet: "Sind Sie durch die anfallende Arbeit überlastet?" Es ist aufgrund der Formulierung zu vermuten, dass hier vor allem eine *quantitative* Überlastung interessiert. Dies wird jedoch nicht zwingend so verstanden. Dies führt bei der Auswertung zum Problem, dass die Antworten auf diese Frage kaum interpretiert werden können.

Die Auswertung der beiden quantitativen Befragungen wird in den Anhängen 4 und 8 dargelegt. Dabei sind die beiden Umfragen wiederum separat zu betrachten:

Auswertung der kantonalen Befragung (vgl. Anhang 4):

- Ein Grundproblem dieser Befragung stellt die unterschiedliche Anzahl auswertbarer Fragebogen pro Kanton dar (siehe SB, Anhang 4: 4, vgl. auch SB: 10). Dadurch erhält der Kanton Aargau (sieben Fragebogen) im Vergleich zum Kanton Jura (ein Fragebogen) ein deutlich höheres Gewicht in der quantitativen Auswertung. Diese Problematik konnte aufgrund der Vorgehensweise zur Bestimmung der Befragten kaum vermieden werden. Unseres Erachtens bedarf dieser Aspekt jedoch einer expliziten Darstellung und einer kritischen Diskussion.
- Die Auswertung erfolgt schwergewichtig mit Häufigkeitsauszählungen und Prozentuierungen. Ergänzt werden diese Auswertungen mit graphischen Darstellungen (Kuchen- und Säulendiagramme). In einigen Fällen werden Subsample-Vergleiche vorgenommen. Weitergehende Auswertungen sind gemäss den Angaben im Schlussbericht nicht durchgeführt worden. Vermutlich wurde aus Zeitgründen auf weitere Analysen verzichtet, obwohl die vorliegenden Daten dies durchaus zugelassen hätten und sich daraus auch zusätzliche Erkenntnisse hätten ergeben können.
- Verschiedentlich zeigen sich Probleme bezüglich der Nachvollziehbarkeit, da zum Teil auf die Wiedergabe der Prozentuierungsbasis (also der absoluten Fallzahlen) verzichtet wird und deshalb die Grundlage für die Datenauswertung nicht transparent ist (vgl. etwa SB, Anhang 4: 7-8 (Tab. 4, 5 und 6) oder 13-14 (Tab. 6 und 7)). Dies ist vor allem deshalb problematisch, da für die verschiedenen Fragen teilweise stark abweichende Fallzahlen zur Verfügung stehen.
- Bei der Auswertung der Frage 14 (SB, Anhang 4: 16) wird folgendes festgestellt: "Leider sind auch diese Antworten durch eine recht hohe Missingquote beeinträchtigt,

was ihren Stellenwert beträchtlich schmälert." Es wird jedoch kein Zusammenhang zum fehlenden Filter zu dieser Frage hergestellt (siehe oben).

- Bei Frage 19 ergibt sich ein ähnliches Problem, jedoch mit umgekehrtem Vorzeichen. Im Fragebogen ist hier ein Filter enthalten. Bei der Auswertung werden nun aber auch Antworten jener Befragten ausgewertet, für die diese Frage aufgrund der Filteranweisung gar nicht zu beantworten gewesen wäre. Dies weist daraufhin, dass der Filter bei der Auswertung unberücksichtigt geblieben ist.
- In einer Vorbemerkung zur Analyse der Fragen 28-32, die sich mit dem Gewässerschutz befassen, heisst es: "Die Fragen zu diesem Abschnitt wurden nur den im Gewässerschutz tätigen Personen vorgelegt." (SB, Anhang 4: 30) Dies widerspricht jedoch den Angaben im Fragebogen, wo es heisst, dass die Fragen in diesem Abschnitt auch von Personen, die sich mit UVP-Fragen befassen, beantwortet werden sollen (SB, Anhang 3: [5]). Im Anschluss an das vorhergehende Zitat folgt folgende Aussage: "Von diesen 33 Fragebogen waren 29 auswertbar." (SB, Anhang 4: 30) Im Gegensatz dazu wird in der nachfolgenden Auswertung aber jeweils von 33 Fällen ausgegangen (SB, Anhang 4: 30-33).
- Folgen wir den Angaben im Text, werden die Grundlagen zum ländlichen Gewässerschutz des BUWAL von den Befragten "mehrheitlich ebenfalls als 'genügend' bewertet" (SB, Anhang 4: 32), in der angefügten Tabelle wird aber ein Wert von 24 % ausgewiesen (SB, Anhang 4: 32, Tabelle 19).
- In der Interpretation zur Frage 11 wird eine problematische Interpretation vorgenommen. Es heisst dort: "Bezüglich seiner internen Koordination werden dem BUWAL strukturelle Mängel vorgeworfen. Etwa die Hälfte der antwortenden Personen ist mit der BUWAL-internen Koordination nicht zufrieden." (SB, Anhang 4: 40) Erstens findet sich in der Umfrage kein Hinweis, dass beim BUWAL *strukturelle* Mängel bestehen würden. Beim zweiten Teil des Zitates stellt sich das Problem, dass diese Frage nur von jenen Personen zu beantworten war, die Probleme bei der Zusammenarbeit mit dem BUWAL bekundeten (SB, Anhang 3: [8], Fragen 10 und 11, vgl. auch SB, Anhang 4: 11-14). Durch diese Interpretation werden die Koordinationsprobleme überhöht dargestellt und zudem in intransparenter Weise mit strukturellen Mängeln in Verbindung gebracht. Ein weiteres Problem bezüglich der Transparenz der Interpretation zeigt sich in folgender Passage: "Die mangelnde Disponibilität des BUWAL-Personals sei vor allem auf den generellen Personalmangel an diesem Amt zurückzuführen." (SB, Anhang 4: 40) Es konnte im Rahmen der quantitativen Befragung kein empirischer Beleg dafür ausfindig gemacht werden, dass die Ursache beim generellen

Personalmangel liegen würde (vgl. dazu die Frage 11, SB, Anhang 3: [9] und Anhang 4: 13-14).

Auswertung der Befragung des BUWAL-Personals (vgl. Anhang 8):

- Die Auswertung im Anhang fällt hier äusserst knapp aus. Es findet sich dazu auch ein Hinweis: "In diesem Dokument geben wir die Ergebnisse der Auswertung ohne irgendwelche Kommentare oder Interpretationen wieder. Soweit wir solche für die Evaluation für nötig erachteten, finden sie sich im Haupttext, wo besonders wichtige Umfrageergebnisse denn auch wiedergegeben werden. Die Interpretation nicht kommentierter Ergebnisse wird der Leserschaft überlassen." (SB, Anhang 8: 4) Diese Bemerkung weist daraufhin, dass der Fragebogen - in Relation zu den vorhandenen zeitlichen, finanziellen und personellen Ressourcen, sowie möglicherweise auch zu der Fragestellung - eher zu umfangreich ausgestaltet wurde.
- Die Interpretation der Umfrageresultate im Hauptbericht ist schwierig zu bewerten, da sehr oft kein präziser Verweis auf die Fundstelle im Anhang vorliegt. Erschwerend kommt hinzu, dass beide Darstellungen, sowohl jene im Anhang, wie auch jene im Hauptbericht unvollständig sind. Während dies beim Hauptbericht aufgrund der Beschränkung auf die wesentlichen Ergebnisse sicher zweckmässig ist, gilt dies für den Anhang nicht. Aus dieser Situation heraus ist es nicht möglich, einzuschätzen, wie systematisch die Datenauswertung hier erfolgte. Weiterhin wird im Hauptbericht auf Stellen im Anhang verwiesen, welche die betreffende Information gar nicht enthalten (so SB: 65 und Anhang 8: 24).
- Die Umfrage enthält verschiedene Fragen, die von Befragten in spezifischen Situationen nicht sachgerecht beantwortet werden konnten. So ist es für das administrative Personal kaum möglich, eine Differenz zwischen administrativen und fachlichen Neuerungen zu erkennen (SB, Anhang 8: 17). Ein ähnliches Problem stellt sich für Mitarbeiter, die zu keiner Hauptabteilung gehören (Direktion, Stabsstellen), wenn sie die Qualität des Informationsaustausches in ihrer 'Hauptabteilung' bewerten sollen (SB, Anhang 8: 23). Schwierigkeiten entstehen auch für Personen, die erst seit kurzer Zeit beim BUWAL tätig sind, wenn sie Auskunft darüber geben sollen, wie oft sie in den letzten 2 Jahren ihr BUWAL-Arbeitsgebäude gewechselt haben (SB, Anhang 8: 8). Die Angaben 'neuer' Mitarbeiter zu dieser Frage sind wenig aussagekräftig.

Generell haben wir bezüglich der quantitativen Datenanalyse den Eindruck erhalten, dass die Auswertung - und möglicherweise auch die Fragebogen - unter grossem zeitlichen Druck erstellt wurde. Darauf weist nicht nur die sehr deskriptive Stossrichtung der Analyse hin, sondern auch die (häufigen) sprachlichen und formalen Unzulänglichkeiten in

den Anhängen (SB, Anhänge 4 und 8). Weiterhin scheint die erhobene Datenmenge die zur Verfügung stehenden Kapazitäten zur Datenauswertung bei weitem übertroffen zu haben.

Der Entscheid, nur die wesentlichsten Ergebnisse der Umfragen in den Haupttext aufzunehmen, war sicherlich zweckmässig. Eine Trennung der Darstellung auf mehrere Berichtsteile (Hauptbericht und Anhänge) setzt aber voraus, dass die Berichterstattung an einer Stelle (also wohl im Anhang) vollständig erfolgt und dass zweitens sichergestellt wird, dass es dem Leser bei Bedarf möglich ist, auf die Darstellung der Ergebnisse zurückzugreifen. Diese beiden Bedingungen sind in der Evaluation nicht erfüllt.

Aufgrund der gemachten Feststellungen kommen wir zum Schluss, dass es nicht immer möglich ist, von einer systematischen Analyse zu sprechen, und dass die quantitativen Analysen nicht immer haltbare Interpretationen sicherstellen können. Deshalb kommen wir zur Einschätzung, dass die Evaluation das Kriterium nur teilweise erfüllt.

5.3.3.4.9. Analyse qualitativer Informationen

Aufgrund der ungenügenden Hinweise zur Vorgehensweise bei der Auswertung qualitativer Informationen (siehe Kapitel 5.3.3.4.3.) muss auf eine Bewertung der Evaluation bezüglich dieses Kriteriums verzichtet werden.

5.3.3.4.10. Gerechtfertigte Folgerungen

Es würde zu weit führen, an dieser Stelle auf alle vierzig in der Evaluation formulierten Empfehlungen einzutreten (für eine Übersicht vgl. SB: 138-148). Wir beschränken uns deshalb auf einige zentrale Aspekte.

Die Empfehlungen stützen sich zumeist auf zwei verschiedene Grundlagen ab. Zum einen sind dies die elf Leitsätze (SB: 57-60), zum anderen die vorgenommenen empirischen Analysen. Die Empfehlungen sind aufgrund dieser Basis ausreichend begründet. Zu den empirischen Analysen haben wir uns bereits an anderer Stelle geäussert (vgl. vor allem Kapitel 5.3.3.4.3., 5.3.3.4.4., 5.3.3.4.8. und 5.3.3.4.9. und die einleitenden Bemerkungen zu Kapitel 5.3.3). Mehrheitlich erlauben die vorgenommenen Analysen eine Rechtfertigung der Empfehlungen. Bei bestimmten Aspekten trifft dies jedoch nicht zu. Beispielsweise formuliert die Evaluation Empfehlungen für das gesamte BUWAL, obwohl 'nur' die drei Teilbereiche Luftreinhaltung, ländlicher Gewässerschutz und UVP einer tiefergehenden Analyse unterzogen wurden (vgl. Kapitel 5.3.3.1.8.). Neben diesen drei Pfaden sind aber in den Gesprächen auch andere Bereiche berücksichtigt worden, wobei hier die

Analyse natürlich nicht so gründlich erfolgen konnte. Dies hat zur Folge, dass zwar generell empirische Evidenz besteht, diese jedoch je nach Bereich in unterschiedlichem Ausmass erhärtet ist. Zu den Leitsätzen haben wir ebenfalls schon einige Hinweise gegeben (vgl. etwa Kapitel 5.3.3.1.4. und 5.3.3.1.8.; vgl. auch die Vorbemerkungen in Kapitel 5.3.3.). Soweit die (erfahrungsgestützten und sorgfältig generalisierten) Leitsätze akzeptiert werden, erachten wir die Empfehlungen als ausreichend gerechtfertigt. Werden diese jedoch abgelehnt, fällt auch vielfach die Begründung zu den Empfehlungen dahin. Zu diesem Problem enthält auch der Schlussbericht einen Hinweis: "Diese [Leitsätze] lassen sich nur beschränkt aus unseren empirischen Befunden herleiten; soweit darin (erfahrungsgestützte) Wertungen der Evaluatoren eingeflossen sind, sollen diese auf ihre politische Wünschbarkeit hin überprüft werden, bevor deren konkrete Anwendung im Rahmen des vorgetragenen Reorganisationsvorschlages diskutiert wird." (SB: 18, Ergänzung TW) Die Evaluation legt damit ihre Prämissen dar. Diese finden, wie aus den Reaktionen hervorgegangen ist, nicht allseits in allen Teilen eine hohe Akzeptanz. Wir erachten das Vorgehen der Evaluatoren aber durchaus als legitim. Was unseres Erachtens im Schlussbericht aber etwas zu wenig deutlich wird, ist der Umstand, dass die Empfehlungen der Evaluatoren auf Prämissen aufbauen, die auch in Frage gestellt werden (können). Es hätte sich unseres Erachtens aufgrund der Situation aufgedrängt, bei der Darstellung der Empfehlungen nochmals an prominenter Stelle auf diese Prämissenabhängigkeit einzugehen.

Aufgrund dieser Feststellungen kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt, dass aber hinsichtlich der empirischen Analyse, wie auch bei der Darstellung der Abhängigkeit der Empfehlungen von den Leitsätzen Verbesserungsmöglichkeiten bestehen würden.

5.3.3.4.11. Objektive Berichterstattung bzw. Bestätigbarkeit

Auf eine Bewertung der Evaluation anhand dieses Kriteriums wird aufgrund der oben angeführten Überlegungen verzichtet (vgl. die Vorbemerkung in Kapitel 5.3.3.).

5.3.4. Gesamteindruck

Die 'Evaluation des BUWAL' ist aufgrund der Bewertung anhand der Kriterien als eine gute Evaluation zu bezeichnen. Die Bewertung in den vier Bereichen (Nützlichkeit, Anwendbarkeit, Korrektheit und Genauigkeit) ist durchwegs positiv, auch wenn sich graduelle Unterschiede feststellen lassen. So fällt die Bewertung der Nützlichkeit sehr positiv aus, während die Evaluation im Bereich der Genauigkeit etwas schwächer beurteilt wird. Die deutlichsten Stärken der Evaluation liegen sodann darin, dass aufgrund einer grossen Informationsfülle eine recht umfassende, politikorientierte Organisationsanalyse gelungen ist, die einen hohen Praxisbezug aufweist. Durch seine hohe Kompetenz im analysierten Politikbereich kann sich das Evaluationsteam auf breite Erfahrungen abstützen. Dadurch wird das Projekt, das doch einen nicht allzu grossen finanziellen und zeitlichen Rahmen einzuhalten hat, überhaupt realisierbar. Die Schwächen der Evaluation liegen unseres Erachtens in drei Bereichen. Erstens handelt es sich um eine 'verordnete' Evaluation, das heisst das Evaluandum hat die Evaluation weder ausgelöst, noch konnte es massgeblichen Einfluss auf die Ausgestaltung der Evaluation nehmen. Möglicherweise hätte sich die Evaluation durch einen stärker partizipativ angelegten Ansatz noch weiter verbessern lassen. Dies steht jedoch im Widerspruch zur Kontrollfunktion, die Überprüfungen durch das Parlament zu erfüllen haben. Zweitens ist bei der Berichterstattung (Stichwort: Rohbericht) eine 'Panne' unterlaufen, die sich für das Projekt (und für die Sache) als wenig förderlich erwiesen hat. Drittens bestehen im methodischen Bereich gewisse Mängel. Neben den Lücken in der methodischen Berichterstattung (vor allem im qualitativen Bereich), haben wir generell den Eindruck erhalten, dass der Umfang der Informationsbeschaffung im Verhältnis zu den Kapazitäten zur Auswertung dieser Informationen zu gross bemessen wurde, so dass die Analyse die erhobenen Informationen nicht voll nutzen konnte.

Gesamthaft resultiert ein durchaus positives Bild dieser Evaluation. Es ist gelungen, eine Evaluation zu realisieren, die über eine starke Praxisorientierung und eine hohe Praxisrelevanz verfügt, ohne die wissenschaftlichen Erfordernisse allzu stark zu vernachlässigen.

5.3.5. Anhang

5.3.5.1. Unterlagen zur Evaluationsstudie

Die Untersuchung dieser Evaluationsstudie beruht auf der Analyse von schriftlichen Materialien und auf Gesprächen mit einigen relevanten Akteuren (siehe Kapitel 5.3.5.2.). Zur Erarbeitung dieser Fallstudie sind die folgenden schriftlichen Unterlagen verwendet worden (Teile der nachfolgenden Dokumente sind anonymisiert worden, um einer Auflage der GPK gerecht zu werden, siehe auch die Vorbemerkung in Kapitel 5.3.2.):

- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 3. und 4. April 1989
- Protokoll zur Plenarsitzung der GPK des Nationalrates vom 23. August 1989
- Auszug aus dem Protokoll zur Plenarsitzung der GPK des Nationalrates vom 20./21. November 1989
- Schreiben des Sekretariats GPK vom 28. November 1989
- Schreiben vom 18. Dezember 1989
- Schreiben vom 25. Januar 1990
- Aktennotiz zur Sitzung der Sektion EDI der GPK des Nationalrates vom 5. Februar 1990
- Schreiben vom 28. Februar 1990
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 28. Februar 1990
- Schreiben vom 7. März 1990
- Schreiben vom 8. März 1990
- Notiz 'Überprüfung BUWAL' vom 20. April 1990
- Auszug aus dem Protokoll zur Sitzung der GPK des Nationalrates vom 10. und 11. Mai 1990
- Schreiben der GPK vom 18. Mai 1990
- Schreiben des Sekretariats der GPK vom 13. Juni 1990
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 28. Juni 1990
- Schreiben des Sekretariats der GPK vom 29. Juni 1990
- Projektskizze vom 10. Juli 1990 (OF)
- Schreiben der GPK vom 15. August 1990
- Auftrag vom 10. September 1990 (EK)
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 29. Oktober 1990
- Schreiben an die Direktion des BUWAL vom 30. Oktober 1990
- Schreiben der GPK an die Mitarbeiter des BUWAL vom 13. November 1990
- Schreiben der GPK an die Kantonsregierungen vom 14. November 1990
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 7. Januar 1991
- Schreiben der GPK vom 16. Januar 1991
- Schreiben Zimmermann an Sekretariat GPK vom 28. Januar 1991
- Schreiben Sekretariat GPK an Knoepfel vom 30. Januar 1991
- Schreiben an die GPK vom 12. Februar 1991
- Schreiben Zimmermann an Knoepfel vom 30. April 1991
- Schreiben Knoepfel an GPK vom 8. Mai 1991
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 22. Mai 1991
- Schreiben an GPK vom 28. Mai 1991
- Schreiben der GPK vom 21. Juni 1991

- Schreiben Knoepfel an GPK vom 28. Juni 1991
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 8. Juli 1991
- Schreiben Knoepfel vom 13. August 1991
- Schreiben an Knoepfel vom 19. August 1991
- Schreiben an GPK vom 10. September 1991
- Erste Stellungnahme zum Rohbericht 'Evaluation des BUWAL' vom 2. September 1991
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 12. September 1991
- Schreiben der GPK vom 16. September 1991
- Schreiben Knoepfel an GPK vom 8. Oktober 1991
- Protokoll zur Sitzung der Sektion EDI der GPK des Nationalrates vom 14. Oktober 1991
- Schreiben (1) der GPK vom 18. Oktober 1991
- Schreiben (2) der GPK vom 18. Oktober 1991
- Schreiben an Sekretariat GPK vom 24. Oktober 1991
- Schreiben an Sekretariat GPK vom 28. Oktober 1991
- Schreiben GPK vom 7. November 1991
- Evaluation des BUWAL. Expertenbericht zur Evaluation der Luftreinhaltung, des ländlichen Gewässerschutzes und der UVP des Bundes. Hauptbericht vom 18. November 1991 (SB)
- Schreiben an GPK vom 19. November 1991
- Stellungnahme zum Expertenbericht 'Evaluation des BUWAL' vom 9. Dezember 1991
- Protokoll zur Sitzung der Vertikalen Sektion 3 der GPK des Nationalrates vom 8. Januar 1992
- Protokoll zur Sitzung der Vertikalen Sektion 3 der GPK Nationalrat vom 19. Januar 1992
- Schreiben an GPK vom 17. Januar 1992
- Protokoll zur Sitzung der Vertikalen Sektion 3 der GPK Nationalrate vom 3. Februar 1992
- Auszug aus dem Protokoll zur Plenarsitzung der GPK des Nationalrates vom 4. Februar 1992
- Protokoll zur Sitzung der Vertikalen Sektion 3 der GPK des Nationalrates vom 26. März 1992
- Schreiben Sekretariat GPK vom 31. März 1992
- Schreiben Sekretariat GPK vom 23. April 1992
- Notiz GPK-Bericht von Knoepfel (ohne Datum)
- Schreiben an die GPK vom 1. Mai 1992
- Schreiben an Sekretariat GPK vom 7. Mai 1992
- Schreiben an GPK vom 18. Mai 1992
- Auszug aus dem Protokoll zur Sitzung der GPK des Nationalrates vom 18. und 19. Mai 1992
- Bericht der GPK des Nationalrates an den Bundesrat zur Evaluation des BUWAL vom 18. Mai 1992
- Evaluation des BUWAL. Zusammenfassung des Berichtes der Geschäftsprüfungskommission vom 18. Mai 1992 (Sperrfrist 19. Mai 1992)
- Schreiben an GPK vom 21. Mai 1992
- Schreiben der GPK vom 25. Mai 1992
- Schreiben vom 31. Dezember 1992
- Schreiben an Sekretariat der GPK vom 2. Februar 1993
- Schreiben an Sekretariat der GPK vom 6. April 1993
- Stellungnahme des Bundesrates vom 7. März 1994
- Protokoll zur Sitzung der Vertikalen Sektion 3 der GPK des Nationalrates vom 20. Juni 1994
- Schreiben GPK an Knoepfel vom 4. Juli 1994
- Auszug aus dem Protokoll zur Plenarsitzung der GPK des Nationalrates vom 18. August 1994
- Schreiben der GPK vom 22. August 1994

5.3.5.2. Liste der kontaktierten Personen

- Gabriella Bardin Arigoni, Stv. Generalsekretärin, EDI, ehem. Stv. Sekretärin der GPK, Bern
- Cyrill Brügger, Nationalrat, SPS, Plaffeien FR
- Peter Knoepfel, Chavannes-près-Renens
- Philippe Mastronardi, Hochschule St. Gallen, ehem. Sekretär der GPK, St. Gallen
- Giorgio Sailer, Bellinzona
- Bruno Wallimann, Stv. Direktor, BUWAL, Bern
- Christian Wanner, Nationalrat, FDP, Messen SO
- Willi Zimmermann, Bern

5.3.5.3. Literaturverzeichnis

- BARONI, Daniela/BRUNNER, Verena/KNOEPFEL, Peter/MOOR, Pierre (1991): Strassenverkehrsrecht im Lichte des Umweltrechts. Le droit de la circulation routière face à la législation de l'environnement. Basel: Helbing & Lichtenhahn.
- BUNDESAMT FÜR UMWELTSCHUTZ (Hrsg.) (1981): Bundesamt für Umweltschutz 1971-1981. Artikelfolge zum zehnjährigen Bestehen. Bern: BUS.
- BUNDESAMT FÜR UMWELTSCHUTZ (Hrsg.) (1986): Schadstoffemissionen des privaten Strassenverkehrs 1950 - 2000. November 1986, Nachtrag September 1988. Schriftenreihe Umwelt, Nr. 55. Bern: BUWAL.
- BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT (BUWAL) (Hrsg.) (1990a): Situation der Trinkwasserversorgung. Februar 1990. Bern: BUWAL.
- BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT (BUWAL) (Hrsg.) (1990b): Handbuch Umweltverträglichkeitsprüfung UVP. Richtlinien für die Ausarbeitung von Berichten zur Umweltverträglichkeit gemäss Umweltschutzgesetz vom 7. Oktober 1983. September 1990. Bern: BUWAL.
- BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT (BUWAL) (Hrsg.) (1991): Umweltbericht 1990. Zur Lage der Umwelt in der Schweiz. Bern: BUWAL/EDMZ.
- BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT (BUWAL) (Hrsg.) (1993): Umweltforschungskatalog der Schweiz, 1990-1992. 5. Auflage. Verfasst von der Forschungsstelle Wissenschaft und Politik (Leitung: Willi Zimmermann). Schriftenreihe Umwelt, Nr. 210. Bern: BUWAL.
- BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT (BUWAL) (Hrsg.) (1994): Umweltbericht 1994. Zur Lage der Umwelt in der Schweiz. Bern: BUWAL/EDMZ.
- BUNDESRAT (Hrsg.) (1995): Bericht des Bundesrates über die Geschäftsführung der Eidgenössischen Verwaltung, Geschäftsbericht 1994 - 2. Teil. Bern: EDMZ.
- CORBIN, Juliet/STRAUSS, Anselm (1990): "Grounded theory research: Procedures, canons, and evaluative criteria." In: Zeitschrift für Soziologie, Jg. 19, Heft 6: 418-427.
- DENTE, Bruno (ed.) (1995): Environmental policy in search of new instruments. Dordrecht: Kluwer.
- ELEKTROWATT INGENIEURUNTERNEHMUNG (EWI) (Hrsg.) (1989): Untersuchungen im Zusammenhang mit dem Luftreinhalte-Konzept des Bundesrates und zusätzlichen Massnahmen zur Reduktion der Luftverschmutzung - Schlussbericht. Juli 1989. 2 Bände. Zürich: EWI.
- EPINEY, Astrid/KNOEPFEL, Peter (1991): Die schweizerische Umweltpolitik vor dem Hintergrund der europäischen Herausforderung. Juin 1991. Cahier de l'IDHEAP, No. 77. Lausanne: IDHEAP.
- FREIBURGHAUS, Dieter/BALTHASAR, Andreas/ZIMMERMANN, Willi/KNÖPFEL, Carlo (1991): Technik-Standort Schweiz. Von der Forschungs- zur Technologiepolitik. Bern/Stuttgart: Verlag Paul Haupt.
- FREIBURGHAUS, Dieter/ZIMMERMANN, Willi (1985): Wie wird Forschung politisch relevant? Erfahrungen in und mit den Schweizerischen Nationalen Forschungsprogrammen. Bern/Stuttgart: Verlag Paul Haupt.
- FREIBURGHAUS, Dieter/ZIMMERMANN, Willi (1987): Forschung und Entwicklung des Bundes: Vorschläge zu einer verbesserten Erhebungsmethode: Expertenbericht. Bern: Bundesamt für Statistik und Forschungsstelle Wissenschaft und Politik.

- FREIBURGHAUS, Dieter/ZIMMERMANN, Willi/BALTHASAR, Andreas (1990): Evaluation der Förderung praxisorientierter Forschung (KWF). Studie Nr. 12 des Bundesamtes für Konjunkturfragen. Lausanne: IDHEAP und Bern: Forschungsstelle Wissenschaft und Politik.
- GESCHÄFTSPRÜFUNGSKOMMISSION DES NATIONALRATES (GPK) (1992): Bericht der Geschäftsprüfungskommission des Nationalrates an den Bundesrat über die Evaluation des BUWAL (Bundesamt für Umwelt, Wald und Landschaft) vom 18.Mai 1992.
- GESCHÄFTSPRÜFUNGSKOMMISSION DES NATIONALRATES (GPK) (1994): Die Kohärenz staatlicher Aktivitäten: das Beispiel des Vollzugs der Luftreinhaltepolitik. Inspektionsbericht der Geschäftsprüfungskommission des Nationalrates zuhanden des Bundesrates vom 5.Mai 1994.
- GRUNDLEHNER, Sibylle/KNOEPFEL, Peter (eds.) (1992): Défis des déchets. Réalités politiques et administratives de la Suisse romande. Bâle: Helbing & Lichtenhahn.
- KISSLING-NAEF, Ingrid/KNOEPFEL, Peter (1992): "Politikflexibilität dank zentralstaatlichem Immobilismus? Handlungsspielräume kantonaler Vollzugspolitiken im schweizerischen politisch-administrativen System." In: Abromeit, Heidrun/Pommerehne, Werner W. (Hrsg.): Staatstätigkeit in der Schweiz. Bern: Haupt: 43-69.
- KISSLING-NAEF, Ingrid/KNOEPFEL, Peter (1993): Politikorientierte Lernprozesse. Konzeptuelle Überlegungen zu Lernprozessen in Verwaltungen. Novembre 1993. Cahier de l'IDHEAP, N° 120. Chavannes-prés-Renens: IDHEAP.
- KNOEPFEL, Peter (1977): Demokratisierung der Raumplanung. Grundsätzliche Aspekte und Modelle für die Organisation der kommunalen Nutzungsplanung unter besonderer Berücksichtigung der schweizerischen Verhältnisse. Berlin: Duncker & Humboldt.
- KNOEPFEL, Peter (1979): Öffentliches Recht und Vollzugsforschung. Beiträge der sogenannten Implementationsforschung zur staats- und verwaltungsrechtlichen Parlamentsdiskussion. Bern: Haupt Verlag.
- KNOEPFEL, Peter (1981): "Strategien und Implementationschancen der Luftreinhaltepolitik." In: Schweizerisches Jahrbuch für Politische Wissenschaft, Band 21: 121-152.
- KNOEPFEL, Peter (1983): Changing local-central relations in environmental policies. January 1983. Cahier de l'IDHEAP, No. 4. Lausanne: IDHEAP.
- KNOEPFEL, Peter (1984): "Die ökologische Herausforderung an die Marktwirtschaft und an das demokratische politische System". In: WWF Schweiz (Hrsg.): Gesunde Umwelt oder gesunde Wirtschaft. Ein Gegensatz, der nicht sein muss. Zürich: WWF.
- KNOEPFEL, Peter (1987): Probleme (international) vergleichender Politikanalysen - ein Werkstattbericht. Octobre 1987. Cahier de l'IDHEAP, No. 39. Lausanne: IDHEAP.
- KNOEPFEL, Peter (1988a): "Schweiz - EG: Chancen eines umweltpolitischen Anforderungsgefälles für die schweizerische Industrie." In: Schwager, Stefan/Knoepfel, Peter/Weidner, Helmut: Umweltrecht Schweiz - EG. Basel: Helbing & Lichtenhahn: 83-100.
- KNOEPFEL, Peter (1988b): Rechtlich noch fassbar? Der dritte Weg ins Vollzugsfeld: Umsetzung landwirtschaftlicher Gewässerschutzprogramme durch Programmumbau und Strukturbildung. Novembre 1988. Cahier de l'IDHEAP, No. 50. Lausanne: IDHEAP.
- KNOEPFEL, Peter (1989a): "Wissenschaftliche Beratung im Politikvollzug. Gefahr für die Demokratie? - Einige Thesen." In: Schweizerisches Jahrbuch für Politische Wissenschaft, Band 29: 81-101.
- KNOEPFEL, Peter (1989b): L'agriculture et la gestion de l'environnement. Les instruments de gestion en Suisse. Décembre 1989. Cahier de l'IDHEAP, No. 60. Lausanne: IDHEAP.
- KNOEPFEL, Peter (1990): Zum Stand des Umweltrechts in der Schweiz. Oktober 1990. Cahier de l'IDHEAP, No. 65. Lausanne: IDHEAP.

- KNOEPFEL, Peter (1991a): "Umweltpolitik zwischen Akzeptanz und Vollzugskrise". In: Lendi, Martin (Hrsg.): *Umweltpolitik. Strukturelemente in einem dynamischen Prozess*. Zürich: Verlag der Fachvereine: 142-172.
- KNOEPFEL, Peter (1991b): "Das eidgenössische Umweltschutzgesetz." In: Vetropack AG (Hrsg.): *Öko-Brevier, ein, praktisches Handbuch für kleinere und mittlere Unternehmen*. Bülach: Vetropack AG: K 2.2.
- KNOEPFEL, Peter (1992a): "La Protection de l'environnement en proie aux problèmes d'acceptation et aux déficits de mise en oeuvre". In: Morand, Charles-Albert (Ed.): *Droit de l'environnement: mise en oeuvre et coordination*. Bâle: Helbing & Lichtenhahn: 3-57.
- KNOEPFEL, Peter (1992b): *Umweltbeobachtung und Umweltberichterstattung in der Schweiz*. Janvier 1992. Cahier de l'IDHEAP, No. 85. Lausanne: IDHEAP.
- KNOEPFEL, Peter (1993a): *Bedingungen einer wirksamen Umsetzung umweltpolitischer Programme - Erfahrungen aus Westeuropäischen Staaten*. Cahier de l'IDHEAP, N°108. Chavannes-près-Renens: IDHEAP. (siehe auch in Dente 1995: 197-233)
- KNOEPFEL, Peter (1993b): *New institutional arrangements for the next generation of environmental policy instruments: Intra- and interpolicy cooperation*. Cahier de l'IDHEAP, N° 112. Chavannes-près-Renens: IDHEAP.
- KNOEPFEL, Peter (1994a): *Le partenariat public - privé dans la politique des déchets*. Cahier de l'IDHEAP, N° 124. Chavannes-près-Renens: IDHEAP.
- KNOEPFEL, Peter (1994b): *See-Sanierung: Rehabilitierung eines Toten. Bedingungen erfolgreicher Umweltpolitik in der Schweiz*. Cahier de l'IDHEAP, N° 126. Chavannes-près-Renens: IDHEAP.
- KNOEPFEL, Peter (1994c): "Chancen und Grenzen des Kooperationsprinzips in der Umweltpolitik." In: Oikos (Hrsg.): *Kooperationen für die Umwelt. Im Dialog zum Handeln*. Chur: Rüegger: 65-92.
- KNOEPFEL, Peter/DESCLOUX, Martin (1988): *Valeurs limites d'immissions: Choix politiques ou déterminations scientifiques?* Novembre 1988. Cahier de l'IDHEAP, N° 48. Lausanne; IDHEAP.
- KNOEPFEL, Peter/HORBER-PAPAZIAN, Katia (1990): "Objets de l'évaluation: essai d'identification à l'aide du concept de politique publique." In: *Evaluation des politiques publiques en Suisse, Pourquoi? Pour qui? Comment? Textes rassemblés par Katia Horber-Papazian*. Lausanne: Presses polytechniques et universitaires romandes.
- KNOEPFEL, Peter/IMHOF, Rita (1991a): *Ökologische Vernetzung versus rechtsstaatliche Handlungsmaximen. Möglichkeiten zur Überwindung eines gespannten Grundverhältnisses*. Janvier 1991. Cahier de l'IDHEAP, No. 67. Lausanne: IDHEAP.
- KNOEPFEL, Peter/IMHOF, Rita (1991b): *Les cycles écologiques e le principe de légalité. De la nécessité d'assouplir les liens de causalité*. Cahier de l'IDHEAP, N° 72. Lausanne: IDHEAP.
- KNOEPFEL, Peter/IMHOF, Rita/MATAFORA, Enzo (1991): "Zum Stand des Vollzugs des USG." In: Vetropack AG (Hrsg.): *Öko-Brevier, ein, praktisches Handbuch für kleinere und mittlere Unternehmen*. Bülach: Vetropack AG: K 2.3.
- KNOEPFEL, Peter/IMHOF, Rita/ZIMMERMANN, Willi (1994): *Massnahmenpläne zur Luftreinhaltung: Wie sich Behörden beim Umweltschutz arrangieren*. Bericht 57 des NFP 25: 'Stadt und Verkehr'. Zürich: NFP 25.

- KNOEPFEL, Peter/KISSLING-NAEF, Ingrid (1993): Transformation öffentlicher Politiken durch Verräumlichung - Betrachtungen zum gewandelten Verhältnis zwischen Raum und Politik. Juin 1993. Cahier de l'IDHEAP, No. 109. Lausanne: IDHEAP. Auch in: Héritier, Adrienne (Hrsg.): Policy Analyse. Kritik und Neuorientierung. PVS Sonderheft 24/1993: 267-288.
- KNOEPFEL, Peter/LARRUE, Corinne (1984a): Les politiques publiques comparées: tourisme intelligent ou vrai progrès? Le cas des politiques comparées de l'environnement. Septembre 1984. Cahier de l'IDHEAP, No. 17. Lausanne: IDHEAP.
- KNOEPFEL, Peter/LARRUE, Corinne (1984b): Les politiques de mise en oeuvre dans la domaine de la lutte contre la pollution atmosphérique: Evaluation comparée de l'action de trois Directions Interdépartementales de l'Industrie en France. Septembre 1984. Cahier de l'IDHEAP, No. 18. Lausanne: IDHEAP.
- KNOEPFEL, Peter/LARRUE, Corinne/ZIMMERMANN, Willi (1988): Regulatory change and institutional rearrangement. Building new policy arenas for ecologization of agriculture. Mai 1988. Cahier de l'IDHEAP, N° 44. Lausanne: IDHEAP.
- KNOEPFEL, Peter/LUNDQVIST, Lennart/PRUD'HOMME, Rémy/WAGNER, Peter (1987): Comparing environmental policies: different styles, similar content. In: Dierkes, Meinolf/Weiler, Hans N./Antal, Ariane B. (Eds.): Comparative policy research. Learning from experience. Hants: Gower House: S.171-185.
- KNOEPFEL, Peter/LEWANSKI, Rodolfo (1983): L'Analisi comparata delle politiche pubbliche di tipo regolativo: il caso dell'inquinamento atmosferico. Cahier de l'IDHEAP, N° 5. Lausanne: IDHEAP.
- KNOEPFEL, Peter/PELLI, Catherine (1989): La manne fédérale et les communes. Tentative d'explication des résultats de la distribution des subventions fédérales dans le domaine de la politique sociale. Le cas de six villes moyennes suisses. Cahier de l'IDHEAP, No. 57. Lausanne: IDHEAP.
- KNOEPFEL, Peter/WEIDNER, Helmut (1980a): Handbuch zur SO₂-Luftreinhaltepolitik. Daten, Konzepte und rechtliche Regelungen in den EG-Staaten und der Schweiz. Teil 1: Vergleichende Analyse. Berlin: Erich Schmidt Verlag.
- KNOEPFEL, Peter/WEIDNER, Helmut (1980b): Handbuch zur SO₂-Luftreinhaltepolitik. Daten, Konzepte und rechtliche Grundlagen in den EG-Staaten und der Schweiz. Teil 2: Länderberichte. Berlin: Erich Schmidt Verlag.
- KNOEPFEL, Peter/WEIDNER, Helmut (1980c): "Normbildung und Implementation. Interessenberücksichtigungsmuster in Programmstrukturen von Luftreinhaltepolitiken". In: Mayntz, Renate (Hrsg.): Implementation politischer Programme. Königstein/Ts.: Athenäum: 82-104.
- KNOEPFEL, Peter/WEIDNER, Helmut (1985): Luftreinhaltepolitik (stationäre Quellen) im internationalen Vergleich. 6 Bände. Berlin: Ed. Sigma Bohn.
- KNOEPFEL, Peter/ZIMMERMANN, Willi (1987a): Oekologisierung von Landwirtschaft. Fünf Geschichten und eine Analyse. Aarau: Sauerländer.
- KNOEPFEL, Peter/ZIMMERMANN, Willi (1987b): Modération du trafic dans les quartiers: l'exemple du quartier de Mattenhof en ville der Berne. Etude de cas préparée pour l'enseignement. Novembre 1987. Cahier de l'IDHEAP, No. 41. Lausanne: IDHEAP.
- KNOEPFEL, Peter/ZIMMERMANN, Willi/DESCLOUX, Martin (1989): L'assainissement d'un lac: l'exemple du lac de Sempach. Cahier de l'IDHEAP, N° 55. Lausanne: IDHEAP.

- KNOEPFEL, Peter/ZIMMERMANN, Willi (1990): Vollzug von Gewässerschutz in der Landwirtschaft: Untersuchung zur Umsetzung der Wegleitung für den Gewässerschutz in der Landwirtschaft in drei Testgemeinden des Kantons Luzern (Flühli, Grosswangen, Hochdorf) erstellt im Auftrag des Bundesamtes für Umwelt, Wald und Landschaft (BUWAL), der eidg. Forschungsanstalt für Agrikultur und Umwelthygiene (FAC) sowie der Volkswirtschaftsdirktion des Kantons Luzern. Lausanne.
- KNOEPFEL, Peter/ZIMMERMANN, Willi (1993): Gewässerschutz in der Landwirtschaft: Evaluation und Analyse des föderalen Vollzugs. Basel: Helbing & Lichtenhahn.
- KNOEPFEL, Peter/ZIMMERMANN, Willi/SAILER, Giorgio/MATAFORA, Enzo (1991): Evaluation des BUWAL. Expertenbericht zur Evaluation der Luftreinhaltung, des ländlichen Gewässerschutzes und der UVP des Bundes. Schlussbericht vom 18. November 1991. Bern: EDMZ.
- LEDERGERBER, Roland (1992): Die UVP im Kanton Bern. Eine vollzugsorientierte Evaluation. Pilotstudie zum Thema 'Evaluation der UVP-Auflagen am Objekt'. Im Auftrag der Koordinationstelle für Umweltschutz des Kantons Bern. (mimeo)
- LEIMBACHER, Jörg (1988): Die Rechte der Natur. Basel: Helbing & Lichtenhahn.
- LEIMBACHER, Jörg/SALADIN, Peter (1988): Die Natur - und damit der Boden - als Rechts-subjekt. Bericht 18 des Nationalen Forschungsprogramms 'Boden'. Liebefeld-Bern: NFP 22.
- LEIMBACHER, Jörg/SALADIN, Peter (1990): Katastrophenschutz: Schutz vor Schädigungen oder Schutz vor Risiken?. Elemente zum Verständnis von Art. 10 USG. Gutachten. Dezember 1990. Schriftenreihe Umwelt, Nr. 137.
- LÜSCHER, Claude/WYSS, Christian (1994): Wirkungen des UVP-Verfahrens in einem Schweizer Kanton. Partizipative Evaluation anhand der Bewilligung von Einkaufszentren/Parkhäusern in einem Schweizer Kanton. Schlussbericht des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen'. Bern: NFP 27.
- SAILER, Giorgio (1990): *Impatto e dialogo delle strade nazionali con l'ambiente. Evoluzione lungo i trent'anni di realizzazione nel territorio ticinese.* Inauguraldissertation der Rechts- und Wirtschaftswissenschaftlichen Fakultät der Universität Bern. Roveredo: Tipografia Mesolcinese.
- SCHWAGER, Stefan/KNOEPFEL, Peter/WEIDNER, Helmut (1988): Umweltrecht Schweiz - EG: das schweizerische Umweltrecht im Lichte der Umweltschutzbestimmungen der Europäischen Gemeinschaften - ein Rechtsvergleich. Basel: Helbing & Lichtenhahn.
- STRAUSS, Anselm (1994): Grundlagen qualitativer Sozialforschung. Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung. München: Wilhelm Fink.
- STRAUSS, Anselm/CORBIN, Juliet (1990): *Basics of qualitative research. Grounded theory procedures and techniques.* Newbury Park: Sage.
- STRAUSS, Anselm/CORBIN, Juliet (1994): "Grounded theory methodology. An overview." In: Denzin, Norman K./Lincoln, Yvonna S. (eds.): *Handbook of qualitative research.* Thousand Oaks: Sage.
- UEBERSOHN, Gerhard (1990): Effektive Umweltpolitik. Folgerungen aus der Implementations- und Evaluationsforschung. Europäische Hochschulschriften Reihe II Rechtswissenschaft, Bd. 1018. Frankfurt: Lang.
- WEIDNER, Helmut/KNOEPFEL, Peter (1983): "Innovation durch international vergleichende Politikanalyse dargestellt am Beispiel der Luftreinhaltspolitik". In: Renate Mayntz (Hrsg.): *Implementation politischer Programme II. Ansätze zur Theoriebildung.* Opladen: Westdeutscher Verlag: 221-255.

- WEIDNER, Helmut/KNOEPFEL, Peter (1985): Luftreinhaltepolitik in städtischen Ballungsräumen. Internationale Erfahrungen. Frankfurt: Campus.
- WIEDEMANN, Peter (1991): "Gegenstandsnahe Theoriebildung." In: Flick, Uwe/Kardorff, Ernst v./Keupp, Heiner/Rosenstiel, Lutz v./Wolff, Stephan (Hrsg.) (1991): Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen. München: Psychologie Verlags Union: 440-445.
- ZIMMERMANN, Willi (1980): Vom Bewusstsein zum Diskurs. Eine Untersuchung zum Verhältnis von Erkenntnistheorie und politischer Theorie am Beispiel der empirisch-analytischen Theorien von John Locke und David Hume. Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Ludwig-Maximilians-Universität München, 1979. Zürich: ADAG.
- ZIMMERMANN, Willi (1988): Umweltbeobachtung und Umweltforschung in der Schweiz. II. Auswertung der Umfrage 'Ist-Zustand der Umweltbeobachtung in der Schweiz'. Schriftenreihe Umweltschutz, Nr 81. Bern: BUS.
- ZIMMERMANN, Willi et al. (1991): Grundlagen der Umweltforschung: Daten, Fakten, Szenarien, unter besonderer Berücksichtigung der Ressortforschung des Bundes. Schriftenreihe Umwelt, Nr. 170. Bern: BUWAL.
- ZIMMERMANN, Willi/AESCHLIMANN, Elsbeth (1991): Angewandte Umweltforschung in der Schweiz: Forschergruppen, Prioritäten und Schwerpunkte, Mängel und Lücken. Eine Erhebung in den Natur- und Ingenieurwissenschaften unter besonderer Berücksichtigung der Ressortforschung des Bundes. Schriftenreihe Umwelt, Nr. 171. Bern: BUWAL.
- ZIMMERMANN, Willi/BALTHASAR, Andreas/NEUENSCHWANDER, Peter (1991): Die UVP im Kanton Bern. Eine vollzugsorientierte Evaluation. Im Auftrag der Koordinationsstelle für Umweltschutz des Kantons Bern. 25. Januar 1991. Bern: Forschungsstelle Wissenschaft und Politik. (mimeo)
- ZIMMERMANN, Willi/KNOEPFEL, Peter (1987): Landwirtschaft und Umwelt im politischen Alltag. Drei Fälle für die Schule. Im Auftrag des Bundesamtes für Forstwesen und Landschaftsschutz, des Bundesamtes für Umweltschutz, des Bundesamtes für Raumplanung und des Bundesamtes für Landwirtschaft. Bern: BUS/EDMZ.
- ZIMMERMANN, Willi/KOLLY, Doris/MAHON, Pascal (1989): Die wissenschaftliche Evaluation von Bundespolitik. Eine bibliographische und methodische Exploration. Eidgenössisches Justiz- und Polizeidepartement, Arbeitsgruppe Gesetzesevaluation. Bern: Forschungsstelle Wissenschaft und Politik.
- ZIMMERMANN, Willi/WYMAN, Christian (1988): Programmatische Aussagen zur Verkehrspolitik. 2 Bde. Bern: (Forschungsstelle Wissenschaft und Politik).

5.3.6. Stellungnahme der Evaluatoren

Die vorliegende Metaevaluation wird unserer vor bald fünf Jahren abgeschlossenen Evaluationsstudie zum Bundesamt für Umwelt, Wald und Landschaft weitgehend gerecht. Die Analyse besticht insbesondere durch ihre Präzision der Textwiedergaben, die bis ins Detail gehenden Kenntnisse des Autors und der Umstand, dass wir uns auch in den kritischen Anmerkungen weitgehend wiederfinden können. Die vorliegende Evaluation erscheint uns korrekt, sachlich richtig und im Hinblick auf künftige Studien ähnlicher Art recht lehrreich. In den folgenden sechs Punkten unserer kritischen Stellungnahme schwingt insgesamt ein leicht kritischer Unterton mit, der sich dahingehend zusammenfassen lässt, dass der Autor wissenschaftlich zwar sauber gearbeitet hat, die politischen Rahmenbedingungen unserer Evaluationsarbeit da und dort indessen zu wenig mitbedacht haben dürfte. Dies ist ihm allerdings kaum anzulasten, hat er doch selbst bisher kaum Gelegenheit gehabt, eine Evaluation unter vergleichbaren Umständen durchzuführen. Im einzelnen nehmen wir zu folgenden Punkten Stellung:

1. *Die "Leitsätze"*: Der Berichterstatter trägt bei der Beurteilung dieser Leitsätze dem Umstand zu wenig Rechnung, dass die iterative Induktion und die gleichzeitige Deduktion dieser Leitsätze aus Erfahrungsbeständen der Evaluatoren unter spezifischen politischen Bedingungen erfolgte, die man als konstruktivistische Dezision der beteiligten Akteure bezeichnen kann. Es handelt sich um stufenweise realisierte Lernschritte, deren Ergebnisse schliesslich in Form von 11 wegleitenden Prinzipien festgehalten wurden. Diese Leitsätze haben eine längere, heute kaum noch rekonstruierbare Geschichte; sie stehen bewusst nicht am Anfang der Studie, weil sie erst im Laufe der Untersuchungen entstanden und sowohl bei den Evaluatoren als auch bei den Mitgliedern der Geschäftsprüfungskommission Wandlungen erfahren haben; aus anfänglich individuellen "Vor-Urteilen" entstanden, wurden sie im Laufe der Evaluationsarbeiten zu kollektiven, von allen Beteiligten gemeinsam geteilten "Wahrheiten". An die Stelle der (zugegebenermassen zu wenig explizit gemachten) Abstützung in der Literatur ist die - notwendige - Abstützung im Wertsystem der am Evaluationsprozess beteiligten Akteure getreten. Es ist für ein politisch zusammengesetztes Gremium weder üblich noch ratsam, diesen (oft im positiven Sinne chaotisch verlaufenden) Entscheidungsprozess in allen seinen Verästelungen zu dokumentieren und zu diskutieren. Die Umschlagstellen von der Empirie zur Dezision sind in diesem stark im politischen Rampenlicht stehenden Evaluationsprozess noch verwischer geblieben, als dies in politisch weniger exponierten Evaluationen üblicherweise der Fall ist. Wie der Berichterstatter denn auch zu recht unterstreicht, wurde mit der expliziten Formulierung der Leitsätze auch die Aufforderung verbunden, diese als politische Dezision anzusehen und ent-

sprechend zu hinterfragen. Damit sollten die politischen Prämissen der Evaluation offengelegt werden. Dass dies weitgehend gelungen ist, zeigt der Bericht der Geschäftsprüfungskommission, in dem einer der Leitsätze (Trennung von Nutzungs- und Schutzfunktionen) nur teilweise übernommen wurde, weshalb auch die darauf abgestützte Handlungsempfehlung (Ausgliederung sämtlicher Funktionen der Waldnutzung aus dem BUWAL) nicht vorbehaltlos akzeptiert wurde.

2. *Methodentransparenz*: Die vom Berichtersteller angeführten methodenkritischen Hinweise zeugen einmal mehr von der Seriosität der vorgelegten Metaevaluation. Sie treffen in den meisten Fällen zu: In den einen Fällen handelt es sich um eigentliche methodische Fehler, die uns in der Hitze des Gefechts unterliefen (z. B. Zitate, einige u. E. für das Ergebnis eher unbedeutende statistische Auswertungen); in anderen Fällen sind es eher Unterlassungen, die daher rühren, dass wir die zwischen den Anhängen und dem Text liegende Arbeitsschritte explizit zu wenig dokumentiert haben. Diese letzteren Unzulänglichkeiten rühren u. a. auch daher, dass verschiedene Personen am Werk waren, die definitive Version der Anhänge oft erst nach jener des Haupttexts vorlag und die Arbeit an den Anhängen mangels Finanzen nicht immer mit der notwendigen Sorgfalt und Kontrolle durchgeführt werden konnte. Tatsächlich forderte unser Auftraggeber keine wissenschaftlichen Abhandlungen, sondern - rasche - Ergebnisse. Diese Haltung der Kommission bewog uns, auch anfänglich im Haupttext geplante methodische Überlegungen und statistische Materialien schliesslich in die Anhänge zu verweisen; einige dieser (intern dokumentierten) Überlegungen fielen schliesslich gänzlich aus dem publizierten Text heraus. Erfahrungsgemäss fordern Politiker methodische Überlegungen nur dann an, wenn sie vom Ergebnis nicht überzeugt sind. Aus dieser Auftragslage kann eine Berichterstattung resultieren, die aus der Sicht der "reinen Wissenschaft" tatsächlich unbefriedigend bleibt.
3. *Die "Panne"*: Tatsächlich hat der federführende Redaktor des Rohberichts (Peter Knoepfel) dessen Inhalt auch dann zu verantworten, wenn er davon ausging, dass der Bericht nur den Mitgliedern der Geschäftsprüfungskommission, auf keinen Fall aber dem BUWAL oder dem Departement vorgelegt werden sollte. Der Rohbericht folgte der bei Organisationsanalysen üblichen Logik, wonach festgestellte Unzulänglichkeiten beim Produkt zunächst auf unzulängliche Strukturen, daraufhin auf Mängel in den Abläufen und schliesslich auf Führungsschwächen bei den entscheidenden Akteuren zurückgeführt werden. Evaluationen können niemals auf die Identifikation solcher Führungsschwächen der Akteure verzichten. Dass die Akteure dabei nicht als Personen, sondern als Funktionsträger in Organisationen angesehen werden, ist selbstverständlich. Daraus resultiert für jede Evaluation das Dilemma, dass die kritisierten Ak-

teure prinzipiell nicht entpersonalisiert werden können. Insofern ist das Gebot der Entpersonalisierung bei Schlüsselakteuren nicht nur praktisch nicht einlösbar, sondern es verstösst gegen das Prinzip der Transparenz und der funktionalen Zurechenbarkeit kritischer Entscheidungen zu entsprechenden Funktionsträgern. Es erscheint uns deshalb nicht statthaft, generell Entpersonifizierung als Qualitätsstandards von Evaluationen zu betrachten.

Richtig ist indessen, dass solche personenbezogenen Aussagen in Evaluationsprozessen vorsichtig zu formulieren und hieb- und stichfest zu begründen sind. Ausserdem sollen sie in der Berichterstattung nur dann aufgeführt werden, wenn sie vom Auftrag miterfasst sind. Über diese Bedingung bestand bis zur Vorlage des Rohberichts keine Klarheit. Der ursprüngliche Auftrag liess bei den Evaluatoren zwar keine Zweifel darüber, dass er auch akteurspezifische Aussagen mitbeinhalten sollte. Dementsprechend enthielt der Rohbericht einen Hinweis auf den Stellenwert solcher Aussagen, die denn auch in den Text Eingang fanden. Im Bewusstsein, dass sich der Auftraggeber in Kenntnis dieser u. E. personen- und nicht struktur- bzw. prozessbedingten Unzulänglichkeiten darüber auszusprechen hätte, ob und inwiefern diese im Bericht zu berücksichtigen seien, bezeichneten wir diese erste, evaluative Berichterstattung als "Rohbericht", der nur den Kommissionsmitgliedern zur Kenntnis gebracht werden sollte. Er sollte (und wurde denn auch) daraufhin wieder eingesammelt und entsprechend den Direktiven der Kommission umformuliert werden.

Wie der Berichtersteller richtig festhält, ist die eingetretene Panne der breiteren Streuung dieses Rohberichts nicht den Evaluatoren zuzuschreiben; sie gaben dem Sekretariat der GPK ausdrücklich zu verstehen, dass dieses Dokument nur an die Kommissionsmitglieder zu verteilen sei. Für künftige Evaluationen ähnlicher Art kann man aus dieser Panne allerdings den Schluss ziehen, dass potentiell jede schriftliche Äusserung von Evaluatoren den Weg zum Evaluandum finden kann; mitunter wird daher Mündlichkeit der Schriftlichkeit vorzuziehen sein. Wir teilen im übrigen die Einschätzung, dass diese "Panne" der Akzeptanz der Empfehlungen nicht gerade dienlich war, obgleich die wenigen nicht entpersonalisierbaren Aussagen jedem halbwegs initiierten Mitarbeiter der Verwaltung bekannt waren.

4. *Partizipation*: Wie der Berichtersteller richtig festhält, stand das Unternehmen "Evaluation des BUWAL" unter dem schwierigen Vorzeichen eines eigentlichen "Inspektions- bzw. Kontrollverfahrens". Entgegen der Auffassung verschiedener Vertreter der neuesten Evaluationsliteratur (z. B. Guba und Lincoln 1987: 203ff.) gehört diese Evaluationssituation auch heute noch zur praktischen Realität. Auch wenn wir die Auffassung teilen, dass Evaluationsprozesse möglichst partizipative Lernprozesse

sein sollen, so kann nicht wegdiskutiert werden, dass dabei konkurrierende politische Macht- und Herrschaftsansprüche weiterhin eine Rolle spielen werden. In der Hand der Parlamente bleibt das Instrument der Evaluation ein politisches Machtmittel, das der kritischen Kontrolle über die Verwaltung dienen soll; sie ist und bleibt eine Evaluation des Parlaments und nicht eine Selbstevaluation der Verwaltung. Wir müssen im Hinblick auf die zunehmende Nutzung der Evaluation für die Verwaltungskontrolle durch die Parlamente lernen, mit dieser bez. Partizipation des Evaluandums prinzipiell beschränkten Kategorie der Evaluation umzugehen. Die "Evaluation des BUWAL" war die erste, sicherlich nicht die letzte systematische Politikevaluation, die in der Schweiz im Auftrag eines Parlamentes durchgeführt wurde. Für die Evaluatoren ist diese Situation bez. ihrer eigenen Unabhängigkeit, ihrer Entscheidungsfähigkeit und ihrer persönlichen und/oder institutionellen Stellung schwieriger zu ertragen als die Mitwirkung an einer von der Verwaltung selbst initiierten Evaluation.

Der Berichterstatter würdigt diese Ausgangslage u. E. in korrekter Weise (5.3.3.2.2.). Er berücksichtigt richtigerweise den von uns befolgten Grundsatz, dass keine Organisation als homogene Einheit auftritt. Ein Grossteil der strukturellen und prozeduralen Ursachen für positive und negative Befunde auf der Ebene der Outputs wurden uns in der Tat von Mitarbeiterinnen und Mitarbeitern der unteren und mittleren Kader des BUWAL selbst genannt. Diesen Einschätzungen legten wir besonderes Gewicht bei, war doch davon auszugehen, dass diese Personen in der künftigen Organisation des Amtes wesentliche Spitzenpositionen einnehmen würden. Naturgemäss gingen wir demgegenüber vorsichtiger um mit Aussagen seitens der seinerzeit verantwortlichen Spitzenfunktionäre, soweit sich diese auf mögliche Ursachen negativer Befunde auf der Ebene der Politikoutputs bezogen.

5. *Datenmenge*: Die Frage, "ob in Anbetracht der zur Verfügung stehenden Ressourcen für die Datenauswertung nicht eher eine zu grosse Datenmenge erhoben wurde" (5.3.3.2.1.) ist berechtigt. Sie gilt nach unseren Erfahrungen für die meisten empirischen Evaluationsstudien, weil oft zu Beginn nicht feststeht, welche Daten für das schlussendliche Raisonement von Bedeutung sein werden. Vermutlich ist diese Unsicherheit um so höher, je kürzer der Projektzeitlauf und insbesondere die Phase der konzeptionellen Vorbereitung der Studie bemessen ist. In unserem Falle betrug diese Phase nur ein bis zwei Monate, und die Studie betrat auch für uns Neuland, weil wir erstmals auf breiter Front eine politikorientierte Organisationsanalyse durchführten. Aus diesem Grunde mussten die "Datennetze" weit ausgeworfen werden (z. B. Kantons- und MitarbeiterInnenbefragung), obwohl der Ertrag nicht von vornherein feststand. Im nachhinein ist man immer klüger. Die - löbliche - Absicht, einige der anläss-

lich der Berichterstattung nicht ausgewerteten Datensätze später einmal weiterzubearbeiten, konnten wir, ähnlich wie manch eine andere BerufskollegIn, nur teilweise in die Tat umsetzen. Dies geschah etwa in unseren Lehrveranstaltungen oder in realisierten oder noch geplanten wissenschaftlichen Publikationen. Vom Parlament in Auftrag gegebene Evaluationen sollten indessen in Zukunft längere Konzeptionsphasen aufweisen, als dies bei der Studie des BUWAL der Fall war. Zu diesem Zwecke müssten die von politischen Konjunkturen geprägte Ungeduld parlamentarischer Kommissionen etwas gebändigt werden. Allerdings wird es immer dabei bleiben, dass die wissenschaftlichen Evaluatoren nicht Herren der Terminkalender sind.

6. *Kontextanalyse*: Es trifft zu, dass unser Bericht wenig Angaben über den internationalen Kontext und über die betroffenen Interessengruppen (Wirtschaftsverbände, Umweltschutzorganisationen, Automobilverbände) enthält (5.3.3.4.2.). Die diesbezügliche Kritik des Berichterstatters erscheint uns indessen nicht gerechtfertigt. Bezüglich internationalem Kontext ist festzuhalten, dass die analysierten Teilpolitiken (Luftreinhaltung, ländlicher Gewässerschutz, UVP) damals nur einen relativ unbedeutenden internationalen Aspekt aufwiesen, der ausserdem amtsintern als eine Art "Umweltaussenpolitik" vom inländischen Kontext organisatorisch isoliert abgewickelt wurde. Eine Analyse der Öffentlichkeit und der Presse wurde in Anbetracht der beschränkten zeitlichen und finanziellen Ressourcen bewusst nicht in das Mandat aufgenommen; die Interessenverbände und die Presse erscheinen im Bericht allerdings dort explizit, wo sie für die betroffenen Politiken eine Rolle spielten. Eine darüber hinausgehende Kontextanalyse wäre ein akademischer Luxus gewesen; es wäre technisch ein leichtes gewesen, diese aus zahlreichen (teilweise parallel zur BUWAL-Studie verfassten) Zeitschriften und Bücherartikeln der Evaluatoren zusammenzuschneiden. Wir sind im übrigen der Meinung, dass Kontextanalysen oft ins Unendliche ausufern (Frage der Systemgrenze) und daher nur einen geringen funktionellen Erklärungsgehalt für die zu erklärenden Sachverhalte aufweisen.

Trotz dieser kritischen Anmerkungen halten wir zum Schluss zusammenfassend fest, dass die vorliegende Metaevaluation eine saubere und wissenschaftlich korrekte Würdigung unserer Evaluation darstellt. Unsere Einwände richten sich letztlich nicht gegen die Art der Durchführung der Evaluation, sondern gegen die (dem amerikanischen Kontext entnommenen) Kriterien. Wir hoffen, dass diese Zeilen dazu beitragen, insbesondere die Kriterien "politische Überlebensfähigkeit", "bewertende Interpretationen" (unter den besonderen Bedingungen einer parlamentsinduzierten Evaluation) und "Kontextanalyse" noch mal zu diskutieren. Wir sind der Überzeugung, dass es der Scientific Community und der Qualität der Evaluation in der Schweiz wohl ansteht, wenn Evaluationen ihrer-

seits von Zeit zu Zeit anspruchsvollen Metaevaluationen unterstellt werden. Dass dabei gelegentlich ein teilweise realitätsferner Kriterienkatalog zur Anwendung gelangt, tut der - guten - Sache keinen Abbruch.

Literatur

Guba, Egon G. und Lincoln, Yvonna S. 1987: *Effective evaluation: improving the usefulness of evaluation results through responsive and naturalistic approaches!* San Francisco/London: Jossey-Bass Publ.

Peter Knoepfel und Willi Zimmermann

6. VERGLEICH ZWISCHEN DEN FALLSTUDIEN

Das nachfolgende Kapitel ist folgendermassen strukturiert. Der erste Abschnitt (Kapitel 6.1.) befasst sich mit dem Vergleich der zehn untersuchten Evaluationsstudien auf der Ebene der dreissig Einzelkriterien.

In einem zweiten Teil folgt die vergleichende Analyse innerhalb der vier Kriteriengruppen (Kapitel 6.2.), die auch eine Zusammenfassung der wesentlichsten Punkte aus dem vorangehenden Kapitel enthält.

Der dritte Abschnitt (Kapitel 6.3.) befasst sich mit dem Vergleich der Fallstudien unter Berücksichtigung des gesamten Kriterienrasters.

In einem vierten Teil (Kapitel 6.4.) werden die Interdependenzen zwischen den Einzelkriterien bzw. den Kriteriengruppen untersucht. Dabei wird speziell darauf geachtet, welche Kriterien(gruppen) stark verknüpft sind. Auf der Ebene der Kriteriengruppen werden auch die Eigen- oder Selbstbeziehungen (i.e. Beziehungen zu Kriterien derselben Kriteriengruppe) thematisiert. Auch dieser Vergleich basiert auf den empirischen Fallstudien, wobei sich die Analyse insbesondere auf die in den Fallstudientexten enthaltenen Verweise zwischen den Kriterien abstützt. Neben einer bivariaten Analyse der Interdependenzen zwischen den Kriterien werden mit einer Clusteranalyse und einer Multidimensionalen Skalierung (MDS) auch multivariate Verfahren für die Netzwerkanalyse eingesetzt.

Der letzte Abschnitt (Kapitel 6.5.) enthält einige grundsätzliche Überlegungen zum Stellenwert der in diesem Kapitel enthaltenen vergleichenden Analysen.

6.1. Vergleich entlang der Einzelkriterien

Der Aufbau dieses Kapitels folgt der Struktur des Kriterienrasters. Zu jedem Kriterium erfolgt zu Beginn eine Kurzübersicht zu den Bewertungsergebnissen. Dabei wird für die Bewertung der zehn Evaluationsstudien folgende Symbolik eingesetzt:

+ : positiv	(+) : eher positiv	+/- : neutral
(-) : eher negativ	- : negativ	/ : keine Bewertung

Eine *positive* Bewertung erhalten Studien, die unseres Erachtens das Kriterium, das als Idealkriterium formuliert ist (vgl. dazu Kapitel 4.1.), erfüllt haben. In den Fallstudien werden Formulierungen wie 'vollständig erfüllt', 'ausreichend erfüllt' oder schlicht 'erfüllt' verwendet.

Eine *eher positive* Bewertung bedeutet, dass die Studie nahe an das Idealkriterium herankommt, ohne dies jedoch vollständig zu erreichen. Zu dieser Kategorie gehören die in den Fallstudien benutzten Begriffe wie 'grösstenteils erfüllt', 'mehrheitlich erfüllt', 'mit Vorbehalt' oder 'mit Abstrichen erfüllt'.

Eine *neutrale* Bewertung erfolgt dann, wenn sich positive wie negative Aspekte etwa die Waage halten. Hier wurden in den Fallstudien Formulierungen, wie 'teilweise' oder 'zum Teil' erfüllt benutzt.

Eine *eher negative* Bewertung resultiert in jenen Fällen, in denen zwar negative Punkte überwiegen, trotzdem aber einige wesentliche positive Elemente festgestellt wurden. In den Fallstudien wurde in diesen Fällen die Formulierung 'mehrheitlich nicht erfüllt' eingesetzt.

Bei einer *negativen* Bewertung dominieren ganz eindeutig problematische Aspekte. In diesen Fällen hat das zur Diskussion stehende Kriterium keine oder viel zu wenig Beachtung erhalten. In den Fallstudien verwendeten wir in diesen Situationen die Formulierung 'nicht erfüllt' oder 'nicht beachtet'.

Wird auf eine Bewertung verzichtet, fehlten uns dazu die entsprechenden Informationen oder dann war das Kriterium auf keine zweckmässige Weise einzusetzen. Im ersten Fall wäre durch eine vertiefte Meta-Evaluation sehr wohl eine Bewertung möglich. Im zweiten Fall hingegen, hat sich das Kriterium zur Bewertung der spezifischen Evaluationsstudie nicht geeignet.

Die nachfolgende vergleichende Analyse geht auf die Gründe für positive wie auch negative Bewertungen hinsichtlich der einzelnen Kriterien ein. Dabei gilt ein besonderes Augenmerk der zur Auswahl der zehn Evaluationsstudien eingesetzten Dimensionen (vgl. auch oben, Kapitel 4.2.):

Dimensionen zum Evaluandum:

- *Politische Ebene* (Bund, Kanton, Gemeinde)
- *Struktur der Policy-Arena*: Mit der Struktur der Policy-Arena sind primär die Konflikt- und Konsensusprozesse gemeint. Ein Politikbereich, in dem ein Evaluandum zu lokalisieren ist, kann sich diesbezüglich stark unterscheiden. Neben unterschiedlichen Konfliktivitätsgraden, also der Umstrittenheit einer Politik (vgl. Widmer 1987), kann sich die Policy-Arena auch dadurch unterscheiden, ob sie über längere Zeit konstante Koalitionen verfügt, oder ob diese nur sehr kurzlebig sind. Der dritte Aspekt betrifft die Polarität einer Policy-Arena, also die Frage durch wieviele Koalitionen (bzw. Konfliktlinien) sie geprägt ist.
- *Art des Evaluandums* (Einzelmassnahme, Gesamtprogramm, Organisation)
- *Steuerungsprinzip(ien) des Evaluandums*: (Gebot/Verbot (regulativ), positive/negative Anreize ('incentives'), Angebot, Überzeugung ('persuasion') und Vorbild)

Dimensionen zur Evaluation:

- *Umfang der Evaluation* (finanzielle, personelle und zeitliche Ressourcen)
- *Erkenntnistheoretische Ausrichtung*: Darunter verstehen wir den erkenntnistheoretischen Ansatz, der in der Evaluation verfolgt wird. Es lassen sich dabei zwei verschiedene Grundausrichtungen ausmachen. Einmal handelt es sich dabei um das von uns als 'traditionelles Paradigma' bezeichnete Verständnis. Darunter fallen alle realistischen, (post-)positivistischen, an Objektivität und Wertfreiheit orientierten Ansätze. Die zweite Grundausrichtung vereint konstruktivistische, partizipativ-orientierte, responsive und perspektivistisch angelegte Ansätze, welche die Objektivität und Wertfreiheit als wissenschaftliche Zielorientierung verneinen.
- *Vergleichsdimension* (Längsschnitt- oder Querschnittuntersuchungen bzw. Kombinationen)
- *Methodik*: Darunter verstehen wir alle Ansätze, Verfahren, Techniken und Instrumente, die dazu dienen, die empirische Umsetzung einer Evaluation mit wissenschaftlichen Mitteln zu gewährleisten. Dazu zählen Evaluationsansätze und Untersuchungsanlagen ebenso wie Verfahren zur Datenerhebung und Datenauswertung.

Dimensionen zu den Evaluatoren:

- *Kontext der Evaluatoren*: Hier sind primär zwei Formen zu unterscheiden, nämlich Evaluatoren, die als Selbständige oder als Angestellte eines privatwirtschaftlichen Unternehmens Evaluationsaufträge bearbeiten, sowie jene, die an einer Universität oder Hochschule tätig sind. Neben diesen beiden Formen sind auch Mischkategorien möglich, also gemischt zusammengesetzte Evaluatorenteams oder Einzelpersonen, die sowohl privatwirtschaftlich wie universitär aktiv sind.
- *Disziplinäre Herkunft der Evaluatoren*

Dimensionen zu den Auftraggebern:

- *Politische Ebene* (Bund, Kanton, Gemeinde)
- *Art des Auftraggebers* (Exekutive/Legislative/Forschungsförderung)

Bei der auf diesen Dimensionen basierenden Analyse ist darauf zu achten, dass keine Überinterpretationen vorgenommen werden. Weder unsere Vorgehensweise noch die Zahl der untersuchten Evaluationen würden dazu ausreichen, weitergehende Schlüsse zu begründen. Es wird lediglich versucht, Anhaltspunkte für zu formulierende Thesen zu gewinnen.

Weiterhin wird thematisiert, inwiefern sich die Kriterien in der empirischen Analyse bewährt haben. Daraus sollen sich Hinweise darauf ergeben, ob sich die eingesetzten Kriterien zur Bewertung von Evaluationen, die in der Schweiz ausserhalb des Bildungssektors entstanden sind, eignen.

Auf eine Darstellung der Datenmatrix, die Auskunft darüber geben würde, welche Studie welches Einzelkriterium erfüllt bzw. nicht erfüllt hat, wurde bewusst verzichtet. Die nachfolgende Analyse ist auf die Gesamtheit der zehn untersuchten Evaluationsstudien ausgerichtet. Dafür ist eine summarische Zusammenfassung der Ergebnisse aus den zehn Fallstudien als Ganzes besser geeignet. Die Informationen zu den einzelnen Evaluationen sind in den Fallstudien enthalten (siehe Kapitel 5 sowie Widmer 1996a-g). Weiterhin könnte eine derartige Darstellung dazu verleiten, die Resultate in unzulässiger Weise miteinander zu verbinden. Einerseits ist darauf hinzuweisen, dass die im nachfolgenden Text angegebenen Bewertungen ein ordinales Messniveau aufweisen und dass es aus diesem Grund nicht zulässig ist, die Werte der einzelnen Kategorien zu einem Kriterium in irgendeiner Form zusammenzufassen. Ebenso wenig zweckmässig ist die Zusammenfassung der Bewertungen zu einer Studie über die Einzelkriterien hinweg, da dafür das relative Gewicht der einzelnen Kriterien zu bestimmen wäre (zur Problematik der Gewichtung, vgl. auch Kapitel 7.).

6.1.1. Nützlichkeit

6.1.1.1. Identifikation der beteiligten Gruppen

+: 2	(+): 5	+/-: 0	(-): 2	-: 0	/: 1
------	--------	--------	--------	------	------

Die Mehrheit der untersuchten Evaluationsstudien kommen dem Erfordernis nach, die Interessen der beteiligten und betroffenen Personen zu identifizieren. Gerade in zwei Fäl-

len, fällt die Bewertung eher negativ aus. Hier sind relevante Interessen nicht oder zu wenig berücksichtigt worden. In einer Fallstudie war es nicht möglich, eine Bewertung hinsichtlich dieses Kriteriums vorzunehmen, da uns nicht ausreichende Informationen zur Verfügung standen. Diese Lücke wäre durch eine vertiefende Analyse (die im Rahmen dieser Studie nicht möglich war) durchaus zu füllen. Hinsichtlich der in Kapitel 4.2. umrissenen Dimensionen zur Charakterisierung der zehn ausgewählten Fallstudien sind keine strukturellen Regelmässigkeiten erkennbar.

6.1.1.2. Glaubwürdigkeit des Evaluators

+: 5	(+): 1	+/-: 0	(-): 0	-: 1	/: 3
------	--------	--------	--------	------	------

Insgesamt betrachtet ergibt sich ein sehr positives Bild. Die Glaubwürdigkeit der Evaluatoren ist aufgrund unserer Einschätzung grösstenteils gegeben. In einem Fall traf dies nicht zu. Bei dieser Evaluation kumulierten sich verschiedene Aspekte. Neben dem schwierigen politischen Umfeld bestand das Evaluatorenteam zu einem Grossteil aus Personen mit wenig oder gar keiner praktischer Forschungserfahrung. Erschwerend kam hinzu, dass die Ziele der Evaluation den Adressaten der Studie in ungeeigneter Weise kommuniziert wurden. Diese Problematik wurde zwar im Verlauf des Evaluationsprozesses bemerkt, konnte jedoch nicht mehr vollständig beseitigt werden. In drei anderen Fällen war es uns nicht möglich, eine fundierte Einschätzung der Glaubwürdigkeit der Evaluatoren vorzunehmen, weil dazu zusätzliche Abklärungen erforderlich gewesen wären. Es hat sich im Rahmen der Meta-Evaluation gezeigt, dass die Glaubwürdigkeit des Evaluators schwieriger zu erreichen ist, wenn eine politisch hart umstrittene Massnahme evaluiert wird. Weil in unserer Untersuchung ausschliesslich externe Evaluationen betrachtet wurden, sind keine fundierten Aussagen zur Glaubwürdigkeit bei internen Evaluationen möglich.

6.1.1.3. Informationsrahmen und -auswahl

+: 5	(+): 3	+/-: 0	(-): 1	-: 1	/: 0
------	--------	--------	--------	------	------

Zu einem grossen Teil resultiert im Bereich des Informationsrahmens und der Informationsauswahl eine positive Bewertung. In zwei Fällen kam es zu einer (eher) negativen Bewertung. Bei diesen zwei Studien sind wir bereits im Rahmen der Fallstudien zum Ergebnis gelangt, dass sie in unserem Verständnis weniger als Evaluationsstudien, als vielmehr als Beratung oder Begleitforschung zu bezeichnen sind (vgl. dazu auch Kapitel 7.2.). In beiden Fällen reichen die erfassten Informationen nicht aus, um evaluative Fra-

gestellungen zu bearbeiten. Bezüglich der oben angeführten Dimensionen ist keine strukturelle Systematik erkennbar. Die zwei genannten 'Evaluationen' sind zwar beide hinsichtlich der zeitlichen und finanziellen Ressourcen eher als kleine Studie zu bezeichnen, verschiedene andere mit knappen finanziellen und zeitlichen Ressourcen ausgestattete Evaluationen, vermögen jedoch hinsichtlich dieses Kriteriums durchaus zu überzeugen.

6.1.1.4. Bewertende Interpretationen

+: 2	(+): 4	+/-: 0	(-): 0	-: 3	/: 1
------	--------	--------	--------	------	------

Mehrheitlich erfolgt eine positive Bewertung. In den drei Fällen mit negativer Bewertung lassen es die in den Schlussberichten angeführten Beschreibungen der verwendeten Perspektiven, Verfahren und Begründungen nicht zu, die Grundlagen für die Werturteile abzuschätzen. Hinsichtlich der unterschiedlichen Dimensionen lassen sich keine Regelmässigkeiten erkennen. In einem Fall war eine Bewertung nicht möglich, da in dieser Studie mehrheitlich darauf verzichtet wurde, Werturteile zu fällen. Das Kriterium geht jedoch von der Voraussetzung aus, dass in einer Evaluation Werturteile gefällt werden.

6.1.1.5. Klarheit des Berichts

+: 4	(+): 3	+/-: 0	(-): 0	-: 3	/: 0
------	--------	--------	--------	------	------

Im Gegensatz zum überaus positiven Gesamteindruck vermögen drei Studien dieses Kriterium nicht zu erfüllen. In allen drei Fällen handelt es sich um Projekte, die als verhältnismässig 'unpapieren' bezeichnet werden können. Die Schriftlichkeit spielte in diesen Fällen, im Gegensatz zur mündlichen Interaktion, eine geringere Rolle. Im Vergleich zu anderen Evaluationen kam dem Schlussbericht hier eine kleinere Bedeutung zu. Möglicherweise könnte im einen oder anderen Fall auch der Zeitdruck bei der Fertigstellung des Berichts ein wichtiges Moment dargestellt haben. Im Hinblick auf die angeführten Dimensionen lassen sich keine Regelmässigkeiten erkennen.

6.1.1.6. Verbreitung des Berichts

+: 7	(+): 1	+/-: 1	(-): 0	-: 0	/: 1
------	--------	--------	--------	------	------

Die zehn Fallstudien werden hinsichtlich der Verbreitung des Berichts insgesamt betrachtet sehr positiv bewertet. In einem Fall reichten die uns vorliegenden Informationen nicht

aus, um eine Bewertung abzugeben. Der Zugang zu relevanten Angaben war im Rahmen der vorliegenden Untersuchung nicht herzustellen. Die äusserst positive Bewertung in bezug auf die Verbreitung des Berichts ist mitunter darauf zurückzuführen, dass für die Meta-Evaluation ausschliesslich Studien ausgewählt wurden, die publiziert worden sind (vgl. oben, Kapitel 4.2.). Die nur marginalen Bewertungsdifferenzen lassen keine weitergehenden Interpretationen zu. In diesem Zusammenhang ist jedoch noch eine ergänzende Bemerkung anzuführen. Es ist a priori nicht von der Hand zu weisen, dass Evaluationsstudien, die nicht publiziert werden, im Gegensatz zu den hier untersuchten Evaluationen andere Charakteristiken aufweisen. Deshalb können die hier präsentierten Ergebnisse auf keinen Fall auf nicht-publizierte Evaluationen übertragen werden.

6.1.1.7. Rechtzeitigkeit des Berichts

+: 3	(+): 4	+/-: 2	(-): 0	-: 0	/: 1
------	--------	--------	--------	------	------

Auch bei diesem Kriterium sind keine negativen Bewertungen festzustellen, auch wenn die Bewertung insgesamt nicht ganz so positiv ausfällt, wie beim vorangehenden Kriterium. Bei grosszügiger Auslegung sind die Abschlusstermine eingehalten worden. Bei drei Studien gelang dies ohne irgendwelche Verzögerungen. Verschiedentlich mussten gewisse Verzögerungen in Kauf genommen werden. Die diesbezüglichen Unterschiede lassen sich nicht in systematischer Weise mit den Dimensionen in Verbindung bringen. In einem Fall war eine Einschätzung nicht möglich, da uns die dazu erforderlichen Informationen nicht zur Verfügung standen.

6.1.1.8. Wirkung der Evaluation

+: 1	(+): 0	+/-: 0	(-): 0	-: 1	/: 8
------	--------	--------	--------	------	------

Sehr oft - in acht von zehn Fällen - war eine Bewertung hinsichtlich der Wirkung der Evaluation nicht möglich. Im Rahmen der empirischen Analyse hat sich gezeigt, dass dieses Kriterium mit der hier gewählten Vorgehensweise im Regelfall nicht ausreichend gründlich untersucht werden konnte. Aufgrund des Forschungsstandes zur Praxisverwendung sozialwissenschaftlicher Wissensbestände stellt sich zudem die Frage, ob eine derartige Bewertung - auch bei einer vertieften Analyse - überhaupt möglich ist. Auch bei der hier negativ bewerteten Evaluation ist nicht auszuschliessen, dass mittel- oder langfristig indirekte Wirkungen auftreten könnten. Nur entziehen sich indirekte Wirkungen oft einer empirischen Beobachtung. Hinzu kommt der Umstand, dass verschiedene der untersuchten Evaluationsstudien erst vor relativ kurzer Zeit abgeschlossen worden sind, was

die Abklärung der Wirkungen zum jetzigen Zeitpunkt zusätzlich relativiert. Die wenigen inhaltlichen Angaben zu diesem Kriterium - nur zwei Evaluationen konnten bewertet werden - sind für weitergehende Interpretationen ungeeignet.

6.1.2. Anwendbarkeit

6.1.2.1. Praktische Verfahren

+: 5	(+): 3	+/-: 2	(-): 0	-: 0	/: 0
------	--------	--------	--------	------	------

Insgesamt resultiert aufgrund der empirischen Analysen ein sehr positives Bild. Bei einer jener Evaluationen, welche das Kriterium nur teilweise erfüllen, hat man sich zu Beginn des Evaluationsprozesses dazu entschlossen, auf primäre Datenerhebungen im quantitativen Bereich zu verzichten, da die zur Verfügung stehenden Ressourcen dafür zu knapp erschienen sind. Da die vorliegenden quantitativen Daten, auf die man in einer Sekundäranalyse zurückgreifen wollte, nicht im gewünschten Umfang oder in der erhofften Qualität vorlagen, fehlten für die Analyse erforderliche Informationen. Da dies vor allem auch den Wirkungsaspekt betraf, konnte die Wirkungsevaluation nicht wie geplant durchgeführt werden. Bei der zweiten Studie, die das Kriterium nur teilweise erfüllt hat, lag dies primär an den äusserst engen zeitlichen wie finanziellen Ressourcen. Die Abstriche bei der Bewertung dieses Kriteriums betreffen gesamthaft betrachtet in erster Linie quantitative Untersuchungsteile und dabei insbesondere die Datenlage. Im Gegensatz dazu haben sich qualitative Vorgehensweisen als sehr praktisch und äusserst flexibel erwiesen. Bewährt haben sich zudem Untersuchungsanlagen, die methodisch breit abgestützt sind und sich nicht auf die Leistungsfähigkeit einer einzelnen Methodik verlassen.

6.1.2.2. Politische Überlebensfähigkeit

+: 1	(+): 4	+/-: 0	(-): 0	-: 0	/: 5
------	--------	--------	--------	------	------

Auch hinsichtlich dieses Kriteriums stellten sich ähnlich gelagerte Probleme, wie sie bereits beim Kriterium zur Wirkung (vgl. oben Kapitel 6.1.1.8.) dargestellt wurden. In fünf Fällen musste auf eine Bewertung verzichtet werden. Die übrigen Evaluationen sind mit graduellen Unterschieden positiv bewertet worden. Aufgrund dieser Konstellation erübrigen sich an dieser Stelle weitere Bemerkungen.

6.1.2.3. Kostenwirksamkeit

+: 2	(+): 2	+/-: 2	(-): 0	-: 0	/: 4
------	--------	--------	--------	------	------

Die Bewertungen zur Kostenwirksamkeit sind mit leicht positiver Tendenz recht aus geglichen ausgefallen. Bei vier Studien musste aber auf eine Bewertung verzichtet werden, da ausreichende Informationen bezüglich der Wirksamkeit fehlten. Generell gesehen stellt sich natürlich auch bei diesem Kriterium dieselbe Problematik, wie wir sie bereits oben in bezug auf die Evaluationswirkungen diskutiert haben (siehe Kapitel 6.1.1.8. und 6.1.2.2.). Verstärkend kommt hinzu, dass eine Einschätzung der Kostenwirksamkeit weiter dadurch erschwert wird, da hier zusätzlich der Wert der erarbeiteten Informationen in Relation zu den Kosten des Projekts gestellt werden muss. Deshalb sind die nachfolgenden Hinweise mit Vorbehalten zu versehen. Betrachten wir uns diese Ergebnisse auf dem Hintergrund der oben angeführten Dimensionen, fällt auf, dass hinsichtlich Zeit und Finanzen kleine Evaluationsstudien das Kriterium tendenziell eher zu erfüllen vermögen, als dies grössere Untersuchungen tun. Dies könnte auch damit in Zusammenhang stehen, dass die aufwendigeren Evaluationsstudien im Rahmen der Forschungsförderung entstanden sind, die sich auch an methodischen und akademischen Zielen orientiert, währenddem die eher knapp ausgestatteten Studien durch die Praxis initialisiert wurden. Es zeigt sich aufgrund der zehn untersuchten Studien weiter, dass traditionell orientierte Evaluationen bezüglich dieses Kriteriums leicht besser abschneiden, als dies alternativ orientierte Studien tun. Dies muss wohl in Zusammenhang damit gestellt werden, dass alternative Ansätze häufig recht aufwendige Datenerhebungstechniken einsetzen, dass die traditionellen Studien eher pragmatisch orientiert sind und dass es möglicherweise schwerer fällt, die Wertschätzung für die Erkenntnisse alternativ positionierter Studien empirisch zu erfassen.

6.1.3. Korrektheit

6.1.3.1. Formale Verpflichtung

+: 3	(+): 4	+/-: 2	(-): 0	-: 0	/: 1
------	--------	--------	--------	------	------

In einem Fall war hier keine Bewertung möglich, da uns kein Evaluationskontrakt oder eine vergleichbare Vereinbarung vorlag. Auch jene Evaluationen, die das Kriterium nur teilweise erfüllt haben, sind in grossen Teilen den formalen Verpflichtungen gefolgt. Im Gegensatz zu den besser bewerteten Evaluationen wiesen diese zwei Studien jedoch in gewissen Bereichen massgebliche Abweichungen oder Lücken auf. Oft war dies eine

Folge der Rahmenbedingungen, unter denen die Evaluationen durchgeführt wurden. Zum Teil haben wir den Eindruck erhalten, dass die Angaben im Evaluationskontrakt oft - unter Einhaltung eines hohen Qualitätsstandards - über den Möglichkeiten lagen, die sich unter den gegebenen Restriktionen realisieren liessen. In diese Richtung weist auch der Umstand, dass eine grosse Zahl der Evaluationen nicht im festgelegten finanziellen Rahmen abgeschlossen werden konnten. Auch wenn in einigen Fällen seitens des Auftraggebers Zusatzkredite gesprochen wurden, kam es in den meisten Fällen auch zu (zum Teil massiven) Kostenüberschreitungen, die durch die Evaluatoren selbst getragen wurden (vgl. Kapitel 7.2.).

6.1.3.2. Deklaration von Interessenkonflikten

+: 1	(+): 3	+/-: 0	(-): 3	-: 0	/: 3
------	--------	--------	--------	------	------

Die Bewertung der zehn Evaluationen bezüglich der Deklaration von Interessenkonflikten fällt ausgeglichen aus. In drei Fällen war eine Bewertung nicht möglich, da die dafür erforderlichen Informationen im Rahmen der vorliegenden Untersuchung nicht beschafft werden konnten. Strukturelle Zusammenhänge zu den verschiedenen Charakteristiken der Evaluationsstudien sind nicht festzumachen. Die negativeren Bewertungen stehen primär damit in Zusammenhang, dass zentrale Interessenkonflikte nicht oder kaum thematisiert worden sind, obwohl diese sowohl das Evaluandum als auch die Evaluation massgeblich betroffen haben.

6.1.3.3. Vollständige und aufrichtige Offenlegung

+: 3	(+): 3	+/-: 2	(-): 1	-: 1	/: 0
------	--------	--------	--------	------	------

Die festgestellten Mängel liegen bei diesem Kriterium vor allem darin, dass die Grenzen der Aussagekraft der Evaluationsergebnisse zu wenig deutlich dargelegt wurden. Im Kontrast dazu haben vor allem jene Studien positive Bewertungen erhalten, die deutlich auf die Grenzen der Evaluation hinweisen. Zu den verschiedenen Dimensionen zur Charakterisierung der Evaluationsstudien ergeben sich keine systematischen Zusammenhänge.

6.1.3.4. Herstellung von Öffentlichkeit

+: 6	(+): 0	+/-: 1	(-): 0	-: 2	/: 1
------	--------	--------	--------	------	------

Die zwei negativen Bewertungen sind primär darauf zurückzuführen, dass in diesen beiden Fällen die Evaluationsberichte weniger dazu geeignet sind, Öffentlichkeit herzustellen, da diese hinsichtlich ihres Inhalts für ein breiteres Publikum nicht verständlich sind, weil Teile der Berichterstattung nicht öffentlich zugänglich sind oder weil in bezug auf die Evaluation eine sehr zurückhaltende Informationspolitik betrieben wurde bzw. wird. Gesamthaft betrachtet ergibt sich aber ein positives Bild, das unter anderem damit im Zusammenhang steht, dass für die vorliegende Untersuchung ausschliesslich Evaluationen ausgewählt wurden, die, dadurch dass sie im Prinzip publiziert wurden, als öffentlich bezeichnet werden können (vgl. dazu oben, Kapitel 6.1.1.6.). In einem Fall konnte keine Bewertung vorgenommen werden, da die uns vorliegenden Informationen dazu nicht ausreichten. Hinsichtlich der verschiedenen Dimensionen zur Charakterisierung der Evaluationen fällt auf, dass Evaluationen mit knapp bemessenem Zeitbudget eher dazu neigen, bei diesem Kriterium schwächer abzuschneiden. Diese Beobachtung zu verallgemeinern wäre jedoch verfehlt, liegen doch aus unserer Stichprobe auch Studien vor, die über sehr enge zeitliche Restriktionen verfügten und hinsichtlich dieses Kriteriums eine positive Bewertung erhalten haben. Umgekehrt zeigt sich aber, dass unter den länger konzipierten Studien alle eine positive Bewertung erhalten haben. Ausserhalb der angeführten Dimensionen fällt in diesem Zusammenhang auf, dass Evaluationen, die als Projekte zur Forschungsförderung bezeichnet werden können (also die NFP 27-Projekte (ohne Kurzevaluationen), sowie die im Rahmen der Arbeiten der AGEVAL entstandenen Studien) - im Gegensatz zu einigen Auftragsevaluationen -, ausnahmslos positiv bewertet werden.

6.1.3.5. Schutz individueller Menschenrechte

+: 9	(+): 0	+/-: 0	(-): 0	-: 0	/: 1
------	--------	--------	--------	------	------

Hinsichtlich des Kriteriums 'Schutz individueller Menschenrechte' sind neun der zehn Evaluationen positiv bewertet worden. Lediglich in einem Fall konnte keine Bewertung vorgenommen werden, da dazu zusätzliche Abklärungen erforderlich wären. Aufgrund dieser Konstellation erübrigen sich hierzu weitere Anmerkungen.

6.1.3.6. Menschliche Interaktionen

+: 2	(+): 1	+/-: 1	(-): 0	-: 0	/: 6
------	--------	--------	--------	------	------

Die wenigen kritischen Bemerkungen betreffen primär die Verwendung der Informationen aus mündlichen oder schriftlichen Befragungen (Wahrung der Anonymität, Publikation der Ergebnisse). Dass in sechs Fällen keine Bewertung möglich war, stand primär damit im Zusammenhang, dass im Rahmen dieser Meta-Evaluation häufig kein ausreichender Einblick in die Interaktionen im Rahmen der evaluierten Evaluationen möglich war. Dies wäre einerseits dadurch zu verbessern, dass vertiefende Interviews mit den beteiligten Personen geführt würden. Andererseits würde sich diese Problematik reduzieren lassen, indem nicht eine ex post-Untersuchungsanlage gewählt, sondern eine begleitende Meta-Evaluation durchgeführt würde (vgl. dazu auch Kapitel 7.3.).

6.1.3.7. Ausgewogene Berichterstattung

+: 3	(+): 4	+/-: 2	(-): 1	-: 0	/: 0
------	--------	--------	--------	------	------

Die Mehrheit der zehn Evaluationen wird in bezug auf die Ausgewogenheit der Berichterstattung positiv bewertet. Die negativen Aspekte betreffen zum einen eine eher unkritische Perspektive gegenüber dem Evaluandum und andererseits den Ansatz der Schwachstellenanalyse, der dazu beitragen kann, negative Punkte überzubewerten und positive Feststellungen zu vernachlässigen. Zusätzlich kann die Abgrenzung des Untersuchungsgegenstandes oder die Definition der Fragestellung (etwa Verzicht auf die Untersuchung (negativer) Nebenwirkungen) die Ausgewogenheit der Berichterstattung tangieren. Systematische Bezüge zu den Dimensionen zur Charakterisierung der Evaluationen sind keine festzustellen.

6.1.3.8. Finanzielle Verantwortlichkeit

+: 5	(+): 0	+/-: 0	(-): 1	-: 1	/: 3
------	--------	--------	--------	------	------

Entgegen dem positiven Gesamteindruck konnten zwei Evaluationen das Kriterium nicht erfüllen. Während in einem dieser Fälle die Ausrichtung der Projektarbeiten in einer ersten Evaluationsphase nicht mit den eigentlichen Intentionen des Projekts übereinstimmten, wurden im anderen Fall relativ kostenintensive Vorgehensweisen gewählt, die wenig zur Klärung einer evaluativen Fragestellung beitragen konnten oder zu qualitativ unbefriedigenden Ergebnissen führten. In drei Fällen war uns eine Bewertung nicht mög-

lich, da die uns zur Verfügung stehenden Angaben zur Verwendung der finanziellen Ressourcen zu spärlich vorlagen. Strukturelle Zusammenhänge zu den genannten Dimensionen sind bei der Bewertung dieses Kriteriums keine festzustellen.

6.1.4. Genauigkeit

6.1.4.1. Identifikation des Gegenstandes

+ : 2	(+) : 4	+/- : 2	(-) : 2	- : 0	/ : 0
-------	---------	---------	---------	-------	-------

Bei der Bewertung durch das Kriterium 'Identifikation des Gegenstandes' resultiert ein leicht positives Gesamtbild. Die Schwächen beziehen sich auf sehr unterschiedliche Aspekte. So gelingt es nicht immer, den Untersuchungsgegenstand ausreichend präzise und konsequent von anderen Gegenständen abzugrenzen. Eine Deskription des Evaluandums und dessen Vollzug wird bei einigen Studien nicht oder nicht ausreichend detailliert erarbeitet und dargestellt. Weiterhin zu Problemen führte der Umstand, dass in bestimmten Fällen erforderliche Differenzierungen nicht vorgenommen wurden. Zum Teil sind diese Mängel auf zu wenig präzise definierte Fragestellungen zurückzuführen. Eine weitere Ursache sehen wir darin, dass eine fundierte Identifikation des Gegenstandes, wie sie vom Kriterium postuliert wird, unter den teilweise gegebenen Rahmenbedingungen (zeitliche und finanzielle Ressourcen, aber auch die Komplexität des zu evaluierenden Gegenstandes) nicht immer in vollem Umfang zu leisten war. Im Hinblick auf die oben angeführten Dimensionen ergeben sich keine systematischen Bezüge.

6.1.4.2. Kontextanalyse

+ : 3	(+) : 1	+/- : 1	(-) : 0	- : 5	/ : 0
-------	---------	---------	---------	-------	-------

Die Hälfte der untersuchten Evaluationen erfüllt dieses Kriterium mehrheitlich oder vollständig. Die andere Hälfte erfüllt das Kriterium nicht. Die negativen Aspekte sind von Studie zu Studie sehr unterschiedlich. Aufgrund dieser Spezifität fällt es schwer, allgemeinere Aussagen zu formulieren. Die Lücken und Mängel der Kontextanalyse beziehen sich unter anderem auf institutionelle, wirtschaftliche, politische, föderale, soziale aber auch historische Elemente des Kontextes. Möglichen Beeinflussungen des Evaluandums und seiner Wirkungen durch den Kontext wird in verschiedenen Evaluationen zu wenig Beachtung geschenkt. Teilweise stand wiederum der Umfang der zu leistenden Abklärungen nicht in einer vernünftigen Relation zu den vorhandenen Mitteln, so dass in einigen

Fällen aus diesem Grund auf eine vertiefte Kontextanalyse verzichtet werden musste. Im Hinblick auf die verschiedenen Dimensionen zur Charakterisierung der Evaluationsstudien sind keine systematischen Muster zu beobachten.

6.1.4.3. Beschreibung von Zielen und Vorgehen

+: 0	(+): 3	+/-: 0	(-): 3	-: 4	/: 0
------	--------	--------	--------	------	------

Das negative Gesamtergebnis deutet darauf hin, dass in der Mehrzahl der Fälle die Nachvollziehbarkeit nicht ausreichend gegeben ist. Die Mängel beziehen sich in erster Linie auf die Beobachtung und Beschreibung des Vorgehens und weniger auf jene der Ziele. Dies betrifft in besonders starkem Ausmass die qualitativen Untersuchungsteile, währenddem bei quantitativen Analysen die Lücken, allgemein betrachtet, weniger gross sind. Weiter ist festzustellen, dass dabei insbesondere die *Datenauswertung* betroffen ist. Zur *Datenerhebung* sind in den Schlussberichten zumeist mehr Informationen enthalten. Bezüglich der oben angeführten Dimensionen sind keine systematischen Zusammenhänge zu beobachten.

6.1.4.4. Verlässliche Informationsquellen

+: 0	(+): 5	+/-: 1	(-): 0	-: 4	/: 0
------	--------	--------	--------	------	------

Die Abstriche bezüglich der Verlässlichkeit der Informationsquellen gestalten sich in den verschiedenen tangierten Evaluationen wiederum sehr unterschiedlich. In qualitativen Untersuchungen, in denen persönliche Interviews oder Gruppendiskussionen als Erhebungsinstrumente eingesetzt werden, zeigt sich das Problem, dass es äusserst schwierig ist, die Relevanz der erhobenen Informationen (zumeist unter Wahrung der Anonymität der Quellen) und deren Bezug zu den Evaluationsresultaten transparent darzustellen. Neben dieser speziell bei qualitativen Untersuchungsteilen präsenten Problematik, sind jedoch auch Mängel festzustellen, die allgemeiner Art sind. In einigen Fällen sind innerhalb des Schlussberichts Inkonsistenzen zu beobachten, welche die Verlässlichkeit der Informationsquellen in Frage stellen. Bei bestimmten Evaluationen sind wir bei der Überprüfung der Informationsquellen verhältnismässig häufig auf Ungenauigkeiten oder Unstimmigkeiten gestossen. Da die hierfür notwendigen Angaben fehlen, ist die Verlässlichkeit der Informationsquellen gegen Aussen verschiedentlich nicht gegeben. Ausreichend deutliche Zusammenhänge zu den verschiedenen, oben angeführten Dimensionen sind nicht zu eruieren.

6.1.4.5. Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit

+: 0	(+): 1	+/-: 1	(-): 3	-: 3	/: 2
------	--------	--------	--------	------	------

Der Vergleich entlang dieses Kriteriums wird dadurch erschwert, dass nicht für alle zehn Evaluationsstudien identische Kriterien eingesetzt wurden. Aufgrund der paradigmatischen Position wurden die Evaluationen entweder hinsichtlich des Kriteriums 'Valide Messung' (traditionelles Paradigma) oder dann mit den zwei Kriterien 'Glaubwürdigkeit' und 'Übertragbarkeit' (alternatives Paradigma) bewertet. In zwei Fällen musste ganz auf eine Bewertung verzichtet werden. Bei einer dieser beiden Studien reichten die für unsere Untersuchung vorliegenden Informationen nicht aus, um eine ausreichend abgestützte Bewertung vorzunehmen. Mittels weitergehender Abklärungen - die im Rahmen der vorliegenden Untersuchung nicht zu realisieren waren - wäre hier aber grundsätzlich eine Bewertung möglich. Im zweiten Fall liegt die Ursache für das Fehlen einer Bewertung tiefer. Da es uns nicht möglich war, die paradigmatische Orientierung dieser Evaluation zu bestimmen, musste hier auf eine Bewertung verzichtet werden. Diese Problematik zeigte sich für diese Studie in gleicher Weise auch bei zwei weiteren Kriterien (vgl. unten Kapitel 6.1.4.6. und 6.1.4.11.).

6.1.4.6. Reliable Messung bzw. Verlässlichkeit

+: 1	(+): 0	+/-: 2	(-): 0	-: 3	/: 4
------	--------	--------	--------	------	------

Da auch hier verschiedene Kriterien eingesetzt wurden und uns in vier Fällen keine Bewertung möglich war, erscheint ein Vergleich nicht zweckmässig (vgl. für weitere Ausführungen Kapitel 6.1.4.5.).

6.1.4.7. Systematische Datenüberprüfung

+: 3	(+): 1	+/-: 1	(-): 0	-: 2	/: 3
------	--------	--------	--------	------	------

Die Bewertung der zehn Fallstudien präsentiert sich mit leicht positiver Tendenz ausgeglichen. Die negativen Bewertungen stehen vor allem in Zusammenhang mit einer teilweise recht hohen Zahl an Unstimmigkeiten, die in den betreffenden Schlussberichten enthalten sind. Dieser Umstand lässt indirekt darauf schliessen, dass keine oder auf alle Fälle keine ausreichend gründliche systematische Datenüberprüfung durchgeführt wurde. Auffallend ist, dass in den Studien kaum explizite Angaben zu einer systematischen Datenüberprüfung erfolgen. Hier besteht bezüglich der Berichterstattung Handlungsbedarf

(siehe dazu Kapitel 7.2.). Aus diesem Grund, aber auch weil es im Rahmen dieser Untersuchung nicht möglich war, weitere Abklärungen vorzunehmen, musste in vier Fällen auf eine Bewertung verzichtet werden. In Relation zu den genannten Dimensionen ergeben sich keine ausreichend abgestützten Bezugsmuster.

6.1.4.8. Analyse quantitativer Informationen

+: 0	(+): 4	+/-: 1	(-): 1	-: 3	/: 1
------	--------	--------	--------	------	------

Der Vergleich hinsichtlich des Kriteriums 'Analyse qualitativer Informationen' ergibt ein tendenziell negativ gefärbtes Bild. Bei einer Studie musste auf eine Bewertung verzichtet werden, da sich diese praktisch ausschliesslich auf den Einsatz qualitativer Analyseverfahren konzentriert und damit eine Bewertung anhand dieses Kriteriums nicht adäquat wäre. Die festgestellten Schwachpunkte betreffen sehr unterschiedliche Aspekte. Unter anderem fehlen teilweise wichtige Angaben, um das quantitative Datenmaterial beurteilen zu können (beispielsweise Prozentuierungen ohne Angabe der zugrundeliegenden Basis). In einigen Studien werden bestimmte statistische Masszahlen falsch eingesetzt (bspw. Standardabweichung anstatt Variationskoeffizient). Teilweise entsprechen die Ausführungen im Text nicht den Angaben in Tabellen oder Darstellungen. In einem Fall werden grosse Datenmengen in Tabellen angeführt, ohne dass eine adäquate Interpretation und Diskussion erfolgen würde. Verschiedentlich basieren Analysen auf sehr kleinen Fallzahlen, ohne dass auf diesen Umstand hingewiesen wird. Ein weiteres Beispiel stellen unzulässige oder zumindest problematische Datenmodifikationen dar (wie etwa die Multiplikation ordinaler Werte oder die Zusammenfassung sachfremder Kategorien).

6.1.4.9. Analyse qualitativer Informationen

+: 0	(+): 1	+/-: 2	(-): 0	-: 0	/: 7
------	--------	--------	--------	------	------

Sieben der zehn Studien konnten mit diesem Kriterium nicht bewertet werden. In sechs dieser Fälle ist dies auf den Umstand zurückzuführen, dass die Berichterstattung zum empirischen Vorgehen insbesondere im qualitativen Bereich zum Teil erhebliche Lücken aufweist. Besonders stark ist dies bei der Auswertung - weniger stark bei der Erhebung - qualitativer Daten zu beobachten. Da in einer Studie aufgrund der uns zur Verfügung stehenden Informationen gänzlich auf qualitative Analysen verzichtet wurde, konnte hier auch zu dieser Studie keine zweckmässige Bewertung erarbeitet werden. Aufgrund der

kleinen Zahl an bewertbaren Studien verzichten wir an dieser Stelle auf weitere Hinweise.

6.1.4.10. Gerechtfertigte Folgerungen

+: 0	(+): 4	+/-: 2	(-): 1	-: 3	/: 0
------	--------	--------	--------	------	------

Bei vier der zehn evaluierten Evaluationsstudien resultiert aufgrund der empirischen Analyse eine negative oder eher negative Bewertung. Dies ist primär darauf zurückzuführen, dass es in den Evaluationen verschiedentlich nicht gelingt, die angeführten Folgerungen mit den durchgeführten empirischen Analysen zu verbinden. Sehr oft bestehen diesbezüglich Lücken. Ein weiteres Problem besteht darin, dass in den Schlussfolgerungen teilweise Aussagen enthalten sind, die weniger als Folgerungen sondern zutreffender als weitergehende Überlegungen, Vermutungen oder Thesen zu bezeichnen wären. Da dies nicht - oder nicht ausreichend gründlich - getan wird, kommt der falsche Eindruck auf, dass es sich bei diesen Aussagen um empirisch abgestützte Evaluationsergebnisse handelt. Um Missverständnissen vorzubeugen, sei an dieser Stelle nochmals deutlich hervorgehoben, dass das Kriterium postuliert, dass Folgerungen *explizit begründet* sein müssen. Es ist also nicht zulässig, aufgrund der insgesamt eher negativen Bewertung zu folgern, dass in den Evaluationen mehrheitlich unzulässige Folgerungen formuliert worden wären. Es lässt sich lediglich sagen, dass in einigen Studien explizite Begründungen zu den Folgerungen fehlen.

6.1.4.11. Objektive Berichterstattung bzw. Bestätigbarkeit

+: 1	(+): 3	+/-: 1	(-): 0	-: 1	/: 4
------	--------	--------	--------	------	------

Auch hier sind, je nach paradigmatischer Orientierung der Evaluationen, zwei verschiedene Kriterien zur Bewertung der Evaluationsstudien eingesetzt worden. Zudem konnte in vier Fällen aufgrund der uns zur Verfügung stehenden Informationen keine Bewertung abgegeben werden. Aus diesen Gründen ist ein Vergleich nicht zweckmässig (vgl. dazu ausführlicher Kapitel 6.1.4.5.).

6.2. Vergleich entlang der Kriteriengruppen

In diesem Kapitel wird in einem ersten Schritt jeweils untersucht, wie sich die Bewertung bezüglich einer Kriteriengruppe im allgemeinen (also über alle zehn Fallstudien hinweg) gestaltet. In einem zweiten Schritt wird besonders darauf geachtet, ob bezüglich einzelner Kriterien einer Gruppe besondere Abweichungen festzustellen sind. Im dritten Schritt wird das Augenmerk darauf gerichtet, welche Evaluationsstudien bezüglich einer bestimmten Kriteriengruppe stark abweichende Bewertungen erhalten haben. Zum Schluss wird bei jeder Kriteriengruppe untersucht, ob sich bezüglich der Anwendung einzelner Kriterien in unserer Studie besondere Schwierigkeiten ergaben.

6.2.1. Nützlichkeit

Im Bereich der Nützlichkeit wird die Mehrheit der zehn untersuchten Evaluationsstudien positiv bewertet. Dies gilt in besonderem Masse für die Einzelkriterien 'Verbreitung des Berichts' (Kriterium 1.6.), 'Informationsrahmen und -auswahl' (1.3.), 'Rechtzeitigkeit des Berichts' (1.7.) und 'Glaubwürdigkeit des Evaluators' (1.2.). Bei der Verbreitung des Berichts ist dies direkt rückführbar auf unsere Vorgehensweise bei der Auswahl der Studien. Dort wurde zur Bedingung gestellt, dass die zu berücksichtigenden Evaluationen in irgendeiner Form publiziert sein müssen. Dadurch sind Studien, die nie veröffentlicht wurden, in unserer Stichprobe gar nicht vertreten. 'Informationsrahmen und -auswahl' sind in den von uns untersuchten Studien zumeist in adäquater Weise bestimmt worden. Die Berichterstattung zu den Evaluationen ist mehrheitlich zeitgerecht erfolgt. Die 'Glaubwürdigkeit des Evaluators' ist meistens gegeben, auch wenn möglicherweise hier eine vielleicht etwas zu unkritische Bewertung unsererseits zum positiven Ergebnis beigetragen haben könnte. Im Vergleich dazu eher schwächer fällt die Bewertung hinsichtlich der Kriterien 'Wirkung der Evaluation' (Kriterium 1.8.), 'Bewertende Interpretation' (1.4.), 'Klarheit des Berichts' (1.5.) und 'Identifikation der beteiligten Gruppen' (1.1.) aus. Unsere Bewertung der 'Wirkung der Evaluation' ist wenig aussagekräftig, stützt sie sich doch lediglich auf die Bewertung zweier Studien ab. In den übrigen Fällen musste auf eine Bewertung verzichtet werden. Bezüglich der Identifikation der 'stakeholders' fallen unseres Erachtens zwei Faktoren ins Gewicht. Erstens enthält unsere Stichprobe einige Studien, die lediglich mit relativ bescheidenen finanziellen und zeitlichen Ressourcen auszukommen hatten. Hier fehlte vermutlich oftmals die Zeit, um sich mit den relevanten Interessen auseinanderzusetzen. Zweitens wird auffällig häufig darauf verzichtet, die Identifikation der beteiligten und betroffenen Individuen, Gruppen oder Institutionen des evaluierten Programms ausreichend gründlich vorzunehmen.

Besonders positiv bewertet wurde die Nützlichkeit der folgenden Studien:

- *'Umweltpolitik und technische Entwicklung'* (Studie Nr. 1): Diese Evaluation überzeugt in hohem Masse durch eine breit abgestützte und zweckmässige Informationsauswahl sowie durch eine gelungene Berichterstattung.
- *'Schneeräumung auf Urner Passstrassen'* (Studie Nr. 6): Dieser (mit nur sehr kleinem Budget ausgestatteten) Evaluation gelingt es, durch eine geradlinige Vorgehensweise und eine verhältnismässig gut handhabbare Fragestellung sowie einem vorzüglichen Schlussbericht die notwendigen Informationen zu beschaffen und zu vermitteln.
- *'Evaluation Kehrichtsackgebühr'* (Studie Nr. 7): Trotz einer recht umfangreichen Fragestellung vermag diese Studie durch eine pragmatisch angelegte Informationsbeschaffung und dem praxisnahen Schlussbericht, eine gute Abdeckung des Untersuchungsgegenstandes sowie eine kommunizierbare Untersuchung sicherzustellen.
- *'Evaluation des BUWAL'* (Studie Nr. 9): Durch ihren aktionsorientierten Ansatz, die vielfältige Informationsbeschaffung und die gelungene Schlussberichterstattung ist die Evaluation in der Lage, zielgerichtete Information zu vermitteln.

Weniger gut bewertet werden bezüglich der Nützlichkeit vor allem jene Projekte, bei denen wir zur Ansicht gekommen sind, dass es sich dabei weniger um Evaluationen, als vielmehr um andere Formen angewandter Forschung bzw. Beratung handelt. Dies gilt speziell für einige Studien mit einer starken Prozess- sowie Beratungsorientierung. Sehr oft sind die Projekte, die hier weniger gut abschneiden, wegen ihrer Grundkonzeption nicht in der Lage, verschiedene der hier relevanten Nützlichkeits-Kriterien zu erfüllen.

Hinsichtlich der Kriterien zur Nützlichkeit zeigte sich im Rahmen dieser Meta-Evaluation, dass die Bewertung der *Wirkung* empirischer Evaluationen mit der gewählten Vorgehensweise zumeist nicht geleistet werden konnte. Die grosse Mehrheit der Studien konnte hinsichtlich ihrer Wirkung nicht bewertet werden. Zumeist würde dies nach einer weitergehenden Untersuchung verlangen, die vertiefter auf eine einzelne Evaluation eingehen kann, als dies im Rahmen dieses Projekts möglich war. (Vgl. dazu unten, Kapitel 7.3.)

6.2.2. Anwendbarkeit

Zur Bewertung der Anwendbarkeit einer Evaluation standen lediglich drei Einzelkriterien zur Verfügung, was die nachfolgenden Ausführungen etwas relativiert, ist doch die empi-

rische Grundlage aufgrund der kleinen Zahl von Kriterien schwächer als bei den übrigen Kriteriengruppen.

Generell resultiert auch hier eine positive Bewertung. Dies gilt in besonders ausgeprägter Weise für das Kriterium 'Praktische Verfahren' (Kriterium 2.1.). Allgemein betrachtet ist man darauf bedacht, in Evaluationen Verfahren einzusetzen, die verhältnismässig robust und mit kleinen Risiken verbunden sind.

Bezüglich der Anwendbarkeit hat vor allem die *'Evaluation des BUWAL' (Studie Nr. 9)* sehr positive Bewertungen erhalten. Hier ist es mit einem konzeptionell durchdachten Design gelungen, eine Untersuchung zu realisieren, die Informationen beschaffen konnte, die im Vergleich zum Aufwand einen grossen Wert aufweisen. Die Untersuchung konnte dabei aber auch in zweierlei Hinsicht von positiven Rahmenbedingungen profitieren: Erstens war das Evaluatorenteam aus Personen zusammengesetzt, die bereits bei Projektbeginn einen sehr hohen Wissensstand aufwiesen. Zweitens hat die Art des Auftraggebers sicherlich dazu beigetragen, die politische Überlebensfähigkeit zu erhöhen.

Im Gegensatz dazu haben sich die Evaluationen *'Verbrauchsabhängige Heizkostenabrechnung - Förderung erneuerbarer Energien' (Studie Nr. 8)* und *'Wohneigentumsförderung durch den Bund' (Studie Nr. 4)* hinsichtlich der Anwendbarkeit weniger bewährt. Bei der ersten Studie wirkte sich der Umstand aus, dass sie zumindest zu Beginn als Implementations- und nicht als Evaluationsstudie angelegt war. Weiter hatte sich diese Studie in einem, der sozialwissenschaftlichen Arbeitsweise gegenüber eher kritisch eingestellten und zum Zeitpunkt der Evaluation stark technisch dominierten Umfeld zu bewegen. Bei der zweiten Studie fiel vor allem ins Gewicht, dass sie sich mit einem recht komplexen Themenbereich zu befassen hatte, dessen Bearbeitbarkeit durch eine weite Fragestellung zusätzlich erschwert wurde. Zudem wurde ein Untersuchungsdesign eingesetzt, das verhältnismässig aufwendig und mit recht hohen Risiken verbunden ist, was sich, da Schwierigkeiten im Bereich der Datenlage bestanden, in diesem Fall negativ auswirkte.

Bei der Bewertung der Kriterien 'Politische Überlebensfähigkeit' (Kriterium 2.2.) und 'Kostenwirksamkeit' (2.3.) musste recht oft auf eine Einschätzung verzichtet werden. Dies steht auch in Zusammenhang damit, dass verschiedene der untersuchten Evaluationen erst vor relativ kurzer Zeit beendet wurden, was zu Schwierigkeiten bei der Bewertung dieser Kriterien führen kann. Einen weiteren Grund sehen wir darin, dass die zur Einschätzung der Studien hinsichtlich der beiden genannten Kriterien erforderlichen Informationen nur sehr schwierig beschafft werden können. Im Rahmen unserer Untersuchung konnten wir dies teilweise nicht realisieren. (Vgl. Kapitel 7.3.)

6.2.3. Korrektheit

Die Bewertungen bezüglich der Korrektheit fielen insgesamt positiv aus. Dies gilt vor allem für das Kriterium 'Schutz individueller Menschenrechte' (Kriterium 3.5.). Aufgrund unserer Untersuchung hat hier keine der untersuchten Evaluationen eine negative Bewertung erhalten. Nicht ganz so stark, aber immer noch überdurchschnittlich fielen die Bewertungen zu den Kriterien 'Formale Verpflichtung' (Kriterium 3.1.), 'Ausgewogene Berichterstattung' (3.7.) und 'Herstellung von Öffentlichkeit' (3.4.) aus. In allen untersuchten Evaluationen bestand eine in schriftlicher Form abgefasste gegenseitige Verpflichtung, die zu einem sehr grossen Teil auch eingehalten wurde. Die positiven Resultate zum Kriterium 'Ausgewogene Berichterstattung' deuten darauf hin, dass die evaluierten Studien in der Regel keine stark einseitige Sichtweise verfolgten. Zur 'Herstellung von Öffentlichkeit' lässt sich darauf hinweisen, dass die Bewertung aufgrund unserer Vorgehensweise bei der Auswahl der Studien in positiver Weise beeinflusst wurde (vgl. auch oben 6.2.1.).

Hinsichtlich der Kriterien zur Korrektheit besonders positiv bewertet wurden die folgenden zwei Studien:

- *'Umweltpolitik und technische Entwicklung' (Studie Nr. 1)*: Die Evaluation überzeugt durch eine offene und ausgewogene Berichterstattung ebenso wie hinsichtlich der Einhaltung der formellen Verpflichtung sowie einem klugen Einsatz der Ressourcen. Positiv beeinflusst hat dies sicherlich auch der Umstand, dass der im Rahmen der Forschungsförderung bestehende Freiraum genutzt wurde, ohne die auch in diesem Bereich bestehenden Grenzen der Korrektheit zu überschreiten.
- *'Schneeräumung auf Urner Passstrassen' (Studie Nr. 6)*: Obwohl die Studie in äusserst knapper Zeit und mit bescheidenen Mitteln realisiert wurde, gelingt es ihr, sowohl in der Durchführung, wie auch in der Berichterstattung einen hohen Grad an Korrektheit zu erreichen. Dies wurde sicherlich dadurch gefördert, dass die Studie bei der Bearbeitung der vergleichsweise technischen Fragestellung einem klaren Konzept folgte, das in der Lage ist, die Korrektheit der Evaluation positiv zu beeinflussen.

Weniger gut bewertet wird in diesem Bereich die *'Evaluation der Zürcher Gemeinschaftszentren' (Studie Nr. 5)*. Besonders ins Gewicht fällt dabei, dass sie - im Vergleich zu den übrigen hier untersuchten Studien - am wenigsten einen öffentlichen Charakter aufweist. Weiterhin wurde diese Untersuchung in einem konfliktiven Umfeld im Auftrag eines Kontrahenten realisiert. Zudem stellt sich die Frage, ob es korrekt ist, diese Studie mit den von uns eingesetzten Kriterien zu bewerten, da es sich bei diesem

Projekt unseres Erachtens mehr um eine Organisationsberatung handelt, als um eine Evaluation.

Bei unserer Analyse stellten sich in dieser Gruppe bezüglich dreier Kriterien - 'Menschliche Interaktionen' (Kriterium 3.6.), 'Deklaration von Interessenkonflikten' (3.2.) sowie 'Finanzielle Verantwortlichkeit' (3.8.) - Probleme. Hier konnte verschiedentlich keine Bewertung vorgenommen werden. In allen drei Fällen war dafür der nicht ausreichende Informationsstand verantwortlich. Im Falle der 'Menschlichen Interaktionen' und der 'Deklaration von Interessenkonflikten' war es uns im Rahmen dieser Studie vielfach nicht möglich, eine fundierte Einschätzung abzugeben. Diese Problematik könnte dadurch bewältigt werden, indem vertiefte Gespräche mit in die Evaluation involvierten Personen geführt würden. Eine Alternative bestände darin, dass eine Meta-Evaluation nicht ex-post, sondern begleitend konzipiert würde. Dies dürfte die Informationsbeschaffung in diesem Bereich deutlich erleichtern. Das Kriterium 'Finanzielle Verantwortlichkeit' konnte in einigen Fällen deshalb nicht zufriedenstellend untersucht werden, da uns die detaillierten finanziellen Schlussabrechnungen nicht zur Verfügung standen. (Vgl. Kapitel 7.3.)

6.2.4. Genauigkeit

Die Bewertung zu den Kriterien der Genauigkeit ist, im Gegensatz zu den anderen drei Kriteriengruppen, generell nicht sehr positiv ausgefallen. Dies betrifft in besonders starker Weise die Bewertungen zum Kriterium 'Beschreibung von Zielen und Vorgehen' (Kriterium 4.3.). Hier erhält die Mehrzahl der untersuchten Studien eine negative Bewertung. Dabei ist besonders auffällig, dass die entsprechenden Schwächen vor allem bei qualitativen Untersuchungsteilen auftreten. Dies hängt einerseits sicher damit zusammen, dass eine transparente methodische Berichterstattung in diesem Bereich schwieriger zu realisieren ist, da hier, im Gegensatz zur mehrheitlich hochstrukturierten und -standardisierten quantitativen Vorgehensweise, offenere Verfahren dominieren. Wir kommen jedoch aufgrund unserer Analyse zum Schluss, dass in diesem Bereich noch deutliche Verbesserungsmöglichkeiten bestehen. Zwei weitere Kriterien erhalten eine überdurchschnittlich negative Bewertung, nämlich 'Valide Messung' bzw. 'Glaubwürdigkeit' sowie 'Übertragbarkeit' (Kriterium 4.5.) und 'Reliable Messung' bzw. 'Verlässlichkeit' (4.6.). Bei beiden Kriterien, die sich mit der Qualität der Messung befassen, spielt unseres Erachtens eine wesentliche Rolle, dass es vielfach nicht möglich ist, unter den gegebenen Bedingungen einer Evaluationsstudie (beschränkte finanzielle und zeitliche Ressourcen, hohe Erwartungen seitens der Auftraggeber und oft nicht-ausreichende Datenlage) den wissen-

schaftlichen Ansprüchen gerecht zu werden. Hier scheint unseres Erachtens ebenfalls Handlungsbedarf zu bestehen. (Vgl. Kapitel 7.2.)

Folgende drei Studien heben sich hinsichtlich der Genauigkeit in positiver Weise von den übrigen Studien ab:

- *'Lohnleichheitspostulat des Heimarbeitsgesetzes'* (Studie Nr. 3),
- *'Schneeräumung auf Urner Passstrassen'* (Studie Nr. 6) und
- *'Verbrauchsabhängige Heizkostenabrechnung - Förderung erneuerbarer Energien'* (Studie Nr. 8).

Für alle drei Studien gilt, dass sie, trotz teilweise schwieriger Rahmenbedingungen entweder darauf verzichten, problematische Analysen vorzunehmen und gewagte Folgerungen zu ziehen, oder dies klar deklarieren und auf die Grenzen der Aussagekraft hinweisen.

Im Gegensatz dazu erhalten folgende zwei Studien eine negativ abweichende Bewertung hinsichtlich der Genauigkeit:

- *'Evaluierung der EWR-Informationenkampagne'* (Studie Nr. 10): Diese Evaluation konnte aufgrund der Konstellation (enge zeitliche und finanzielle Restriktionen, analytisch verhältnismässig schwer fassbares, persuasives Evaluandum, hohe Erwartungen des Auftraggebers, breit angelegte Fragestellung und zudem Vermischung formativer und summativer Funktionen) keine positive Bewertung im Bereich der Genauigkeit erzielen. Dafür war die oben beschriebene Ausgangslage zu ungünstig. Zudem sind wir aufgrund unserer Analyse zum Schluss gekommen, dass dieses Projekt nicht mit den für Evaluationen bestimmten Kriterien zu bewerten ist, weil es eher eine Begleitforschung mit starker Beraterfunktion darstellt.
- *'Evaluation der Zürcher Gemeinschaftszentren'* (Studie Nr. 5): Auch bei diesem Projekt sind wir zum Schluss gelangt, dass es nicht statthaft ist, dieses Projekt anhand der hier eingesetzten Kriterien zu messen. Zudem waren die Rahmenbedingungen auch bei dieser Studie für die Erfüllung der Kriterien zur Genauigkeit nicht förderlich. Verschärfend kommt hinzu, dass in diesem Projekt eine Methodik eingesetzt wird, die sich nur beschränkt an sozialwissenschaftlichen Gütekriterien orientiert.

Bezüglich des Bewertungsvorgangs in der Meta-Evaluation ergaben sich vor allem bei folgenden Kriterien Schwierigkeiten:

- *'Analyse qualitativer Informationen' (Kriterium 4.9.):* Hier stellte sich das Problem, dass aufgrund der fehlenden, lückenhaften oder ungenauen Berichterstattung zur qualitativen Datenanalyse häufig unklar blieb, wie die qualitativen Informationen erhoben und ausgewertet wurden. Sicherlich würde auch in diesem Bereich die Möglichkeit bestehen im Rahmen einer Meta-Evaluation (beispielsweise mit Interviews) weitere methodische Hinweise zu erhalten. Unseres Erachtens ist es jedoch zweckmässiger, darauf zu achten, dass Evaluationsberichte auch in diesem Bereich Transparenz herstellen und damit die Nachvollziehbarkeit erleichtern.
- *'Reliable Messung' bzw. 'Verlässlichkeit' (Kriterium 4.6.), 'Objektive Berichterstattung' bzw. 'Bestätigbarkeit' (4.11.) und 'Systematische Datenüberprüfung' (4.7.):* Bei diesen Kriterien stellte sich - in abgeschwächter Weise - eine analoge Problematik, wie sie oben beschrieben wurde. Auch hier liesse sich durch eine Verbesserung der methodischen Berichterstattung die Bewertbarkeit verbessern.

Bei der Bewertung mit den Kriterien zur Genauigkeit kam es zu zwei unterschiedlichen Situationen, in denen bestimmte Kriterien nicht eingesetzt werden konnten. In einem Fall, es handelte sich um die *'Evaluation des BUWAL' (Studie Nr. 9)*, war es uns nicht möglich festzulegen, ob eher das traditionelle oder das alternative Kriterienset einzusetzen sei. Die uns vorliegenden Informationen zur paradigmatischen Ausrichtung dieser Studie liessen eine Auswahl nicht zu. So mussten wir auf eine Bewertung dieser Studie bei den drei betreffenden Kriterien (4.5., 4.6. und 4.11.) verzichten. Die zweite problematische Situation, mit der wir konfrontiert wurden, bestand darin, dass zwei Studien - *'Unternehmerische Innovationsprozesse' (Studie Nr. 2)* und *'Evaluierung der EWR-Informationenkampagne (Studie Nr. 10)* - einen rein qualitativen bzw. rein quantitativen Ansatz verfolgten. Daraus ergab sich das Problem, dass in diesen Fällen die Kriterien *'Analyse quantitativer Informationen' (Kriterium 4.8.)* bzw. *'Analyse qualitativer Informationen' (4.9.)* nicht adäquat waren. Wir haben in diesen Fällen auf eine Bewertung verzichtet. (Vgl. Kapitel 7.1.)

6.3. Vergleich aufgrund des gesamten Kriterienrasters

Aufgrund einer Häufigkeitsauszählung der Bewertungen zu den dreissig Kriterien in den Fallstudien zu den zehn Evaluationen lassen sich Aussagen über die strukturellen Stärken und Schwächen der untersuchten Evaluationsstudien gewinnen. Da es nicht statthaft wäre, die in den Kategorien enthaltenen Werte zu einem einzelnen Wert zusammenzufassen, der das Bewertungsniveau eines einzelnen Kriteriums repräsentieren würde, enthält die nachfolgende Zusammenstellung (siehe unten, Tabelle 2) die Häufigkeitsangaben zu jeder Kategorie jedes Kriteriums.

6.3.1. Kriterien mit positiver Bewertung

Unter den Kriterien mit positiver Gesamtbewertung finden sich ausschliesslich solche aus den Gruppen 'Nützlichkeit', 'Anwendbarkeit' und 'Korrektheit'. Die folgenden Kriterien sind in der empirischen Analyse am positivsten bewertet worden:

- Glaubwürdigkeit des Evaluators (Kriterium 1.2.)
- Informationsrahmen und -auswahl (Kriterium 1.3.)
- Verbreitung des Berichts (Kriterium 1.6.)
- Rechtzeitigkeit des Berichts (Kriterium 1.7.)
- Praktische Verfahren (Kriterium 2.1.)
- Formale Verpflichtung (Kriterium 3.1.)
- Herstellung von Öffentlichkeit (Kriterium 3.4.)
- Schutz individueller Menschenrechte (Kriterium 3.5.)
- Ausgewogene Berichterstattung (Kriterium 3.7.)

Bei diesen neun Kriterien bestehen zwar auch im einen oder anderen Fall negative Bewertungen. Diese sind jedoch eindeutig als Ausnahmen zu bezeichnen. Positive Bewertungen sind hier deutlich in der Mehrheit. Unter diesen Kriterien sind gewisse thematische Schwerpunkte festzulegen. Zum einen handelt es sich um Kriterien, die einen Zusammenhang zur Beziehung zwischen Auftraggeber und Evaluator aufweisen ('Rechtzeitigkeit des Berichts', 'Formale Verpflichtung'). Zweitens finden sich Kriterien, die einen Bezug zur Umsetzbarkeit und Zweckmässigkeit des Vorgehens in der Evaluation beinhalten ('Informationsrahmen und -auswahl', 'Praktische Verfahren'). Ein weiterer Schwerpunkt zeigt sich im Bereich der Berichterstattung ('Verbreitung des Berichts', 'Rechtzeitigkeit des Berichts', 'Herstellung von Öffentlichkeit', 'Ausgewogene Berichterstattung').

Tabelle 2: Übersicht über die Bewertungen entlang der dreissig Kriterien (zur Symbolik vgl. Kapitel 6.1.)

1.1. Identifikation der beteiligten Gruppen	+: 2	(+): 5	+/-: 0	(-): 2	-: 0	/: 1
1.2. Glaubwürdigkeit des Evaluators	+: 5	(+): 1	+/-: 0	(-): 0	-: 1	/: 3
1.3. Informationsrahmen und -auswahl	+: 5	(+): 3	+/-: 0	(-): 1	-: 1	/: 0
1.4. Bewertende Interpretationen	+: 2	(+): 4	+/-: 0	(-): 0	-: 3	/: 1
1.5. Klarheit des Berichts	+: 4	(+): 3	+/-: 0	(-): 0	-: 3	/: 0
1.6. Verbreitung des Berichts	+: 7	(+): 1	+/-: 1	(-): 0	-: 0	/: 1
1.7. Rechtzeitigkeit des Berichts	+: 3	(+): 4	+/-: 2	(-): 0	-: 0	/: 1
1.8. Wirkung der Evaluation	+: 1	(+): 0	+/-: 0	(-): 0	-: 1	/: 8
2.1. Praktische Verfahren	+: 5	(+): 3	+/-: 2	(-): 0	-: 0	/: 0
2.2. Politische Überlebensfähigkeit	+: 1	(+): 4	+/-: 0	(-): 0	-: 0	/: 5
2.3. Kostenwirksamkeit	+: 2	(+): 2	+/-: 2	(-): 0	-: 0	/: 4
3.1. Formale Verpflichtung	+: 3	(+): 4	+/-: 2	(-): 0	-: 0	/: 1
3.2. Deklaration von Interessenkonflikten	+: 1	(+): 3	+/-: 0	(-): 3	-: 0	/: 3
3.3. Vollständige und aufrichtige Offenlegung	+: 3	(+): 3	+/-: 2	(-): 1	-: 1	/: 0
3.4. Herstellung von Öffentlichkeit	+: 6	(+): 0	+/-: 1	(-): 0	-: 2	/: 1
3.5. Schutz individueller Menschenrechte	+: 9	(+): 0	+/-: 0	(-): 0	-: 0	/: 1
3.6. Menschliche Interaktionen	+: 2	(+): 1	+/-: 1	(-): 0	-: 0	/: 6
3.7. Ausgewogene Berichterstattung	+: 3	(+): 4	+/-: 2	(-): 1	-: 0	/: 0
3.8. Finanzielle Verantwortlichkeit	+: 5	(+): 0	+/-: 0	(-): 1	-: 1	/: 3
4.1. Identifikation des Gegenstandes	+: 2	(+): 4	+/-: 2	(-): 2	-: 0	/: 0
4.2. Kontextanalyse	+: 3	(+): 1	+/-: 1	(-): 0	-: 5	/: 0
4.3. Beschreibung von Zielen und Vorgehen	+: 0	(+): 3	+/-: 0	(-): 3	-: 4	/: 0
4.4. Verlässliche Informationsquellen	+: 0	(+): 5	+/-: 1	(-): 0	-: 4	/: 0
4.5. Valide Messung / Glaubwürdigkeit und Übertragbarkeit	+: 0	(+): 1	+/-: 1	(-): 3	-: 3	/: 2
4.6. Reliable Messung / Verlässlichkeit	+: 1	(+): 0	+/-: 2	(-): 0	-: 3	/: 4
4.7. Systematische Datenüberprüfung	+: 3	(+): 1	+/-: 1	(-): 0	-: 2	/: 3
4.8. Analyse quantitativer Informationen	+: 0	(+): 4	+/-: 1	(-): 1	-: 3	/: 1
4.9. Analyse qualitativer Informationen	+: 0	(+): 1	+/-: 2	(-): 0	-: 0	/: 7
4.10. Gerechtfertigte Folgerungen	+: 0	(+): 4	+/-: 2	(-): 1	-: 3	/: 0
4.11. Objektive Berichterstattung / Bestätigbarkeit	+: 1	(+): 3	+/-: 1	(-): 0	-: 1	/: 4

Im Gegenstandsbereich dieser neun Kriterien resultiert aus unserer Analyse ein sehr positives Gesamtbild. Obwohl die Kriterien als Idealkriterien formuliert und eingesetzt wurden und die Analyse somit recht kritisch angelegt war, sind in diesen Bereichen keine strukturellen Schwächen festzumachen. Vielmehr sind die obengenannten Aspekte als ausgesprochene Stärken der untersuchten Evaluationen zu bezeichnen (vgl. Kapitel 7.2.).

6.3.2. Kriterien mit negativer Bewertung

Alle Kriterien, die sich durch eine überdurchschnittlich negative Bewertung in den empirischen Analysen auszeichnen, stammen aus der Kriteriengruppe 'Genauigkeit'. Im einzelnen betrifft dies folgende Kriterien:

- Beschreibung von Zielen und Vorgehen (Kriterium 4.3.)
- Verlässliche Informationsquellen (Kriterium 4.4.)
- Valide Messung bzw. Glaubwürdigkeit und Übertragbarkeit (Kriterium 4.5.)
- Reliable Messung bzw. Verlässlichkeit (Kriterium 4.6.)
- Analyse quantitativer Informationen (Kriterium 4.8.)
- Gerechtfertigte Folgerungen (Kriterium 4.9.)

Bei diesen sechs Kriterien haben sich aufgrund der Fallstudien zu den zehn Evaluationsstudien massgebliche Schwächen ergeben. Neben den inhaltlich eng miteinander verbundenen Kriterien zu den wissenschaftlichen Gütekriterien (Validität bzw. Glaubwürdigkeit und Übertragbarkeit, Reliabilität bzw. Verlässlichkeit) sind auch in der methodischen Berichterstattung (Beschreibung von Zielen und Vorgehen) und bei der Begründung der Folgerungen (Gerechtfertigte Folgerungen) eindeutige Schwachpunkte zu konstatieren. Hier überwiegen negative Bewertungen. Positive Einschätzungen treten nur äusserst selten auf. Alle sechs Kriterien sind im Grunde genommen Postulate der empirischen Sozialwissenschaften. Die beiden Kriterien 'Beschreibung von Zielen und Vorgehen' und 'Gerechtfertigte Folgerungen' befassen sich primär mit der Vermittlung der empirischen Analyse. Im Bereich der methodischen Berichterstattung liegt damit ein Ansatzpunkt zur Verbesserung. Der zweite Bereich ist aufgrund der festgestellten Schwächen bei der Bewertung mit den vier anderen, obengenannten Kriterien im Bereich der Analysequalität selbst zu lokalisieren (vgl. dazu Kapitel 7.2.).

6.3.3. Kriterien mit eingeschränkter Bewertbarkeit

Aufgrund der Häufigkeitsauszählung der Bewertungen über die dreissig Kriterien und die zehn Fallstudien hinweg (siehe dazu Tabelle 2) lassen sich auch jene Kriterien eruieren,

bei denen in den empirischen Fallstudien verhältnismässig häufig auf eine Bewertung verzichtet werden musste. Dieser Vergleich stützt sich lediglich auf die Zahl der fehlenden Bewertungen ab und ignoriert somit die spezifischen Ursachen für das Ausbleiben einer Bewertung (vgl. dazu die Hinweise in Kapitel 6.1.). Besonders häufig musste bei folgenden Kriterien auf eine Bewertung verzichtet werden:

- Wirkung der Evaluation (Kriterium 1.8.)
- Politische Überlebensfähigkeit (Kriterium 2.2.)
- Menschliche Interaktionen (Kriterium 3.6.)
- Analyse qualitativer Informationen (Kriterium 4.9.)

Die ersten beiden angeführten Kriterien befassen sich im weitesten Sinne mit den Folgen einer Evaluation. Sie fordern, dass eine Evaluation angepasst auf ihren Kontext und auf die Interessen, die sich auf das Evaluandum (oder die Evaluation) beziehen, erfolgen soll, um zu erreichen, dass die Evaluation auch Wirkungen auslösen kann. Die Kriterien sollen somit vermeiden, dass irrelevante Evaluationsstudien entstehen. Die Auswirkungen einer Evaluation stellen einen zentralen Aspekt bei der Evaluation einer Evaluation dar, ebenso wie dies die Wirkungen des evaluierten Gegenstandes bei einer Evaluation tun. Eine Erfassung der Wirkungen ist damit für eine Meta-Evaluation von grosser Wichtigkeit. Aufgrund der komplexen Fragestellung - die Wirkungen sozialwissenschaftlicher Wissensbestände sind gleichzeitig ein vielbehandeltes und ein hart umstrittenes Themengebiet der (deutschen) Wissenssoziologie (vgl. dazu Hitzler/Honer/Maeder 1994, Dewe 1991, Stehr 1991, Nowotny 1990, Wingers 1988 und 1990, Beck/Bonss 1984 und 1989) und der 'utilization research' im englischsprachigen Raum (siehe Albaek 1995, Straton 1995, Stevens/Dial 1994a, Tomlinson et al. 1994, Pröpper 1993, Boggs 1992, Lampinen 1992, Lester 1992, Lipton 1992, Leeuw 1991, MacRae 1991, Stenzel 1991, Huberman/Cox 1990, Lester/Wilds 1990, McLaughlin et al. 1988, Leviton/Hughes 1981 und den 'Klassiker' von Carol H. Weiss 1977) -, aber auch wegen der hier gewählten Untersuchungsanlage, konnte diese Thematik im Rahmen dieser Studie nicht ausreichend gründlich untersucht werden (vgl. dazu Kapitel 7.3.).

Bei den beiden anderen Kriterien stellen sich weniger analytische Probleme, als vielmehr der mangelnde Informationsstand einer Bewertung entgegen. Beim Kriterium 'Menschliche Interaktionen' wäre mit zusätzlichen Abklärungen durchaus eine Bewertung möglich. Der Aufwand dafür hätte jedoch die Möglichkeiten dieses Projektes überschritten. Als Alternative bestünde die Möglichkeit, durch eine begleitende Meta-Evaluation den Datenzugang zu verbessern (vgl. Kapitel 7.3.). Dabei müsste jedoch gleichzeitig eine Reduktion der Bewertungsmöglichkeiten im Wirkungsbereich in Kauf genommen werden (siehe oben). Auch beim anderen Kriterium, das sich mit der 'Analyse qualitativer

Informationen' befasst, würden sich die Bewertungsmöglichkeiten durch ergänzende Abklärungen sicher verbessern lassen. Unseres Erachtens scheint es uns hier jedoch angemessener zu sein, durch eine Qualitätssteigerung bei der methodischen Berichterstattung eine Verbesserung der Bewertungsmöglichkeiten zu erreichen. Damit könnte nämlich gleichzeitig auch eines der diagnostizierten Problemfelder (vgl. Kapitel 6.3.2.) einer Verbesserung zugeführt werden (vgl. dazu auch Kapitel 7.2.).

6.4. Interdependenzen der Kriterien

Bei der Erarbeitung der empirischen Fallstudien zu den zehn Evaluationsstudien (siehe Kapitel 5.) haben wir darauf geachtet, dass Beziehungen zwischen verschiedenen Kriterien im Text explizit ausgewiesen werden. Aufgrund der in den Fallstudientexten verfügbaren Verweise zwischen den Kriterien lässt sich ein empirisches Beziehungsnetzwerk der Kriterien darstellen. Die nachfolgende Analyse dieses Beziehungsnetzwerkes erfolgt aus verschiedenen Gründen:

- Die Analyse der Interdependenzen soll dazu beitragen, die Struktur des verwendeten Kriterienrasters zu beschreiben.
- Die Netzwerkanalyse soll Rückschlüsse auf die Frage erlauben, inwiefern sich der Kriterienraster als Bewertungsinstrument in Meta-Evaluationen eignet.
- Die Untersuchung der Beziehungen der Einzelkriterien und der Kriteriengruppen soll Aufschlüsse darüber ermöglichen, ob die Struktur des Kriterienrasters (also die Einteilung in die vier Gruppen) verändert werden sollte.
- Die Netzwerkanalyse soll weiter dazu beitragen, Hinweise auf thematische Überlappungen zu finden, die eine Änderung (oder eine Eliminierung) eines Einzelkriteriums nahelegen würden.

Die gewählte Vorgehensweise ist aus methodischer Perspektive nicht unproblematisch, handelt es sich doch dabei im Prinzip um eine Inhaltsanalyse selbst verfasster Texte. Dieses Material verfügt nicht über die Authentizität, wie sie sonst in der empirischen Inhaltsanalyse üblich ist - normalerweise werden ja dort Konstruktionen erster Ordnung analysiert. Trotz diesem Vorbehalt erhoffen wir uns aber Hinweise, die zur Klärung der vier oben angeführten Punkte beitragen können.

Zur Analyse kann, wenn die Beziehungen zwischen zwei Einzelkriterien betrachtet werden, auf die *Anzahl der Beziehungen der Diade* (des Kriterienpaars) abgestellt werden. Werden jedoch Beziehungen zwischen einem Einzelkriterium und einer Kriteriengruppe bzw. solche zwischen zwei Kriteriengruppen betrachtet, kann die absolute Zahl der Verbindungen nicht als Indikator für die Beziehungsstärke eingesetzt werden, da die Kriteriengruppen eine unterschiedliche Zahl von Einzelkriterien umfassen. Aus diesem Grund wird in diesen Fällen mit der von diesen Beeinflussungen befreiten *Beziehungsdichte* gearbeitet.

Die nachfolgende Datenmatrix (Tabelle 3) zeigt die Anzahl der Beziehungen unter den einzelnen Kriterien. Da die Richtung der Beziehungen dabei nicht berücksichtigt wurde,

entsteht eine symmetrische Datenmatrix. Die Zahlen in der ersten Zeile bzw. in der ersten Spalte stellen die Nummern der Einzelkriterien dar. Die nachfolgenden Analysen werden auf diesem Material basieren.

Tabelle 3: Interdependenzen der Kriterien

	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	2.1.	2.2.	2.3.	3.1.	3.2.	3.3.	3.4.	3.5.	3.6.	3.7.	3.8.	4.1.	4.2.	4.3.	4.4.	4.5.	4.6.	4.7.	4.8.	4.9.	4.10.	4.11.				
1.1.	X																																	
1.2.		X						1		2																								
1.3.			X						2																									
1.4.				X	5																													
1.5.					X	2		1		2																								
1.6.						X	3	3																										
1.7.							X	5	1			4																						
1.8.								X		8	3		2																					
2.1.									X			4	1																					
2.2.										X		1	2																					
2.3.											X																							
3.1.												X	1																					
3.2.													X																					
3.3.														X																				
3.4.															X																			
3.5.																X																		
3.6.																	X																	
3.7.																		X																
3.8.																			X															
4.1.																				X														
4.2.																					X													
4.3.																						X												
4.4.																							X											
4.5.																								X										
4.6.																									X									
4.7.																										X								
4.8.																											X							
4.9.																												X						
4.10.																													X					
4.11.																														X				

6.4.1. Beziehungen der Einzelkriterien

Die Vernetzung der insgesamt dreissig Einzelkriterien ist aufgrund der empirischen Analyse stark unterschiedlich ausgefallen. Eine tabellarische Übersicht (Tabelle 4) zeigt die Einzelkriterien in abnehmender Beziehungsstärke. In den ersten zwei Spalten der Tabelle sind die Nummer und die Bezeichnung des Kriteriums sowie der Rang, den das Kriterium hinsichtlich seiner Beziehungsstärke einnimmt, aufgeführt. In der dritten Spalte wird die Beziehungsstärke des Kriteriums zu den übrigen 29 Einzelkriterien angeführt. Neben einer nominalen Beschreibung der Beziehungsstärke enthält die Tabelle auch die Zahl der beobachteten Beziehungen. Die vierte Spalte führt jene Einzelkriterien auf, zu denen das spezifische Kriterium gemäss der empirischen Analyse häufig Beziehungen hat. Die Klammer enthält die Zahl der Bezugspunkte einer Diade. In der fünften Spalte sind die Kriteriengruppen angeführt, zu welchen seitens eines bestimmten Kriteriums eine starke Beziehung besteht. In der Klammer steht die zugehörige Beziehungsdichte. Diese entspricht dem Quotienten aus der Anzahl der empirisch erfassten Bezugspunkte des Kriteriums mit einer Kriteriengruppe und der Zahl der Kriterien in dieser Gruppe. Diese Berechnung ist erforderlich, da die absolute Anzahl der Beziehungen von der Zahl der Kriterien in einer Gruppe beeinflusst wird. Durch die Quotientenbildung kann diese Verzerrung korrigiert werden. Diese Vorgehensweise ist aus der Netzwerkanalyse bekannt (Scott 1991: 72-79). In der sechsten und damit letzten Spalte finden sich sodann Angaben, die über das Verhältnis des Kriteriums zur Kriteriengruppe, zu der es selbst gehört, Aufschlüsse ermöglichen. Neben einer nominalen Umschreibung haben wir wiederum die Beziehungsdichte angeführt.

Tabelle 4a: Übersicht über die Kriterien und deren Beziehungen aufgrund der empirischen Analyse - nach Beziehungsstärke (Teil 1)

Nr.	Kriterium [Rang]	Beziehungsstärke [Anzahl Bezüge]	Starke Beziehungen zu Einzelkriterien [Anzahl Bezüge]	Beziehungen zu Gruppen [Dichte der Beziehungen]	Beziehungen zur eigenen Gruppe [Dichte der Beziehung]
4.5.	Valide Messung [1]	ausgesprochen stark [80]	<ul style="list-style-type: none"> 4.6. Reliable Messung [11] 4.8. Analyse quantitativer Informationen [8] 4.3. Beschreibung von Zielen und Vorgehen [7] 1.3. Informationsrahmen und -auswahl [6] 2.1. Praktische Verfahren [5] 4.11. Objektive Berichterstattung [5] 2.3. Kostenwirksamkeit [4] 4.4. Verlässliche Informationsquellen [4] 4.7. Systematische Datenüberprüfung [4] 	<ul style="list-style-type: none"> 4. Genauigkeit [4,9] 2. Anwendbarkeit [3,0] 1. Nützlichkeit [1,9] 	sehr stark [4,9]
4.10.	Gerechtfertigte Folgerungen [2]	sehr stark [63]	<ul style="list-style-type: none"> 1.4. Bewertende Interpretationen [8] 4.8. Analyse quantitativer Informationen [8] 3.3. Vollständige & aufrichtige Offenlegung [5] 1.8. Wirkung der Evaluation [4] 2.3. Kostenwirksamkeit [4] 3.1. Formale Verpflichtung [4] 3.7. Ausgewogene Berichterstattung [4] 	<ul style="list-style-type: none"> 2. Anwendbarkeit [2,7] 4. Genauigkeit [2,4] 3. Korrektheit [2,0] 1. Nützlichkeit [1,9] 	stark [2,4]
4.3.	Beschreibung von Zielen und Vorgehen [3]	sehr stark [62]	<ul style="list-style-type: none"> 1.4. Bewertende Interpretation [13] 4.4. Verlässliche Informationsquellen [7] 4.5. Valide Messung [7] 1.5. Klarheit des Berichts [4] 4.6. Reliable Messung [4] 4.8. Analyse quantitativer Informationen [4] 	<ul style="list-style-type: none"> 4. Genauigkeit [3,0] 1. Nützlichkeit [2,8] 2. Anwendbarkeit [1,0] 	stark [3,0]
4.8.	Analyse quantitativer Informationen [4]	sehr stark [58]	<ul style="list-style-type: none"> 4.5. Valide Messung [8] 4.10. Gerechtfertigte Folgerungen [8] 2.1. Praktische Verfahren [7] 4.6. Reliable Messung [5] 1.3. Informationsrahmen und -auswahl [4] 1.4. Bewertende Interpretationen [4] 1.5. Klarheit des Berichts [4] 4.3. Beschreibung von Zielen und Vorgehen [4] 	<ul style="list-style-type: none"> 4. Genauigkeit [3,2] 2. Anwendbarkeit [3,0] 1. Nützlichkeit [1,6] 	stark [3,2]

Tabelle 4b: Übersicht über die Kriterien und deren Beziehungen aufgrund der empirischen Analyse - nach Beziehungsstärke (Teil 2)

Nr.	Kriterium [Rang]	Beziehungsstärke [Anzahl Bezüge]	Starke Beziehungen zu Einzelkriterien [Anzahl Bezüge]	Beziehungen zu Gruppen [Dichte der Beziehungen]	Beziehungen zur eigenen Gruppe [Dichte der Beziehung]
1.3.	Informationsrahmen und -auswahl [5/1]	stark [46]	1.1. Identifikation der beteiligten Gruppen [7] 4.5. Valide Messung [6] 4.4. Verlässliche Informationsquellen [5] 4.1. Identifikation des Gegenstandes [4] 4.8. Analyse quantitativer Informationen [4]	4. Genauigkeit [2,7] 1. Nützlichkeit [1,4] 2. Anwendbarkeit [1,3]	mittel [1,4]
4.6.	Reliable Messung [5/2]	stark [46]	4.5. Valide Messung [11] 4.8. Analyse quantitativer Informationen [5] 2.1. Praktische Verfahren [4] 4.3. Beschreibung von Zielen und Vorgehen [4]	4. Genauigkeit [2,8] 2. Anwendbarkeit [2,3]	stark [2,8]
1.4.	Bewertende Interpretationen [7]	stark [42]	4.3. Beschreibung von Zielen und Vorgehen [13] 4.10. Gerechtfertigte Folgerungen [8] 1.5. Klarheit des Berichts [5] 4.8. Analyse quantitativer Informationen [4]	4. Genauigkeit [2,9] 1. Nützlichkeit [1,0]	eher schwach [1,0]
4.4.	Verlässliche Informationsquellen [8]	stark [39]	4.3. Beschreibung von Zielen und Vorgehen [7] 1.3. Informationsrahmen und -auswahl [5] 4.5. Valide Messung [4]	4. Genauigkeit [2,0] 1. Nützlichkeit [1,1] 2. Anwendbarkeit [1,0]	stark [2,0]
1.5.	Klarheit des Berichts [9]	recht stark [36]	1.4. Bewertende Interpretationen [5] 4.3. Beschreibung von Zielen und Vorgehen [4] 3.4. Herstellung von Öffentlichkeit [4] 4.8. Analyse quantitativer Informationen [4]	4. Genauigkeit [1,7] 1. Nützlichkeit [1,4]	mittel [1,4]
1.8.	Wirkung der Evaluation [10]	recht stark [34]	1.7. Rechtzeitigkeit des Berichts [5] 2.2. Politische Überlebensfähigkeit [8] 4.10. Gerechtfertigte Folgerungen [4]	2. Anwendbarkeit [3,7] 1. Nützlichkeit [1,4]	mittel [1,4]
2.1.	Praktische Verfahren [11]	recht stark [33]	4.8. Analyse quantitativer Informationen [7] 4.5. Valide Messung [5] 3.1. Formale Verpflichtung [4] 4.6. Reliable Messung [4]	4. Genauigkeit [2,2]	schwach [0,5]

Tabelle 4c: Übersicht über die Kriterien und deren Beziehungen aufgrund der empirischen Analyse - nach Beziehungsstärke (Teil 3)

Nr.	Kriterium [Rang]	Beziehungsstärke [Anzahl Bezüge]	Starke Beziehungen zu Einzelkriterien [Anzahl Bezüge]	Beziehungen zu Gruppen [Dichte der Beziehungen]	Beziehungen zur eigenen Gruppe [Dichte der Beziehung]
1.1.	Identifikation der beteiligten Gruppen [12/1]	recht stark [32]	1.3. Informationsrahmen und -auswahl [7]	1. Nützlichkeit [1,4] 3. Korrektheit [1,1] 4. Genauigkeit [1,0]	mittel [1,4]
2.3.	Kostenwirksamkeit [12/2]	recht stark [32]	4.5. Valide Messung [4] 4.10. Gerechtfertigte Folgerungen [4]	4. Genauigkeit [2,5]	schwach [0,5]
4.1.	Identifikation des Gegenstandes [14]	recht stark [31]	1.3. Informationsrahmen und -auswahl [4] 4.2. Kontextanalyse [4]	4. Genauigkeit [1,4] 1. Nützlichkeit [1,0] 2. Anwendbarkeit [1,0]	mittel [1,4]
3.7.	Ausgewogene Berichterstattung [15/1]	mittel [28]	4.10. Gerechtfertigte Folgerungen [4]	3. Korrektheit [1,3] 4. Genauigkeit [1,1]	eher schwach [0,9]
4.9.	Analyse qualitativer Informationen [15/2]	mittel [28]	-	2. Anwendbarkeit [1,7] 4. Genauigkeit [1,3]	mittel [1,3]
4.11.	Objektive Berichterstattung [15/3]	mittel [28]	4.5. Valide Messung [5]	1. Nützlichkeit [1,3] 3. Korrektheit [1,0]	eher schwach [0,8]
4.2.	Kontextanalyse [18]	mittel [27]	4.1. Identifikation des Gegenstandes [4]	1. Nützlichkeit [1,1]	eher schwach [0,9]
3.1.	Formale Verpflichtung [19]	mittel [26]	1.7. Rechtzeitigkeit des Berichts [4] 2.1. Praktische Verfahren [4] 4.10. Gerechtfertigte Folgerungen [4]	2. Anwendbarkeit [1,7] 1. Nützlichkeit [1,0]	schwach [0,6]
3.3.	Vollständige und auf-richtige Offenlegung [20]	mittel [25]	4.10. Gerechtfertigte Folgerungen [5]	4. Genauigkeit [1,9]	sehr schwach [0,1]
4.7.	Systematische Datenüberprüfung [21]	mittel [24]	4.5. Valide Messung [4]	4. Genauigkeit [1,5] 2. Anwendbarkeit [1,0]	mittel [1,5]
2.2.	Polit. Überlebensf. [22]	mittel [23]	1.8. Wirkung der Evaluation [8]	1. Nützlichkeit [1,9]	keine [0,0]

Tabelle 4d: Übersicht über die Kriterien und deren Beziehungen aufgrund der empirischen Analyse - nach Beziehungsstärke (Teil 4)

Nr.	Kriterium [Rang]	Beziehungsstärke [Anzahl Bezüge]	Starke Beziehungen zu Einzelkriterien [Anzahl Bezüge]	Beziehungen zu Gruppen [Dichte der Beziehungen]	Beziehungen zur eigenen Gruppe [Dichte der Beziehung]
3.2.	Deklaration von Interessenkonflikten [23]	eher schwach [20]	-	2. Anwendbarkeit [1,0]	eher schwach [0,9]
3.8.	Finanz. Verantwortl. [24]	eher schwach [19]	-	4. Genauigkeit [1,5]	keine [0,0]
1.2.	Glaubwürdigkeit des Evaluators [25]	eher schwach [17]	-	4. Genauigkeit [1,1]	schwach [0,4]
1.7.	Rechtzeitigkeit des Berichts [26]	eher schwach [14]	1.8. Wirkung der Evaluation [5] 3.1. Formale Verpflichtung [4]	1. Nützlichkeit [1,1]	eher schwach [1,1]
1.6.	Verbr. des Berichts [27/1]	schwach [10]	-	1. Nützlichkeit [1,1]	eher schwach [1,1]
3.4.	Herstellung von Öffentlichkeit [27/2]	schwach [10]	1.5. Klarheit des Berichts [4]	keine starken Beziehungen	sehr schwach [0,1]
3.6.	Menschliche Interaktionen [29]	schwach [8]	-	keine starken Beziehungen	schwach [0,6]
3.5.	Schutz individueller Menschenrechte [30]	sehr schwach [1]	-	keine starken Beziehungen	sehr schwach [0,1]

Wie aus der Tabelle hervorgeht, sind die Differenzen zwischen den am stärksten und den am schwächsten verknüpften Kriterien erheblich. Am stärksten mit anderen Kriterien verflochten ist dabei eindeutig das Kriterium 'Valide Messung' (Kriterium 4.5.) mit insgesamt 80 Beziehungen. Aber auch die nächsten drei - 'Gerechtfertigte Folgerungen' (Kriterium 4.10.; 63 Beziehungen), 'Beschreibung von Zielen und Vorgehen' (4.3.; 62) und 'Analyse quantitativer Informationen' (4.8.; 58) - weisen eine hohe Anzahl Beziehungen auf. Im Gegensatz dazu verfügt das Kriterium 'Schutz individueller Menschenrechte' (3.5.; 1) kaum über Bezugspunkte zu anderen Kriterien. Über alle zehn untersuchten Evaluationsstudien hinweg zeigte sich hier gerade einmal ein Bezugspunkt zu einem anderen Kriterium. Aber auch die Kriterien 'Menschliche Interaktionen' (3.6.; 8), 'Verbreitung des Berichts' (1.6.; 10) und 'Herstellung von Öffentlichkeit' (3.4.; 10) weisen nur sehr selten Beziehungspunkte zu anderen Kriterien auf.

Betrachten wir die Zugehörigkeit zu den vier Kriteriengruppen, fällt auf, dass die Kriterien der Gruppe 'Genauigkeit' (4.) sehr häufig Beziehungen zu anderen Einzelkriterien aufweisen, stammen doch fünf der sechs am stärksten vernetzten Kriterien aus dieser Gruppe. Durchschnittlich sind in der empirischen Analyse der zehn Fallstudien in dieser Gruppe pro Kriterium rund 44 Beziehungen festgestellt worden. Währenddem sich die Kriterien der Gruppen 'Nützlichkeit' (1.) und 'Anwendbarkeit' (2.) mit durchschnittlich rund 29 Beziehungen pro Kriterium bezüglich ihrer Beziehungsstärke etwa durchschnittlich verhalten, weisen dagegen die Kriterien zur 'Korrektheit' nur wenige Beziehungen auf. Im Durchschnitt sind wir bei den Kriterien dieser Gruppe auf nur rund 18 Beziehungen gestossen.

Aufgrund der Angaben in der vierten Spalte der Tabelle 4 lässt sich erkennen, welche Diaden (Kriterienpaare) besonders stark miteinander verknüpft sind (vgl. dazu auch Tabelle 3):

- Deutlich am engsten sind die Einzelkriterien '*Bewertende Interpretationen*' (Kriterium 1.4.) und '*Beschreibung von Zielen und Vorgehen*' (Kriterium 4.3.) mit insgesamt dreizehn Beziehungen miteinander verbunden. Aufgrund der Formulierungen zu den beiden Kriterien wird ersichtlich, dass dies damit in Zusammenhang steht, dass die beiden Kriterien partiell denselben Bereich abdecken. So sollen gemäss dem Kriterium 'Bewertende Interpretationen' die verwendeten Perspektiven, Verfahren und Begründungen sorgfältig beschrieben werden, währenddem beim Kriterium 'Beschreibung von Zielen und Vorgehen' postuliert wird, dass die Ziele und das Vorgehen ausreichend genau beobachtet und beschrieben werden sollten (vgl. dazu Kapitel 7.1.).

- Auch eine sehr hohe Zahl von Beziehungen (11) weisen die beiden Kriterien '*Valide Messung*' (4.5.) und '*Reliable Messung*' (4.6.) auf. Die Gründe für diesen Zusammenhang liegen in der inhaltlichen Relation zwischen den beiden Konzepten Validität und Reliabilität.
- Zwischen den Kriterien '*Bewertende Interpretation*' (1.4.) und '*Gerechtfertigte Folgerungen*' (4.10.) sind insgesamt acht Bezüge beobachtet worden. Die Ursache liegt wohl vor allem darin, dass sich in den Gegenstandsbereichen der beiden Kriterien eine Überlappung ergibt, behandeln doch beide Kriterien die Begründungen zu den Ergebnissen bzw. zu den Folgerungen (vgl. dazu Kapitel 7.1.).
- Acht Bezüge bestehen auch zwischen den Kriterien '*Wirkung der Evaluation*' (1.8.) und '*Politische Überlebensfähigkeit*' (2.2.). Beide befassen sich mit den Konsequenzen einer Evaluation und die beiden Kriterien haben eine enge inhaltliche Interdependenz.
- Die Kriterien '*Valide Messung*' (4.5.) und '*Analyse quantitativer Informationen*' (4.8.) verfügen ebenso über acht Beziehungen. Eine angemessene und systematische Analyse, wie sie im Kriterium 4.8. gefordert wird, bildet einen wichtigen Aspekt bei der Etablierung einer hohen Validität einer Untersuchung. Die Gültigkeit der erhaltenen Informationen für den gegebenen Zweck (siehe Kriterium 4.5.) wiederum, hat einen wesentlichen Anteil bei der Sicherstellung haltbarer Informationen (siehe dazu Kriterium 4.8.). Bemerkenswert ist in diesem Zusammenhang der Aspekt, dass im Gegensatz zu den quantitativen Analysen kaum Bezüge zum Kriterium, das sich mit der '*Analyse qualitativer Informationen*' (4.9.) befasst, festzustellen sind. Auch hier würde ja das oben beschriebene Verhältnis in analoger Weise zutreffen. Eine der Ursachen für diese Diskrepanz sehen wir darin, dass im Bereich der qualitativen Analysen aufgrund der deutlich schwereren Nachvollziehbarkeit häufig keine ausführliche Würdigung möglich war (vgl. dazu Kapitel 6.1.4.9. und 6.2.4.). Dies hatte zur Folge, dass eventuell bestehende Beziehungen in unserer Analyse zu wenig eruiert werden konnten.
- Ebenfalls acht Beziehungen wurden in den empirischen Fallstudien zwischen den Kriterien '*Analyse quantitativer Informationen*' (4.8.) und '*Gerechtfertigte Folgerungen*' (4.10.) festgestellt. Unseres Erachtens sind diese Beziehungspunkte vor allem darauf zurückzuführen, dass das Kriterium '*Analyse quantitativer Informationen*' als vorgelagert zum Kriterium '*Gerechtfertigte Folgerungen*' gesehen werden muss. Die Begründungen zu Folgerungen stützen sich unter anderem auf eine quantitative Analyse ab. Deshalb finden sich bei der Analyse verschiedentlich Querverweise, die auf positive oder negative Aspekte hinweisen. Wird eine fundierte quantitative Analyse durchgeführt, besteht eine stabile Basis zur Formulierung der Folgerungen. Bestehen

jedoch in der quantitativen Analyse Mängel (wie dies beispielsweise beim Einsatz unzweckmässiger statistischer Masszahlen der Fall ist), wirkt sich dies zumeist auch negativ auf die Begründung der Folgerungen aus. Auch hier fällt auf, dass im Gegensatz zum Kriterium 'Analyse quantitativer Informationen' bei jenem zur 'Analyse qualitativer Informationen' (4.9.) keine Häufung von Beziehungen feststellbar ist (vgl. dazu die Hinweise im vorangehenden Abschnitt).

- Ein Paar, das über sieben Beziehungen verfügt, besteht aus den Kriterien '*Identifikation der beteiligten Gruppen*' (1.1.) und '*Informationsrahmen und -auswahl*' (1.3.). Diese Beziehungen entstehen dadurch, dass die Informationsbeschaffung und die Ermittlung der relevanten Interessen untrennbar oder zumindest sehr stark miteinander verbunden sind.
- Sieben Beziehungen sind auch zwischen den Kriterien '*Praktische Verfahren*' (2.1.) und '*Analyse quantitativer Informationen*' (4.8.) festzustellen. Dieser Zusammenhang lässt sich dadurch erklären, dass sich Störungen bei der Durchführung der quantitativen Analyse (mangelhafte Datenlage, ungenügende Datenqualität usw.) bei beiden Kriterien in den Bewertungen bemerkbar machen.
- Über ebenfalls sieben Beziehungen verfügt das Kriterienpaar '*Beschreibung von Zielen und Vorgehen*' (4.3.) und '*Verlässliche Informationsquellen*' (4.4.). Die ausreichend genaue Beschreibung des Vorgehens sowie die detaillierte Beschreibung der Informationsquellen, wie sie in den beiden Kriterien gefordert werden, betreffen eine sehr ähnliche Thematik. Aufgrund der inhaltlichen Verwandtschaft der beiden Kriterien kommt es in den Fallstudien häufig zu Querverweisen.
- Insgesamt sieben Interdependenzen sind in den Fallstudien auch zwischen den Kriterien '*Beschreibung von Zielen und Vorgehen*' (4.3.) und '*Valide Messung*' (4.5.) zu konstatieren. Diese Beziehungen sind vor allem darauf zurückzuführen, dass in der empirischen Bewertung der Evaluationsstudien verschiedentlich das Problem bestand, dass aufgrund einer im Detail nicht nachvollziehbaren Beschreibung der Vorgehensweise die Bewertung der Validität nicht ausreichend geklärt werden konnte.

Gehen wir aufgrund der fünften Spalte in Tabelle 4 der Frage nach, wie sich die Beziehungen der Einzelkriterien zu den vier Kriteriengruppen verhalten, lassen sich folgende Aussagen machen:

- Mehrheitlich - in insgesamt 15 Fällen - bestehen die stärksten Beziehungen zur Kriteriengruppe 'Genauigkeit' (4.). Sechsmal treten die stärksten Beziehungen zur Gruppe 'Nützlichkeit' (1.) auf. Bei fünf Kriterien bestehen die stärksten Beziehungen zur

Gruppe 'Anwendbarkeit' (2.). Die Gruppe der Kriterien zur 'Korrektheit' (3.) tritt dagegen nur einmal als stärkster Beziehungspartner in Erscheinung.

- Das Einzelkriterium 'Valide Messung' (4.5.) weist eine äusserst starke Beziehung zur Kriteriengruppe 'Genauigkeit' (4.) auf (Beziehungsdichte 4,9). Sehr stark ist auch der Bezug des Kriteriums 'Wirkung der Evaluation' (1.8.) zur Gruppe 'Anwendbarkeit' (2.; Dichte 3,7). Ebenfalls stark ausgeprägt sind die Bezüge des Kriteriums 'Analyse quantitativer Informationen' (4.8.) zu den Gruppen 'Genauigkeit' (4.; Dichte 3,2) und 'Anwendbarkeit' (2.; Dichte 3,0). Über eine Dichte von 3,0 verfügen die Beziehungen der Kriterien 'Valide Messung' (4.5.) zur Gruppe 'Anwendbarkeit' (2.) und 'Beschreibung von Zielen und Vorgehen' (4.3.) zur Gruppe 'Genauigkeit' (4.). Recht stark sind auch die Bezüge des Kriteriums 'Bewertende Informationen' (1.4.) zur Gruppe 'Genauigkeit' (4.; Dichte 2,9).

Die sechste Spalte der Tabelle 4 informiert darüber, wie stark das einzelne Kriterium in die eigene Kriteriengruppe integriert ist. Auch hier sind wesentliche Unterschiede festzustellen. Besonders stark fällt die Bindung zur eigenen Gruppe bei den Kriterien 'Valide Messung' (4.5.; Dichte 4,9), 'Analyse quantitativer Informationen' (4.8.; Dichte 3,2), 'Beschreibung von Zielen und Vorgehen' (4.3.; Dichte 3,0) und 'Reliable Messung' (4.6.; Dichte 2,8) aus. Im Gegensatz dazu weisen einzelne Kriterien zu den übrigen ihrer Gruppe keine oder nur sehr schwache Beziehungspunkte auf. Dies sind im Einzelnen die folgenden Kriterien: 'Politische Überlebensfähigkeit' (2.2.; Dichte 0,0), 'Finanzielle Verantwortlichkeit' (3.8., Dichte 0,0), 'Vollständige und aufrichtige Offenlegung' (3.3.; Dichte 0,1), 'Herstellung von Öffentlichkeit' (3.4.; Dichte 0,1) und 'Schutz individueller Menschenrechte' (3.5.; Dichte 0,1). Neben dem Aspekt, dass wiederum die der Gruppe 'Genauigkeit' (4.) angehörenden Kriterien starke und jene der Gruppe 'Korrektheit' (3.) schwache Bindungen zur eigenen Gruppe aufweisen, fällt auf, dass verschiedene der nur schwach in die eigene Gruppe integrierten Kriterien, stärkere Beziehungen zu anderen Kriteriengruppen aufweisen:

- Das Kriterium '*Politische Überlebensfähigkeit*' (2.2.) weist keinerlei Beziehungen zu den übrigen Kriterien der eigenen Gruppe auf (Dichte 0,0) hat aber einen recht starken Zusammenhang zu Kriterien der Gruppe 'Nützlichkeit' (1.; Dichte 1,9). Dies lässt sich vor allem auf häufige Beziehungen zu einem Einzelkriterium aus der Nützlichkeits-Gruppe zurückführen (vgl. auch oben, Tabelle 3), nämlich zum Kriterium 'Wirkung der Evaluation' (1.8.) Die zwei Kriterien weisen beide einen starken 'stakeholder'-Bezug auf. Die politische Überlebensfähigkeit und der Wirkungsaspekt beziehen sich zudem beide auf die Folgen einer Evaluation und sind inhaltlich eng miteinander verbunden (vgl. dazu oben).

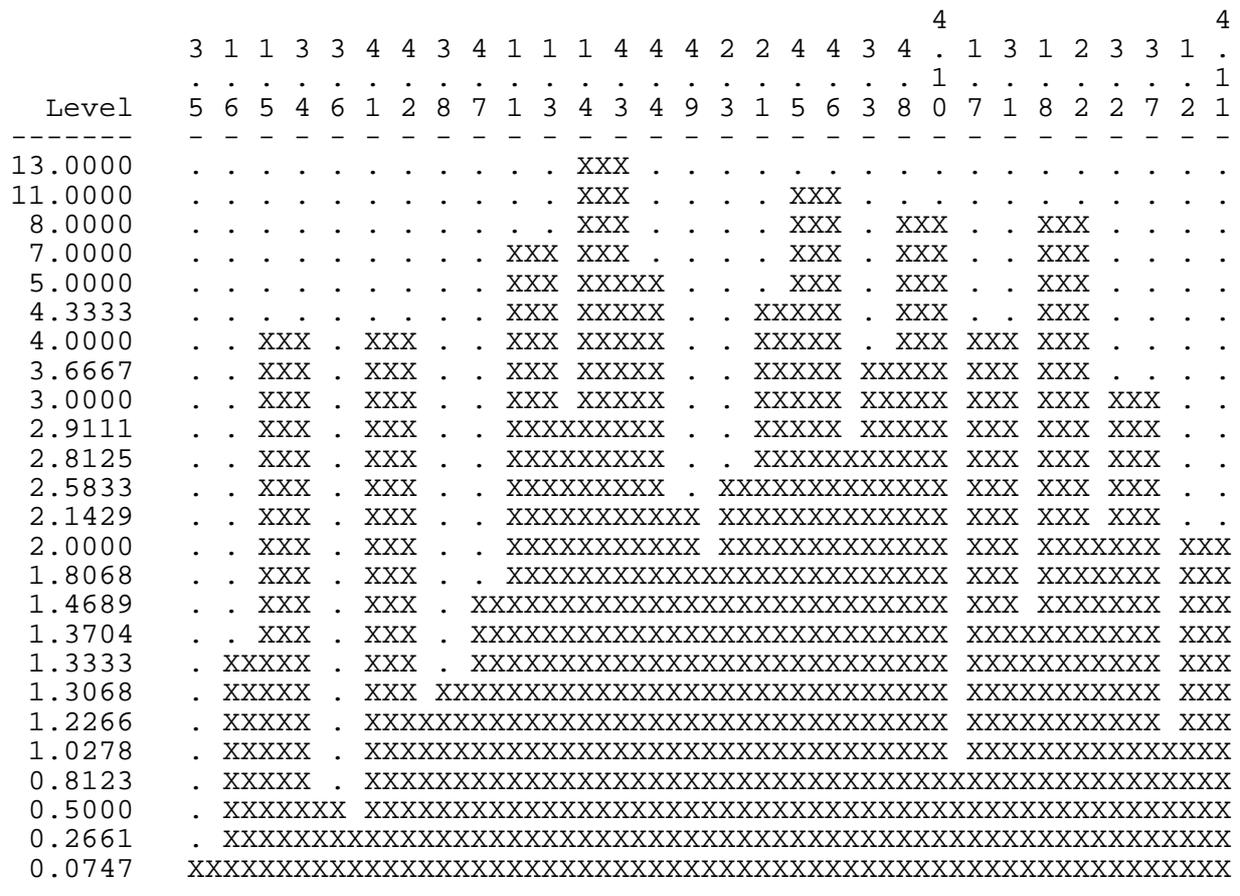
- Das Kriterium '*Vollständige und aufrichtige Offenlegung*' (3.3.) hat zur eigenen Kriteriengruppe nur einen sehr schwachen Bezug (Dichte 0,1), verfügt aber zur Gruppe 'Genauigkeit' (4.) über einen recht starken Zusammenhang (Dichte 1,9). Am stärksten tritt dabei das Kriterium 'Gerechtfertigte Folgerungen' (4.10.) hervor, wobei auch zu den übrigen Kriterien der Gruppe Beziehungen bestehen. Der starke Zusammenhang zu einem gruppenfremden Kriterium kann inhaltlich damit begründet werden, dass das Kriterium zur vollständigen und aufrichtigen Offenlegung eine offene, direkte und ehrliche Darlegung der sachdienlichen Ergebnisse und der Grenzen der Evaluation postuliert, was in direktem Zusammenhang zum Gegenstandsbereich der gerechtfertigten Folgerungen steht.
- Das Kriterium '*Formale Verpflichtung*' (Kriterium 3.1.) weist ebenfalls eine nur schwache Verbindung zur eigenen Kriteriengruppe auf (Dichte 0,6), steht aber zur Gruppe 'Anwendbarkeit' (2.) in einer recht starken Beziehung (Dichte 1,7). Letztere ist primär darauf zurückzuführen, dass recht häufig Bezüge zum Kriterium 'Praktische Verfahren' (Kriterium 2.1.) festgestellt wurden. Die Praktikabilität der eingesetzten Methodik erscheint somit als wichtiger Bestimmungsfaktor zur Einhaltung der formalen Verpflichtungen.

6.4.2. Netzwerkanalyse zu den Interdependenzen der Kriterien

Zur vertiefenden Analyse der zur Verfügung stehenden Angaben zum Verhältnis der Einzelkriterien untereinander, haben wir eine Netzwerkanalyse vorgenommen, die zur Veranschaulichung auch eine hierarchische Clusteranalyse sowie eine Multidimensionale Skalierung umfasst. Als Eingangsdaten für die Analysen diente jeweils die in Tabelle 3 (siehe oben) wiedergegebene symmetrische Datenmatrix, welche die absolute Anzahl der Beziehungen zwischen den Einzelkriterien enthält. Zur Durchführung der Analysen setzen wir die Software UCINET IV (Version 1.0) ein (Borgatti/Everett/Freeman 1992). Dies ist ein Computerprogramm, das speziell zur Implementation von Netzwerkanalysen konzipiert wurde. (Zur Netzwerkanalyse vgl. Wasserman/Faust 1994, Wasserman/Galaskiewicz 1994, Scott 1991, Pappi 1987; zur Clusteranalyse vgl. Backhaus/Erichson/Plinke/Weiber 1994: 260-321, Aldenderfer/Blashfield 1984, Lorr 1983; zur Multidimensionalen Skalierung vgl. Backhaus/Erichson/Plinke/Weiber 1994: 433-497, Coxon 1982, Kruskal/Wish 1978).

Die Ergebnisse der hierarchischen *Clusteranalyse* nach dem 'average-linkage'-Verfahren bestätigen weitgehend die Ergebnisse der bivariaten Ergebnisse in Kapitel 6.4.1. Nachfolgende Darstellung (Graphik 1) zeigt das resultierende hierarchische Cluster-Schema:

Graphik 1: Ergebnisse der hierarchischen Clusteranalyse



Aus Graphik 1 lassen sich einige Kriterien identifizieren, die eine hohe Ähnlichkeit aufweisen. Dies betrifft - mit abnehmender Intensität (vgl. die senkrechte Achse in der Graphik) - die folgenden Kriterien (bei Beziehungen, die bereits in Kapitel 6.4.1. ausführlich diskutiert wurden, verzichten wir auf einen Kommentar):

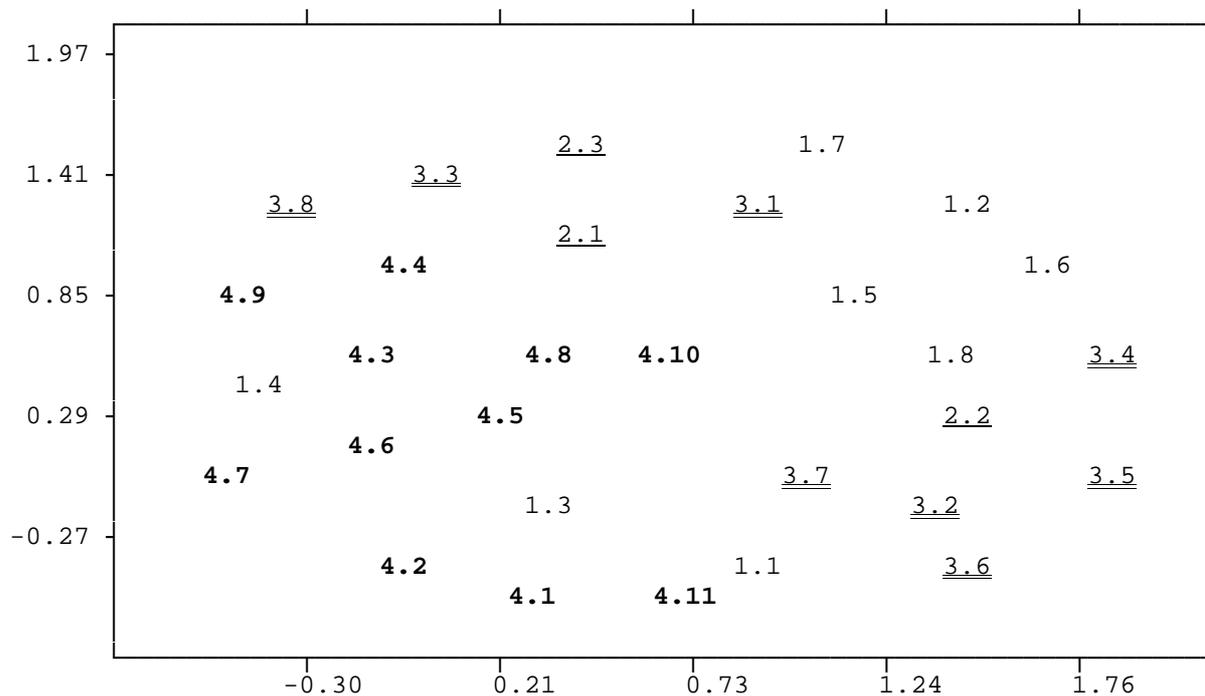
- 'Bewertende Interpretationen' (1.4.) und 'Beschreibung von Zielen und Vorgehen' (4.3.) (vgl. Kapitel 6.4.1.)
- 'Valide Messung' (4.5.) und 'Reliable Messung' (4.6.) (vgl. Kapitel 6.4.1.)
- 'Analyse quantitativer Informationen' (4.8.) und 'Gerechtfertigte Folgerungen' (4.10.) (vgl. Kapitel 6.4.1.)
- 'Wirkung der Evaluation' (1.8.) und 'Politische Überlebensfähigkeit' (2.2.) (vgl. Kapitel 6.4.1.)
- 'Identifikation der beteiligten Gruppen' (1.1.) und 'Informationsrahmen und -auswahl' (1.3.) (vgl. Kapitel 6.4.1.)

- '*Klarheit des Berichts*' (1.5.) und '*Herstellung von Öffentlichkeit*' (3.4.): Der Zusammenhang der beiden Kriterien lässt sich dadurch erklären, dass eine klare Berichterstattung ein Mittel zur Herstellung von Öffentlichkeit darstellt. Beide Kriterien befassen sich mit der Diffusion der Evaluationsergebnisse.
- '*Identifikation des Gegenstandes*' (4.1.) und '*Kontextanalyse*' (4.2.): Die Identifikation des Gegenstandes legt zugleich auch den Gegenstandsbereich der Kontextanalyse fest. Jede Definition stellt implizit immer auch eine Ausgrenzung des 'Nicht-Untersuchten' dar - also die Definition des Kontextes.
- '*Rechtzeitigkeit des Berichts*' (1.7.) und '*Formale Verpflichtung*' (3.1.): Die formalen Verpflichtungen enthalten in der Regel auch zeitliche Vorgaben. Zeitliche Korrektheit respektive zeitliche Verzögerungen wirken sich bei beiden Kriterien aus.
- '*Deklaration von Interessenkonflikten*' (3.2.) und '*Ausgewogene Berichterstattung*' (3.7.): Eine explizite Deklaration von Interessenkonflikten bildet einen wichtigen Aspekt bei der Gewährleistung einer ausgewogenen Berichterstattung.
- '*Glaubwürdigkeit des Evaluators*' (1.2.) und '*Objektive Berichterstattung*' (4.11.): Eine objektive Berichterstattung ist eine der Voraussetzungen, um die Glaubwürdigkeit des Evaluators zu etablieren bzw. zu erhalten. Fällt die Berichterstattung unausgewogen aus, ist damit zu rechnen, dass der Evaluator seine Glaubwürdigkeit zumindest gegenüber einem Teil der Beteiligten oder Betroffenen verliert.

Wie aus dieser Liste der zehn ähnlichsten Kriterienpaare hervorgeht, stimmen die Resultate der Clusteranalyse sehr weitgehend mit den Ergebnisse der bivariaten Interdependenzanalyse (siehe Kapitel 6.4.1.) überein. Die fünf ähnlichsten Paare verfügen ebenso über eine starke bilaterale Beziehung.

Auch die Resultate der *Multidimensionalen Skalierung* (MDS) stimmen mit den bisher referierten Ergebnissen gut überein. Aus der Abbildung in Graphik 2 kann die Lokalisierung der Kriterien im zweidimensionalen Merkmalsraum entnommen werden (zu dieser Vorgehensweise vgl. auch Caracelli/Riggin 1994 und den 'concept mapping'-Ansatz bei Trochim 1989a&b und 1993, Keith 1989 und Lassegard 1992):

Graphik 2: Ergebnisse der Multidimensionalen Skalierung



Die Verteilung der dreissig Kriterien im zweidimensionalen Raum zeigt gewisse Schwerpunkte hinsichtlich der vier Kriteriengruppen:

- Die Kriterien aus der Gruppe *'Nützlichkeit'* (1.) finden sich mehrheitlich rechts oben. Nicht in dieses Muster passen die Kriterien *'Identifikation der beteiligten Gruppe'* (Kriterium 1.1.), *'Informationsrahmen und -auswahl'* (1.3.) und *'Bewertende Interpretation'* (1.4.), die sich inmitten der Kriterien aus der Gruppe Genauigkeit (4.) befinden.
- Die Gruppe *'Anwendbarkeit'* (2.), die lediglich drei Kriterien umfasst, findet sich im Punkteschwarm oben in der Mitte. Das Kriterium *'Politische Überlebensfähigkeit'* (2.2.) weicht davon ab.
- Eine Mehrheit der Kriterien zur Gruppe *'Korrektheit'* (3.) findet sich in der Darstellung rechts unten. Hingegen sind die zwei Kriterien *'Vollständige und aufrichtige Offenlegung'* (3.3.) und *'Finanzielle Verantwortlichkeit'* (3.8.) links oben positioniert, also diametral entgegengesetzt zum Rest der Gruppe.
- Alle Kriterien aus der Gruppe *'Genauigkeit'* (4.) befinden sich im linken, unteren Bereich.

Aufgrund dieser gruppenweisen Betrachtung zeigt sich, dass sich die Kriterien zur *'Genauigkeit'* (4.) in der Multidimensionalen Skalierung am homogensten verhalten. Besonders auffällig ist die Position des Kriteriums *'Bewertende Interpretation'* (1.4.), das vollständig von Kriterien aus der Genauigkeitsgruppe (4.) umgeben wird. Aber auch das Kriterium *'Politische Überlebensfähigkeit'* (2.2.) hat kein angrenzendes Kriterium, das aus derselben Kriteriengruppe stammen würde.

Die in der bivariaten, wie auch in der Clusteranalyse festgestellten Kriterienpaare mit starken Interdependenzen, werden durch die Multidimensionale Skalierung nahe beieinander positioniert. Eine Ausnahme dazu bildet die Diade, bestehend aus den Kriterien 'Glaubwürdigkeit des Evaluators' (1.2.) und 'Objektive Berichterstattung' (4.11.). Bei allen anderen Diaden hingegen stimmen die Resultate der drei verschiedenen Untersuchungstechniken überein.

Das hier präsentierte Ergebnis stellt die beste, zweidimensionale Lösung dar. Einer Empfehlung von Joseph B. Kruskal und Myron Wish (1978: 58) folgend, zogen wir keine Lösungen mit mehr als zwei Dimensionen in Betracht. Auch wenn der Wert für das Stress 1-Mass von Kruskal mit 0.366 gestützt auf die Methodenliteratur (vgl. dazu Backhaus/Erichson/Plinke/Weiber 1994: 454-458, Coxon 1982: 83-88) nicht befriedigend ausfällt und die Anpassung an das empirische Datenmaterial damit nicht besonders gut zu sein scheint, liefert die Multidimensionale Skalierung Illustration zur Beschreibung des Beziehungsnetzes unter den dreissig untersuchten Evaluationskriterien. Aufgrund der Erfahrungen von William Trochim (1993) mit einer Multidimensionalen Skalierung zu einer ähnlich gelagerten Fragestellung ist davon auszugehen, dass die durch die Methodologen gesetzten Anforderungen für Analysen in der vorliegenden Art zu konservativ sind (vgl. dazu auch Lassegard 1992).

6.4.3. Beziehungen der Kriteriengruppen

Um genauere Aufschlüsse darüber zu erhalten, wie sich die Beziehungen der Einzelkriterien innerhalb einer Kriteriengruppe sowie zwischen den vier Gruppen verhalten, haben wir zusätzlich die Beziehungen auf der Ebene der vier Kriteriengruppen 'Nützlichkeit' (1.), 'Anwendbarkeit' (2.), 'Korrektheit' (3.) und 'Genauigkeit' (4.) aggregiert. Hier ergeben sich hinsichtlich der Interdependenzen zwischen den Kriterien die in Tabelle 5 angeführten Konstellationen.

Es resultiert wiederum eine symmetrische Datenmatrix. Im Gegensatz zur Tabelle zu den Einzelkriterien (vgl. oben, Tabelle 3), bestehen hier auch in der Diagonale (von links oben nach rechts unten) Angaben, die für Selbstbeziehungen in den Kriteriengruppen stehen. Neben der Anzahl der Bezüge (N), die aufgrund der unterschiedlichen Kriterienzahl pro Gruppe kaum zu interpretieren sind, enthält die Tabelle auch Angaben zur Beziehungsdichte (Q), die einen - um den Einfluss der unterschiedlichen Kriterienzahl - bereinigten Wert darstellt, der einen unverzerrten Vergleich ermöglicht.

Tabelle 5: Interdependenzen der Kriteriengruppen

Anzahl Bezüge (N) Dichte (Q)	Nützlichkeit (1.)	Anwendbar- keit (2.)	Korrektheit (3.)	Genauigkeit (4.)
Nützlichkeit (1.) N = 198, Q = 0,97	N = 33 Q = 1,18	N = 21 Q = 0,88	N = 29 Q = 0,45	N = 115 Q = 1,31
Anwendbarkeit (2.) N = 87, Q = 1,04		N = 1 Q = 0,33	N = 11 Q = 0,46	N = 54 Q = 1,64
Korrektheit (3.) N = 124, Q = 0,61			N = 13 Q = 0,46	N = 71 Q = 0,81
Genauigkeit (4.) N = 363, Q = 1,38				N = 123 Q = 2,24

Aufgrund der Angaben in der Diagonale der Tabelle 5 lässt sich analysieren, wie stark die zu einer Gruppe gehörenden Kriterien mit den übrigen Kriterien derselben Gruppe vernetzt sind. Diese Werte können somit als Indiz dafür gelten, wie eng die Kriterien einer Gruppe aufeinander bezogen sind. Aufgrund der Angaben zeigt sich, dass die Gruppe der 'Genauigkeit' Kriterien umfasst, die sehr eng miteinander verbunden sind ($Q = 2,24$), währenddem diese Beziehungen in den Gruppen 'Anwendbarkeit' ($Q = 0,33$) und 'Korrektheit' ($Q = 0,46$) nur sehr schwach ausgeprägt sind. Die Kriterien zur 'Nützlichkeit' nehmen diesbezüglich eine Mittelposition ein ($Q = 1,18$). Dass die Kriterien zur Genauigkeit hoch miteinander vernetzt sind, hat sich bereits aufgrund der Analyse zu den Einzelkriterien abgezeichnet (siehe oben, Kapitel 6.4.1. und 6.4.2.).

Neben diesen Beziehungen enthält die Tabelle 5 auch Angaben zu den Verhältnissen zwischen unterschiedlichen Kriteriengruppen. Am stärksten sind dabei die Beziehungen zwischen der Gruppe 'Anwendbarkeit' (2.) und 'Genauigkeit' (4.) ($Q = 1,64$). Recht starke Beziehungen bestehen auch zwischen den Gruppen 'Nützlichkeit' (1.) und 'Genauigkeit' (4.) ($Q = 1,31$). Im Gegensatz dazu sind nur schwache Beziehungen zwischen 'Nützlichkeit' (1.) und 'Korrektheit' (3.) ($Q = 0,45$), zwischen 'Anwendbarkeit' (2.) und 'Korrektheit' (3.) ($Q = 0,46$) sowie zwischen 'Nützlichkeit' (1.) und 'Anwendbarkeit' (2.) ($Q = 0,88$) feststellbar. Aufgrund der Angaben in der ersten Spalte der Tabelle 5 bestätigt sich der Eindruck, dass die Kriterien zur 'Korrektheit' (3.) deutlich am schwächsten vernetzt sind ($Q = 0,61$). Auch hier fällt die hohe Beziehungsdichte der Genauigkeitskriterien auf ($Q = 1,38$).

6.5. Abschliessende Bemerkungen zum Fallstudienvergleich

Die vorangehende Analyse befasste sich mit einem Vergleich der zehn in dieser Untersuchung durchgeführten Fallstudien zu empirischen Evaluationsstudien sowie mit den Interdependenzen zwischen den dreissig Kriterien und den vier Kriteriengruppen, wie sie in der empirischen Analyse der zehn untersuchten Evaluationsstudien festgestellt wurden. Die in diesem Zusammenhang referierten Ergebnisse sind *nicht* in dem Sinn zu verstehen, dass damit allgemeingültige Ergebnisse erzielt wurden. Dies ist schon rein deshalb nicht möglich, da dazu die von uns untersuchten zehn Evaluationsstudien in verschiedener Hinsicht nicht als repräsentativ für die gesamte Evaluationsforschung gelten können. Als weiteres Problem stellt sich einer solchen Verallgemeinerung auch entgegen, dass die eingesetzten Kriterien einiges an Ermessensspielraum für deren empirische Umsetzung enthalten (vgl. dazu Kapitel 7.3.). Bei der Erarbeitung der Fallstudien wurde darum spezielles Gewicht darauf gelegt, dass die Argumentation ausführlich dargestellt und damit nachvollziehbar wird. Die Absicht besteht darin, eine 'plausible inference' (Cronbach et al. 1980) zu etablieren. Aus den dargelegten Gründen können die angeführten Feststellungen 'lediglich' als empirisch abgestützte Thesen, nicht aber als bewährte Aussagen im Sinne Poppers (1989: 198-226) gelten. Wir haben bereits an anderer Stelle auf den - in diesem Sinne - explorativen Charakter der Studie hingewiesen (siehe Kapitel 4.6.).

Vergleicht man die Ergebnisse der Kriterienbewertung mit jenen aus der Untersuchung der Interdependenzen der Kriterien, fällt auf, dass Kriterien mit einer überdurchschnittlich positiven Bewertung oder solche, bei denen keine Bewertung möglich war, über schwache Beziehungen zu anderen Kriterien verfügen, währenddem negativ bewertete Kriterien besonders starke Interdependenzen aufweisen. Dies muss als Indiz dafür gelten, dass die Analyse der Interdependenzen durch die Bewertungen beeinflusst wurde. Diese Verzerrung ist wohl primär damit zu erklären, dass sich die Verweise in den Fallstudien häufig auf festgestellte Schwächen, Mängel oder Lücken beziehen, die für die Bewertung verschiedener Kriterien relevant sind. Aufgrund dieser Situation ist damit zu rechnen, dass - wären die Stärken und Schwächen der Evaluationen anders auf die Kriterien verteilt - die Analyse der Interdependenzen leicht anders hätte ausfallen können. Massgebliche Verzerrungen sind unseres Erachtens dadurch aber nicht entstanden,

Das nachfolgende Kapitel 7 diskutiert die Folgerungen, die sich aus den Fallstudien und der vergleichenden Analyse ziehen lassen.

7. FOLGERUNGEN AUS DER EMPIRISCHEN ANALYSE

Die in diesem Kapitel enthaltenen Ausführungen basieren auf den Fallstudien zu zehn Evaluationen (siehe Kapitel 5 und Widmer 1996a-g) und deren vergleichender Analyse (siehe Kapitel 6). Diese Vorgehensweise bringt es mit sich, dass die nachfolgenden Aussagen über keinen universellen Geltungsanspruch verfügen können. Aus diesem Grund ist es zweckmässig, an dieser Stelle nochmals dezidiert auf die Begrenzungen des Gegenstandsbereichs dieser Untersuchung und damit auf den Geltungsbereich ihrer Ergebnisse hinzuweisen (vgl. zur Auswahl der zehn Fallstudien Kapitel 4.2.). Die Meta-Evaluation behandelt:

- ausschliesslich schweizerische Evaluationsstudien. Die Ergebnisse sind deshalb nicht ohne weitere Abklärungen auf die Situation in anderen Ländern übertragbar.
- lediglich externe Evaluationen, also Evaluationen, die durch externe Evaluatoren durchgeführt wurden. Eine Übertragung der hier gemachten Feststellungen auf interne Evaluationen ist deshalb nicht möglich (vgl. dazu Love 1983, 1991 und 1993, Mathison 1991, Torres 1991, Winberg 1991, Duffy 1994, Sonnichsen 1994, Lovell 1995).
- ausschliesslich Studien, deren Resultate in irgendeiner Form publiziert wurden. Die vorgetragenen Schlüsse sind deshalb nicht auf Evaluationen übertragbar, die der Öffentlichkeit nicht durch einen publizierten Bericht zugänglich gemacht wurden oder werden.
- nur Studien, die im weitesten Sinn einem sozialwissenschaftlichen Ansatz folgen. Daraus ergibt sich, dass die Ergebnisse nicht auf Evaluationen anderer Ausrichtungen angewendet werden können.
- Studien, die in den Jahren 1990 bis 1994 abgeschlossen wurden. Rückschlüsse auf Studien, die nicht aus dieser Zeitspanne stammen, sind nicht zulässig.
- keine akademischen Arbeiten (Diplom-, Lizentiatsarbeiten oder Dissertationen). Die Untersuchung lässt deshalb auch keine Aussagen auf Arbeiten mit diesem Hintergrund zu.
- keine Studien aus dem Bereich der Erziehungswissenschaften oder der Sozial- und Präventivmedizin. Eine Übertragung der Resultate der vorliegenden Untersuchung auf diese Gebiete ist deshalb nicht möglich.

Innerhalb dieser Grenzen ist anhand einer Reihe von Dimensionen eine theoretische Stichprobe bestimmt worden. Es wurde somit keine Zufallsauswahl, sondern eine gezielte Auswahl getroffen. Aufgrund der beschränkten Zahl von berücksichtigten Studien und

aufgrund der Vorgehensweise bei der Auswahl, sind die nachfolgenden Folgerungen und Empfehlungen nicht als bewährte Aussagen im Sinne der Popper'schen Erkenntnistheorie (Popper 1989: 198-226), sondern als empirisch abgestützte Thesen zu gewichten.

Das nachfolgende Kapitel befasst sich mit drei verschiedenen Themenkomplexen. Der erste Abschnitt richtet sich auf den in dieser Untersuchung eingesetzten Kriterienraster, thematisiert seine Umsetzbarkeit und formuliert Vorschläge zu seiner Verbesserung (Kapitel 7.1.). Im zweiten Kapitel werden Folgerungen diskutiert, die sich auf die praktische Durchführung von Evaluationen richten und Empfehlungen enthalten, die sich als Beitrag zur Qualitätssicherung bzw. Qualitätsverbesserung verstehen (Kapitel 7.2.). Der dritte Abschnitt resümiert sodann die im Rahmen dieser Meta-Evaluation gewonnenen Erfahrungen und diskutiert Folgerungen, die sich für die zukünftige Durchführung von Meta-Evaluationen ergeben (Kapitel 7.3.).

7.1. Konsequenzen hinsichtlich der Kriterien

7.1.1. Untersuchungsergebnisse zum gesamten Kriterienraster

Aufgrund der Erfahrungen aus der empirischen Analyse in den zehn Fallstudien kommen wir zum Schluss, dass sich der von uns eingesetzte Kriterienraster *grundsätzlich bewährt* hat. Die dreissig Einzelkriterien waren in der Lage, die Bewertung der einzelnen Evaluationsstudien bezüglich ihrer Stärken und Schwächen in den einzelnen Gebieten zu systematisieren. Dabei ergaben sich - sowohl zwischen den einzelnen Studien wie auch unter den Einzelkriterien - zum Teil grössere Differenzen in der Bewertung, was darauf hinweist, dass der Kriterienraster über eine ausreichend hohe Trennschärfe verfügt und genügend selektiv ist. Besonders positiv hervorzuheben ist, dass der Kriterienraster in der Lage ist, alle relevanten Themen vollständig abzudecken. Im Laufe der Umsetzung des Rasters auf die zehn Evaluationsstudien waren wir nie mit der Situation konfrontiert worden, dass wir eine Beobachtung keinem der Kriterien hätten zuordnen können - im Gegenteil bestand die Problematik eher darin, dass sich aufgrund von Überlappungen (vgl. dazu die Hinweise zur Eindeutigkeit des Kriterienrasters in Kapitel 3.5.) häufig eine mehrfache Zuordnung aufdrängte. Bei der Anwendung des Kriterienrasters, der auf den ursprünglich für den US-amerikanischen Bildungsbereich geschaffenen Standards des Joint Committee (Joint Committee 1981a) beruht, auf schweizerische Evaluationsstudien aus ganz unterschiedlichen Politikbereichen stellen sich zwei Fragen:

1. Hat sich der Kriterienraster zur Einschätzung von *Evaluationen im schweizerischen Kontext* bewährt?
2. Ist der Kriterienraster zur Einschätzung von *Evaluationen ausserhalb des Bildungsreichs* geeignet?

Beide Fragen können aufgrund der Erfahrungen in den zehn empirischen Fallstudien positiv beantwortet werden. Weder hat sich gezeigt, dass der Kriterienraster aufgrund seiner Prägung durch US-amerikanische Spezifika auf schweizerische Verhältnisse nicht anwendbar wäre (vgl. dazu im Hinblick auf andere Länder: Lewy 1984, Marklund 1984, Nevo 1984, Smith/Chricop/Mukherjee 1993; siehe auch Hendricks/Conner 1995), noch bestehen Anhaltspunkte dafür, dass der Kriterienraster bildungsspezifische Charakteristiken aufweisen würde, die seine Umsetzung in anderen Politikbereichen beeinträchtigen könnten (siehe hierzu auch Stockdill 1986 und Patton 1994: 194).

Trotz der generell sehr positiven Einschätzung des Kriterienrasters drängen sich unseres Erachtens punktuelle Änderungen auf, die wir im nachfolgenden erläutern und begründen

werden. Dabei wird unter anderem auch der Versuch unternommen, den doch recht umfangreichen (und damit auch schwerfälligen) Kriterienraster etwas zu kürzen. Dies geschieht nicht nur in der Absicht, die Anwendbarkeit des Kriterienrasters zu erhöhen, sondern auch um die bereits früher festgestellten Schwächen - wie die fehlende Eindeutigkeit, Einheitlichkeit, Konsistenz und Unabhängigkeit (vgl. oben, Kapitel 3.5.) - zu reduzieren.

Die nachfolgenden Vorschläge basieren auf den Erfahrungen im Rahmen dieser Meta-Evaluation, wobei auch die Neuerungen der zweiten Ausgabe der 'Program Evaluation Standards' (Joint Committee 1994, vgl. dazu Kapitel 3.6.1.) berücksichtigt werden.

7.1.2. Folgerungen und Vorschläge zu Einzelkriterien

Aufgrund der in der empirischen Analyse gewonnenen Erkenntnisse erscheint es unseres Erachtens zweckmässig, das Kriterium *'Bewertende Interpretationen'* (Kriterium 1.4.) ersatzlos zu streichen. In der Untersuchung der Beziehungen dieses Kriteriums zu den übrigen Einzelkriterien wurde deutlich, dass starke Interdependenzen zu den Kriterien *'Beschreibung von Zielen und Vorgehen'* (Kriterium 4.3.), *'Gerechtfertigte Folgerungen'* (Kriterium 4.10.) und *'Klarheit des Berichts'* (Kriterium 1.5.) bestehen, die auf inhaltliche Überschneidungen zurückzuführen sind (vgl. Kapitel 6.4.1. und 6.4.2.). Da diese drei Kriterien den Gegenstandsbereich des Kriteriums *'Bewertende Interpretationen'* ausreichend abdecken, kommen wir zum Schluss, dass dieses Kriterium aus dem Kriterienraster eliminiert werden kann, ohne dass damit ein substantieller Verlust in Kauf genommen werden muss. Diese Folgerung wird auch durch die Resultate der Clusteranalyse sowie der Multidimensionalen Skalierung im Rahmen der Netzwerkanalyse (vgl. Kapitel 6.4.2.) gestützt.

Die zweite Version der Joint Committee-Standards schlägt beim Kriterium *Klarheit des Berichts* (Kriterium 1.5.) eine sprachliche Vereinfachung vor, die wir für zweckmässig halten. Wir schlagen vor, das Kriterium neu folgendermassen zu formulieren:

Der Evaluationsbericht sollte das Evaluandum und seinen Kontext ebenso beschreiben wie die Absichten, die Verfahren und die Ergebnisse der Evaluation, damit die wesentlichen Informationen zur Verfügung stehen und leicht verstanden werden können.

Eine weitere Neuerung, die in der Überarbeitung der amerikanischen Standards enthalten ist, besteht darin, dass die Kriterien *'Verbreitung des Berichts'* (Kriterium 1.6.) und

'*Rechtzeitigkeit des Berichts*' (Kriterium 1.7.) in einem Kriterium zusammengefasst werden. Dies erscheint deshalb sinnvoll, weil sie inhaltlich eng miteinander verbunden sind. So kann ein Evaluationsbericht, der erst lange Zeit nachdem das Thema aktuell gewesen wäre, weit verbreitet wird, ebensowenig überzeugen, wie ein Bericht, der zwar zeitgerecht fertiggestellt wurde, jedoch keine Verbreitung findet. Wir schlagen für das zusammenfassende Kriterium 'Rechtzeitigkeit und Verbreitung des Berichts' folgende Formulierung vor:

Wichtige Zwischenergebnisse und Evaluationsberichte sollten den Beteiligten und Betroffenen in einer Weise zukommen, so dass sie rechtzeitig genutzt werden können.

Das Kriterium zur '*Wirkung der Evaluation*' (Kriterium 1.8.) hat in der empirischen Analyse insofern zu Problemen geführt, als dass im Rahmen der vorliegenden Untersuchung vielfach keine Bewertung abgegeben werden konnte (vgl. dazu Kapitel 6.1.1.8. und 6.3.3.). Trotz diesem Resultat sind wir, wegen der grossen Bedeutung, die diesem Kriterium zukommt, der Ansicht, dass auf diesen Aspekt keinesfalls verzichtet werden sollte. Die Wirkungen einer Evaluation stellen einen zentralen Aspekt dar, sowohl für die an der Evaluation beteiligten wie auch für die durch sie betroffenen Personen oder Institutionen. Die Schwierigkeiten, die sich bei der empirischen Umsetzung des Kriteriums zeigten, sind unseres Erachtens keinesfalls ein Grund, auf das Kriterium zu den Evaluationswirkungen zu verzichten. Die revidierte Fassung der amerikanischen Standards schlägt eine sprachliche Präzisierung des Kriteriums vor, die dessen Aussagekraft und Eindeutigkeit erhöht (vgl. dazu Kapitel 3.6.1.2.). Das Kriterium 'Wirkung der Evaluation' lautet deshalb neu folgendermassen:

Evaluationen sollten so geplant, durchgeführt und beschrieben werden, dass die Beteiligten und Betroffenen dazu ermuntert werden, der Evaluation zu folgen, damit sich die Wahrscheinlichkeit erhöht, dass die Evaluation genutzt wird.

Die Formulierung des Kriteriums '*Kostenwirksamkeit*' (Kriterium 2.3.) ist in der neuen Version der amerikanischen Standards abgeändert worden. Da wir die Ansicht vertreten, dass diese Veränderungen nicht zu einer Verbesserung des Kriteriums beitragen können (vgl. die Hinweise dazu in Kapitel 3.6.1.3.), schlagen wir vor, die hier eingesetzte Fassung beizubehalten.

Entsprechend der in den amerikanischen Standards vorgenommenen Änderung der Bezeichnung schlagen wir vor, das Kriterium '*Formale Verpflichtung*' (Kriterium 3.1.) neu

mit 'Formale Vereinbarung' zu benennen. So wird klar, dass die Vereinbarung neben Verpflichtungen durchaus auch Berechtigungen der Parteien enthält. Weiter kommt mit der neuen Bezeichnung auch besser zum Ausdruck, dass Evaluationsvereinbarungen ausgehandelt werden, also nicht nur ein Produkt, sondern auch den zum Produkt führenden Prozess umfassen. Das Kriterium ist in der älteren Fassung etwas umständlich formuliert, weshalb wir die folgende, vereinfachte Formulierung vorschlagen:

Die formalen Parteien einer Evaluation sollten den Vereinbarungen in schriftlicher Form zustimmen, damit sie zur Einhaltung aller Bedingungen verpflichtet sind oder damit sie diese aufs neue verhandeln müssen.

Das Kriterium '*Deklaration von Interessenkonflikten*' (Kriterium 3.2.) kann unseres Erachtens um die Zwischenbemerkung '- oft unausweichlich -' gekürzt werden. Diese Änderung ist auch in der amerikanischen Neuauflage vorgenommen worden. Damit erhält das Kriterium neu folgende Formulierung:

Interessenkonflikte sind offen und aufrichtig zu behandeln, damit sie den Prozess und die Resultate der Evaluation nicht kompromittieren.

Zu den beiden Kriterien '*Vollständige und aufrichtige Offenlegung*' (Kriterium 3.3.) und '*Herstellung von Öffentlichkeit*' (Kriterium 3.4.) werden in der neueren Fassung der Standards des Joint Committee neue Bezeichnungen und Formulierungen vorgeschlagen. Die bisherigen Fassungen haben sich jedoch in unserer Analyse bewährt. Die in der älteren Version des Kriteriums '*Vollständige und aufrichtige Offenlegung*' (Kriterium 3.3.) eingesetzten Formulierungen sind unseres Erachtens einfacher und klarer. Die neuere Fassung des Kriteriums '*Herstellung von Öffentlichkeit*' (Kriterium 3.4.) enthält Begriffe (wie 'service orientation', 'organizations' und 'targeted participants', siehe Kapitel 3.6.1.4.), die sich für die Verwendung ausserhalb des Erziehungssektors nicht immer eignen. Diese Änderungen würden den Geltungsbereich des Kriteriums auf Evaluanda einschränken, die über Teilnehmer verfügen. Dies ist zwar im Bildungsbereich zumeist gegeben, in anderen Politikbereichen jedoch sehr oft nicht (vgl. dazu auch die Hinweise in Kapitel 3.6.1.4.). Aus diesen Gründen schlagen wir vor, diese Kriterien in unveränderter Fassung beizubehalten.

Das Kriterium '*Menschliche Interaktionen*' (Kriterium 3.6.) hat in den empirischen Fallstudien teilweise nicht bewertet werden können. Die Gründe dafür liegen weniger in grundsätzlichen Problemen im Zusammenhang mit dem Kriterium, als vielmehr in der im vorliegenden Projekt verfolgten Untersuchungsanlage. So drängen sich aufgrund der em-

pirischen Analyse keine Änderungen auf. Das Joint Committee hat in der neuen Fassung der 'Program Evaluation Standards' einen ergänzenden Nachsatz zur Verdeutlichung der Zielsetzung des Kriteriums hinzugefügt. Durch die explizite Nennung der Absicht des Kriteriums gewinnt dieses an Aussagekraft. Deshalb schlagen wir die folgende, ergänzte Formulierung des Kriteriums vor:

Evaluatoren sollten in den Interaktionen mit anderen Personen im Zusammenhang mit der Evaluation die Würde und den Wert der Menschen respektieren, damit die Beteiligten nicht gefährdet oder geschädigt werden.

Die neue Version der amerikanischen Standards erweitert den Geltungsbereich des Kriteriums 'Ausgewogene Berichterstattung' (Kriterium 3.7.) vom Bericht auf die gesamte Untersuchung. Diese Erweiterung halten wir für zweckmässig, da sich das Kriterium damit besser dazu eignet, die Evaluation als Ganzes zu bewerten. Die Beschränkung auf die Berichterstattung vernachlässigt prozessuale Aspekte. Eine weitere Änderung hingegen, die sich auf eine stärkere Gewichtung der Vollständigkeit zu Lasten der Ausgewogenheit bezieht, halten wir für weniger sinnvoll. Da es in empirischen Evaluationsstudien kaum möglich ist, eine vollständige Untersuchung und eine ebensolche Berichterstattung zu gewährleisten, tendieren wir dazu, die Vollständigkeit nicht allzu stark zu betonen und stattdessen eher die Ausgewogenheit in den Vordergrund zu rücken (siehe dazu die Argumentation bei Patton 1994: 197; vgl. Kapitel 3.6.1.4.). Deshalb schlagen wir vor, das Kriterium neu mit 'Ausgewogene Einschätzung' zu bezeichnen und folgendermassen zu formulieren:

Evaluationen sollten in der Überprüfung und in der Präsentation der Stärken und Schwächen des Evaluandums vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder benannt werden können.

Das Kriterium 'Identifikation des Gegenstandes' (Kriterium 4.1.) hat sich im Rahmen der zehn empirischen Fallstudien bewährt. Das Joint Committee schlägt nun für dieses Kriterium eine neue Bezeichnung und eine neue Formulierung vor. Da die neue Formulierung einfacher und klarer ist und weil die neue Bezeichnung näher an den Inhalt des Kriteriums herankommt, schlagen wir vor, das Kriterium neu mit 'Dokumentation des Programms' zu bezeichnen und folgendermassen zu formulieren:

Das zu evaluierende Programm sollte klar und genau beschrieben und dokumentiert werden, so dass es deutlich identifiziert werden kann.

Die Kriterien zur *'Validen Messung'* (Kriterium 4.5.) und zur *'Reliablen Messung'* (Kriterium 4.6.) haben sich bei der empirischen Analyse der Interdependenzen als stark vernetzt erwiesen. Die Bezüge zwischen den beiden Kriterien sind sowohl in der bivariaten (siehe Kapitel 6.4.1.) wie auch in der multivariaten Betrachtungsweise (siehe Kapitel 6.4.2.) ausgesprochen stark. Da die Konzepte Validität und Reliabilität auch aufgrund der sozialwissenschaftlichen Methodologie stark miteinander verbunden sind und wegen der Wechselwirkungen kaum losgelöst voneinander betrachtet werden können, schlagen wir vor, diese beiden Kriterien in einem Kriterium zusammenzufassen und mit *'Valide und reliable Information'* zu bezeichnen. Die Änderung in der Bezeichnung - von *'Messung'* zu *'Information'* - geht auf den Vorschlag des Joint Committee zurück. Für das zusammenfassende Kriterium resultiert somit folgende Formulierung:

Die Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit und die Zuverlässigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt sind.

Die Parallelkriterien, die wir für Evaluationen mit alternativer Orientierung eingesetzt haben (also die Kriterien *'Glaubwürdigkeit'* sowie *'Übertragbarkeit'*, *'Verlässlichkeit'* und *'Bestätigbarkeit'*), werden unverändert beibehalten. Da in diesem Bereich bisher aus der Literatur keine neueren Fassungen bekannt sind und die in unserer Analyse gewonnenen Erkenntnisse nur eine sehr schmale empirische Basis aufweisen (nur gerade zwei der zehn Studien sind mit dem alternativen Kriterienset evaluiert worden), sind keine fundierten Änderungsvorschläge möglich. Es besteht unseres Erachtens in diesem Bereich jedoch noch grosser Handlungsbedarf, kann doch die von uns gewählte Vorgehensweise, die Vielfalt der erkenntnistheoretischen Positionen auf zwei Kategorien zu beschränken, nicht überzeugen. Es ergaben sich nicht nur Probleme bei der Zuordnung einzelner Fallstudien (so etwa bei der Studie Nr. 9 *'Evaluation des BUWAL'*, siehe Kapitel 5.3.), sondern auch beim Einsatz der betreffenden Kriterien. So bezogen sich Evaluationsstudien vereinzelt auf Qualitätskriterien der anderen methodologischen Ausrichtung (siehe beispielsweise in Studie Nr. 3 die Hinweise zur Validität in Kapitel 5.1.3.4.5), was deutlich darauf hinweist, dass die dichotome Fassung des Kriteriensets die epistemologische Verschiedenartigkeit nicht abzubilden vermag. Hingegen muss auch darauf hingewiesen werden, dass eine Erweiterung der Zahl der eingesetzten Kriterienraster die Vergleichbarkeit zwischen den resultierenden Einschätzungen reduziert.

Das Joint Committee schlägt vor, das Kriterium *'Systematische Datenüberprüfung'* (Kriterium 4.7.) neu mit der Bezeichnung *'Systematische Information'* zu bezeichnen. Wir folgen dieser Änderung nicht, da unseres Erachtens die ältere Bezeichnung den Inhalt des

Kriteriums besser wiedergibt. Dieses befasst sich nämlich ganz klar mit der Datenüberprüfung und nicht mit der Information als solcher.

Die Kriterien zur *'Analyse quantitativer Information'* (Kriterium 4.8.) und zur *'Analyse qualitativer Information'* (Kriterium 4.9.) haben in der empirischen Bewertung in zweierlei Hinsicht zu Problemen geführt. Erstens waren Studien zu bewerten, die sich ausschliesslich Verfahren entweder der quantitativen oder qualitativen Datenanalyse bedienen. In diesen Fällen konnte das jeweils andere Kriterium nicht zweckmässig eingesetzt werden. Diese Problematik erscheint unseres Erachtens nicht von allzu grosser Relevanz zu sein, da wir die Ansicht vertreten, dass zur Bearbeitung der meisten evaluativen Fragestellungen mit Vorteil eine kombinierte Methodik eingesetzt werden sollte (vgl. Cook/Reichhardt 1979, Cook 1985, Bryman 1988, Denzin 1989: 234-247, Patton 1990: 464-472 sowie Hakim 1992: 144-145). Aus dieser Perspektive wäre auch die Variante zu erwägen, dass das Fehlen einer Analyse quantitativer bzw. qualitativer Informationen auch eine ungenügende Bewertung mit sich bringen würde.

Zweitens hat sich in der empirischen Analyse gezeigt, dass aufgrund einer unzureichenden methodischen Berichterstattung vor allem auch im qualitativen Bereich verschiedentlich Bewertungen ausbleiben mussten. Daraus ergeben sich jedoch unseres Erachtens keine Gründe zu Änderungen an den Kriterien. Vielmehr sollte eine deutliche Verbesserung der methodischen Berichterstattung angestrebt werden (vgl. dazu Kapitel 7.2.). Hingegen sehen wir uns durch die vom Joint Committee vorgenommenen Änderungen an den Standards veranlasst, eine leichte Veränderung an der Formulierung vorzunehmen (vgl. dazu auch Kapitel 3.6.1.5.): Anstatt der Zielsetzung, haltbare Interpretationen sicherzustellen, soll sich das Kriterium neu stärker an der evaluativen Funktion der Analyse orientieren. Wir halten diese leichte Verschiebung der Zielsetzung für zweckmässig, wird damit doch deutlich gemacht, dass sich die haltbaren Interpretationen zur Bearbeitung der evaluativen Fragestellung eignen müssen.

Aufgrund unserer Erfahrungen in der empirischen Analyse und wegen den oben angeführten Überlegungen zu uni- und multimethodischen Vorgehensweisen in der Evaluationsforschung schlagen wir vor, die beiden Kriterien zusammenzufassen, mit *'Analyse qualitativer und quantitativer Information'* zu bezeichnen und neu folgendermassen zu formulieren:

Qualitative und quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, so dass die Fragen der Evaluation effektiv beantwortet werden.

In der empirischen Analyse der zehn Evaluationsstudien und vor allem aus den Reaktionen der zur Stellungnahme aufgeforderten Evaluatoren wurde deutlich, dass die Bezeichnung des Kriteriums '*Gerechtfertigte Folgerungen*' (Kriterium 4.10.) zu Missverständnissen führen kann. Das Kriterium beinhaltet das Erfordernis, dass Folgerungen *explizit* zu begründen sind. Es geht also hier nicht um die Frage, ob die gezogenen Schlüsse an und für sich gerechtfertigt wären, sondern um die explizite Darlegung der Begründungen. Um dies auch in der Bezeichnung des Kriteriums deutlicher zu machen, soll das Kriterium neu '*Begründete Folgerungen*' genannt werden.

Das Kriterium '*Objektive Berichterstattung*' (Kriterium 4.11) erhält in den amerikanischen Standards aus dem Jahre 1994 eine neue Bezeichnung und eine leichte Veränderung des Gegenstandsbereichs. Die Bezeichnung lautet neu '*Unvoreingenommene Berichterstattung*'. Die Änderung erlaubt den Verzicht auf den Begriff 'objektiv', der in der erkenntnistheoretischen Debatte stark umstritten ist. Michael Quinn Patton äussert sich dazu folgendermassen: "The original 'Objective Reporting' has been revised to 'Impartial Reporting,' thereby avoiding the rhetorical and philosophical quagmire about whether anyone can be 'objective.' I find this change, subjectively, a welcome improvement." (Patton 1994: 198). Da unseres Erachtens bei der Formulierung von Kriterien der vorliegenden Art die möglichst weite Akzeptanz der verwendeten Begrifflichkeit von grosser Bedeutung sind, folgen wir dem Vorschlag des Joint Committee und bezeichnen das Kriterium neu mit '*Unvoreingenommene Berichterstattung*'.

Die zweite Änderung betrifft eine Einengung des Gegenstandsbereichs des Kriteriums von 'evaluation procedures' (1981) zu 'reporting procedures' (1994). Da wir die Ansicht vertreten, dass Evaluationsverfahren im allgemeinen und nicht nur die eigentlichen Berichtsverfahren zu einer unvoreingenommenen Berichterstattung beitragen können und müssen, behalten wir die herkömmliche Formulierung - im Gegensatz zur Bezeichnung - des Kriteriums bei.

Die zweite Auflage der 'Program Evaluation Standards' umfasst einen gänzlich neuen Eintrag mit einem Kriterium zur Meta-Evaluation. Aufgrund unserer Erfahrungen aus der durchgeführten Meta-Evaluation zu zehn Evaluationsstudien treten wir dafür ein, diesen Standard neu in den Kriterienraster aufzunehmen. Die Erfahrungen haben gezeigt, dass die Durchführung von Meta-Evaluationen einem Bedürfnis entspricht (vgl. Kapitel 7.2. und 7.3.). Aus diesen Überlegungen heraus halten wir es für angebracht, eine Ergänzung des Kriterienrasters um das Kriterium 'Meta-Evaluation' vorzunehmen, das nun folgendermassen formuliert wird:

Die Evaluation selbst sollte formativ und summativ mit den vorliegenden oder anderen Kriterien evaluiert werden, so dass die Durchführung entsprechend angeleitet werden kann und damit die Beteiligten und Betroffenen bei Abschluss der Evaluation ihre Stärken und Schwächen gründlich überprüfen können.

Das Joint Committee hat den Standard zur Meta-Evaluation der Gruppe der Genauigkeitsstandards zugeordnet (siehe Joint Committee 1994: 185-190). Aufgrund der inhaltlichen Bedeutung scheint uns dies nicht unbedingt angebracht zu sein, ist doch die Funktion der Meta-Evaluation in unserem Verständnis nicht auf die 'Genauigkeit' beschränkt. Weil eine Zuteilung zu einer der anderen drei Kriteriengruppen jedoch die gleiche Problematik hervorrufen würde, folgen wir bezüglich der Zuteilung dem Joint Committee. Die Alternative, dieses Kriterium separat aufzuführen, haben wir verworfen, da dadurch eine fünfte Kriteriengruppe (bestehend aus einem Einzelkriterium) entstehen würde, was den Überblick über den Gesamtraster erschwert.

7.1.3. Vorschlag für einen revidierten Kriterienraster

Aufgrund der im vorangehenden Kapitel angeführten Änderungsvorschläge ergibt sich der nachfolgende Kriterienraster. Zur Numerierung der revidierten Kriterien werden die Kurzbezeichnungen 'N' für Nützlichkeit, 'A' für Anwendbarkeit, 'K' für Korrektheit und 'G' für Genauigkeit eingesetzt. Aus der Überarbeitung ergibt sich ein Kriterienraster mit nunmehr 27 Einzelkriterien, wovon sechs Kriterien auf die Gruppe 'Nützlichkeit', drei auf die Gruppe 'Anwendbarkeit', acht auf die Gruppe 'Korrektheit' und zehn auf die Gruppe 'Genauigkeit' entfallen. Der nachfolgend präsentierte Kriterienraster bildet einen Vorschlag aufgrund der empirischen Analysen und neuerer Entwicklungen in der Literatur, der auf seine Tauglichkeit und Akzeptanz hin wieder neu überprüft werden muss.

Revidierter Kriterienraster zur Bewertung von Evaluationen

N. Nützlichkeit

N.1. Identifikation der beteiligten Gruppen

Die Beteiligten oder Betroffenen einer Evaluation sollten identifiziert werden, damit deren Interessen berücksichtigt werden können.

N.2. Glaubwürdigkeit des Evaluators

Die Personen, welche die Evaluation durchführen, sollten sowohl vertrauenswürdig wie auch kompetent sein, damit die Ergebnisse der Evaluation ein Höchstmass an Glaubwürdigkeit und Akzeptanz erreichen.

N.3. Informationsrahmen und -auswahl

Die beschafften Informationen sollten von einem Umfang und einer Auswahl sein, welche die Behandlung sachdienlicher Fragen zum Evaluandum ermöglichen und gleichzeitig die Interessen der Beteiligten und Betroffenen berücksichtigen.

N.4. Klarheit des Berichts

Der Evaluationsbericht sollte das Evaluandum und seinen Kontext ebenso beschreiben wie die Absichten, die Verfahren und die Ergebnisse der Evaluation, damit die wesentlichen Informationen zur Verfügung stehen und leicht verstanden werden können.

N.5. Rechtzeitigkeit und Verbreitung des Berichts

Wichtige Zwischenergebnisse und Evaluationsberichte sollten den Beteiligten und Betroffenen in einer Weise zukommen, so dass sie rechtzeitig genutzt werden können.

N.6. Wirkung der Evaluation

Evaluationen sollten so geplant, durchgeführt und beschrieben werden, dass die Beteiligten und Betroffenen dazu ermuntert werden, der Evaluation zu folgen, damit sich die Wahrscheinlichkeit erhöht, dass die Evaluation genutzt wird.

A. Anwendbarkeit

A.1. Praktische Verfahren

Die Evaluationsverfahren sollten praktisch sein, so dass Störungen minimal gehalten und die benötigten Informationen beschafft werden können.

A.2. Politische Überlebensfähigkeit

Evaluationen sollten mit Voraussicht auf die unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt werden, so dass deren Kooperation erreicht werden kann und so dass mögliche Versuche irgendeiner dieser Gruppen, die Evaluationsoperationen zu beschneiden oder die Resultate zu verzerren oder zu missbrauchen, vermieden werden können.

A.3. Kostenwirksamkeit

Die Evaluation sollte Informationen mit einem Wert produzieren, der die gemachten Aufwendungen rechtfertigt.

K. Korrektheit

K.1. Formale Vereinbarung

Die formalen Parteien einer Evaluation sollten den Vereinbarungen in schriftlicher Form zustimmen, damit sie zur Einhaltung aller Bedingungen verpflichtet sind oder damit sie diese aufs neue verhandeln müssen.

K.2. Deklaration von Interessenkonflikten

Interessenkonflikte sind offen und aufrichtig zu behandeln, damit sie den Prozess und die Resultate der Evaluation nicht kompromittieren.

K.3. *Vollständige und aufrichtige Offenlegung*

Mündliche und schriftliche Evaluationsberichte sollten in der Darlegung der sachdienlichen Ergebnisse und der Grenzen der Evaluation offen, direkt und ehrlich sein.

K.4. *Herstellung von Öffentlichkeit*

Die formellen Parteien einer Evaluation sollten das Recht der Öffentlichkeit auf Information respektieren und sicherstellen, wobei die Begrenzungen durch andere verwandte Prinzipien und Regelungen, wie die öffentliche Sicherheit und das Recht auf Privatsphäre, berücksichtigt werden müssen.

K.5. *Schutz individueller Menschenrechte*

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

K.6. *Menschliche Interaktionen*

Evaluatoren sollten in den Interaktionen mit anderen Personen im Zusammenhang mit der Evaluation die Würde und den Wert der Menschen respektieren, damit die Beteiligten nicht gefährdet oder geschädigt werden.

K.7. *Ausgewogene Einschätzung*

Evaluationen sollten in der Überprüfung und in der Präsentation der Stärken und Schwächen des Evaluandums vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder benannt werden können.

K.8. *Finanzielle Verantwortlichkeit*

Die Zuweisung und Ausgabe von Ressourcen durch den Evaluator sollte unter Berücksichtigung fundierter und verantwortlicher Verfahren und auch anderweitig klug und ethisch verantwortlich erfolgen.

G. *Genauigkeit*

G.1. *Dokumentation des Programms*

Das zu evaluierende Programm sollte klar und genau beschrieben und dokumentiert werden, so dass es deutlich identifiziert werden kann.

G.2. *Kontextanalyse*

Der Kontext des Evaluandums sollte ausreichend detailliert untersucht werden, so dass mögliche Beeinflussungen des Evaluandums identifiziert werden können.

- G.3. Beschreibung von Zielen und Vorgehen*
Die Ziele und das Vorgehen der Evaluation sollten ausreichend genau beobachtet und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.
- G.4. Verlässliche Informationsquellen*
Die Informationsquellen sollten ausreichend detailliert beschrieben werden, so dass die Adäquatheit der Informationen eingeschätzt werden kann.
- G.5. Valide und reliable Information (a) bzw. Glaubwürdigkeit (b), Übertragbarkeit (c) und Verlässlichkeit (d)*
(a) Die Verfahren zur Erhebung von Informationen sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit und die Zuverlässigkeit der erhaltenen Informationen für den gegebenen Zweck sichergestellt sind.
(b) Die Rekonstruktionen durch die Evaluation sollten eine korrekte Abbildung der Konstruktionen der Akteure darstellen.
(c) Eine präzise, detaillierte und vollständige Beschreibung (thick description) des untersuchten Kontextes sollte es erlauben, zu entscheiden, ob die Aussagen der Evaluation auf andere Kontexte übertragbar sein könnten.
(d) Der Evaluationsprozess sollte von aussen erkennbar sein, damit auch die getroffenen Entscheidungen und Interpretationen in der Evaluation untersucht und beurteilt werden können.
- G.6. Systematische Datenüberprüfung*
Die in einer Evaluation gesammelten, bearbeiteten und angegebenen Daten sollten überprüft und korrigiert werden, so dass die Resultate der Evaluation nicht fehlerhaft sind.
- G.7. Analyse qualitativer und quantitativer Information*
Qualitative und quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, so dass die Fragen der Evaluation effektiv beantwortet werden.
- G.8. Begründete Folgerungen*
Die in einer Evaluation erfolgten Folgerungen sollten explizit begründet sein, so dass die Beteiligten und Betroffenen diese einschätzen können.

- G.9. *Unvoreingenommene Berichterstattung (a) bzw. Bestätigbarkeit (b)*
- (a) Die Evaluationsverfahren sollten über Vorkehrungen verfügen, um die Ergebnisse und Berichte einer Evaluation gegen Verzerrungen durch persönliche Gefühle und Vorlieben irgendeiner Partei in der Evaluation zu schützen.
 - (b) Die Konstruktionen der Evaluation sollten auf ihre Ursprünge und Wurzeln im empirischen Material auch von aussen rückführbar sein.
- G.10. *Meta-Evaluation*
- Die Evaluation selbst sollte formativ und summativ mit den vorliegenden oder anderen Kriterien evaluiert werden, so dass die Durchführung entsprechend angeleitet werden kann und damit die Beteiligten und Betroffenen bei Abschluss der Evaluation ihre Stärken und Schwächen gründlich überprüfen können.

7.2. Folgerungen für die Ausgestaltung von Evaluationen

Die empirische Analyse hat ergeben, dass die zehn Evaluationen in den Bereichen der Nützlichkeit, der Anwendbarkeit sowie der Korrektheit deutlich bessere Leistungen erreichen, als im Bereich der Genauigkeit (vgl. zu den detaillierten Resultaten die Kapitel 6.1., 6.2. und vor allem 6.3.). Besondere Stärken ergeben sich bei Kriterien,

- die sich mit dem Verhältnis zwischen Auftraggeber und Evaluator befassen ('Rechtzeitigkeit des Berichts', 'Formale Verpflichtung'),
- die auf die Umsetzbarkeit und Zweckmässigkeit des Vorgehens in einer Evaluation abzielen ('Informationsrahmen und -auswahl', 'Praktische Verfahren') und
- die sich auf die Berichterstattung einer Evaluation beziehen ('Verbreitung des Berichts', 'Rechtzeitigkeit des Berichts', 'Herstellung von Öffentlichkeit', 'Ausgewogene Berichterstattung').

Zu den eindeutigen Stärken der untersuchten Studien sind aber auch die 'Glaubwürdigkeit des Evaluators' und der 'Schutz individueller Menschenrechte' zu zählen. In allen diesen Gebieten erreichen die Evaluationen - sieht man von punktuellen Ausnahmen ab - eine sehr hohe Qualität (siehe dazu ausführlicher Kapitel 6.3.1.). Daraus ergibt sich die folgende Empfehlung:

Empfehlung 1: Es ist darauf zu achten, dass die Stärken in den Bereichen der Nützlichkeit, der Anwendbarkeit sowie der Korrektheit auch in Zukunft erhalten bleiben oder sogar noch weiter ausgebaut werden können.

Im Bereich der Genauigkeit sind im Gegensatz dazu einige erhebliche Schwächen festgestellt worden. Besonders negativ sind die Bewertungen

- im Bereich der Gütekriterien empirischer Sozialwissenschaften, also 'Valide Messung' (bzw. 'Glaubwürdigkeit' und 'Übertragbarkeit'), 'Reliable Messung' (bzw. der 'Verlässlichkeit'), 'Verlässliche Informationsquellen' sowie 'Analyse quantitativer Informationen' und
- bei der Vermittlung der empirischen Analyse ('Beschreibung von Zielen und Vorgehen', 'Gerechtfertigte Folgerungen')

ausgefallen (vgl. die detaillierteren Angaben in Kapitel 6.3.2.). Diese Schwächen sind nicht bei allen untersuchten Evaluationen zu beobachten, sie sind jedoch im Rahmen der Fallstudien ausgesprochen häufig anzutreffen.

Aufgrund dieses Befundes ergibt sich für eine zukünftige Verbesserung der Evaluationspraxis ein deutlicher Ansatzpunkt. Gestützt auf die Erfahrungen aus den zehn Fallstudien bestehen für die angeführten Schwächen im Bereich der Genauigkeit verschiedene Ursachen, auf die wir im nachfolgenden eintreten werden.

Evaluationen werden üblicherweise unter einer Reihe von Rahmenbedingungen durchgeführt, die den Handlungsspielraum des Evaluators unterschiedlich stark einschränken können. Im Rahmen der Meta-Evaluation sind wir verschiedentlich auf Evaluationen gestossen, bei denen der zur Verfügung stehende Handlungsspielraum deutlich zu eng definiert war, so dass keine ausgeglichene Zweck-Mittel-Relation bestand. Unter den wichtigsten Rahmenbedingungen sind unter anderem zu nennen:

- die eng begrenzten finanziellen und zeitlichen Ressourcen,
- die ausgesprochen hohen Erwartungen des Auftraggebers hinsichtlich des Leistungspotentials einer Evaluation,
- das in vielen Politikbereichen fehlende oder mangelhafte Grundlagenwissen über soziale und politische Strukturen und Prozesse,
- die als überaus problematisch zu bezeichnende Datenlage (nicht nur, aber vor allem auch im quantitativen Bereich),
- die teilweise äusserst hohe Komplexität der Untersuchungsgegenstände, die eine fundierte Evaluation erschweren sowie
- analytisch ausserordentlich schwierig zu behandelnde, und zumeist sehr weite Fragestellungen

Diese Umstände führten bei einem grossen Teil der untersuchten Studien zu (zum Teil massiven) Kostenüberschreitungen, die zwar in einigen Fällen teilweise vom Auftraggeber, zumeist jedoch vor allem oder ausschliesslich vom Evaluator getragen werden mussten. Neben dieser Folge der äusserst schwierigen Umstände, in denen Evaluationen oft entstehen, sehen wir die oben angeführten Schwachpunkte im Bereich der Genauigkeit als weitere Konsequenz. Der Bereich der Genauigkeit ist bei der Abwicklung einer Evaluation deutlich weniger stark einer Kontrolle ausgesetzt, als dies bei Nützlichkeit, Anwendbarkeit oder Korrektheit der Fall ist. Deshalb werden in diesem Bereich wohl eher Mängel in Kauf genommen. Im Bereich der Genauigkeit manifestiert sich die teilweise Überstrapazierung des Leistungspotentials bei gegebenen Ressourcen. Im Gegensatz zu den anderen drei Bereichen, in denen sich die Artikulation der Interessen des Auftraggebers aber auch anderer Beteiligter auf den Evaluationsprozess auswirken, bestehen hier kaum Mechanismen, die zu einer Qualitätssicherung beitragen könnten. Kommt hinzu,

dass Evaluatoren sich mit der Anforderung konfrontiert sehen, auch dann handlungsrelevante Empfehlungen zu formulieren, wenn dies aufgrund der wissenschaftlichen Regeln im konkreten Fall nicht zulässig ist. Wie die empirische Analyse gezeigt hat, kann der Evaluator in dieser Situation dazu neigen, eher die Erwartungen der Praxis zu erfüllen und weniger auf eine genaue Einhaltung der in der Kriteriengruppe 'Genauigkeit' enthaltenen Anforderungen zu achten. Einen weiteren Grund für den Befund sehen wir darin, dass die Autorität des Evaluators im Verhandlungsprozess mit dem Auftraggeber tendenziell eher schwach ausgebildet ist.

Die nachfolgenden Empfehlungen sind darauf ausgerichtet, den Handlungsspielraum der Evaluatoren zu erweitern, um damit eine Qualitätsverbesserung zu erreichen.

Verschiedentlich sind wir bei der Durchführung der zehn Fallstudien auf die Situation gestoßen, dass die finanzielle und zeitliche Planung von Evaluationen nicht aufgrund der für die Durchführung der Evaluation erforderlichen Arbeiten, sondern vielmehr in Hinblick auf die zur Verfügung stehenden Mittel erfolgte. Dies hatte in mehreren Fällen dazu beigetragen, dass die finanziellen wie zeitlichen Rahmenbedingungen nicht den anforderungsreichen Fragestellungen und den aufwendigen empirischen Arbeiten entsprach. Daraus ergibt sich die folgende Empfehlung:

Empfehlung 2: Ausgangspunkt für die Planung und Konzipierung einer Evaluation sollte vermehrt die Fragestellung und die Thematik bzw. die für deren Bearbeitung erforderlichen Ressourcen bilden.

Die finanzielle und zeitliche Planung einer Evaluation sollte primär aufgrund der Fragestellung und der vorgesehenen Vorgehensweise und dem damit in Zusammenhang stehenden Aufwand erfolgen. Sollte das auf diese Weise entwickelte Auftragsvolumen die Möglichkeiten übersteigen, ist die Fragestellung und die Abgrenzung des Untersuchungsgegenstandes zu überprüfen und entsprechend anzupassen. Sowohl bei zeitlichen, wie auch bei finanziellen Anpassungen ist darauf zu achten, dass der Untersuchungsbe- reich nicht in einer Weise beschnitten wird, die eine ausgewogene und faire Evaluation verhindern würde. Entspricht das Auftragsvolumen nicht den zu leistenden Arbeiten, ist damit zu rechnen, dass die Evaluation nicht die vorgesehene Thematik bearbeiten kann und / oder - und dies ist aufgrund unserer Untersuchung häufig der Fall - dass die wissenschaftliche Qualität der Evaluation nicht optimal ausfällt.

Empfehlung 3: Wenn das Auftragsvolumen zeitlich und finanziell definiert ist, muss die inhaltliche Dimension in vernünftiger und ausgewogener Weise an die vorhandenen Ressourcen angepasst werden.

Evaluationsstudien sind, auch wenn erprobte Verfahren zur Anwendung gelangen, wie jede andere empirische Forschungstätigkeit auch, einer Vielzahl von Einflussfaktoren ausgesetzt, die ihre Umsetzung in erheblichem Umfang tangieren können. Unter anderem besteht zumindest vor Beginn einer Evaluation kaum eine Möglichkeit, die Frage des Datenzugangs abzuklären, die Qualität der zur Verfügung stehenden Daten zu überprüfen oder die notwendig werdenden Datenanalyseverfahren zu bestimmen und bezüglich ihres Aufwands gründlich zu untersuchen. Weiter ist davon auszugehen, dass in vielen Fällen im Verlauf des Evaluationsprozesses Erkenntnisse gewonnen werden, die zusätzliche Abklärungen erforderlich machen, oder dass es zu Veränderungen im Kontext der Evaluation kommt, die neue Fragestellungen aufkommen lassen oder die Bearbeitung der Evaluation behindern können. Auch wenn derartige Phänomene häufig nicht vorausgesehen werden können, sollte zumindest versucht werden, alle voraussehbaren Aspekte in die Planung zu integrieren:

Empfehlung 4: Schon vor Beginn einer Evaluation sollte versucht werden, die potentiellen Schwierigkeiten bei der Durchführung einer Evaluation so weit wie möglich zu antizipieren und in die Planung einzubeziehen.

Zur Vermeidung von übermässigen Kostenüberschreitungen und zur Sicherung einer qualitativ hochstehenden Durchführung von Evaluationen kann auch die nachfolgende Empfehlung, die sich vor allem auch an die Auftraggeber von Evaluationen richtet, beitragen:

Empfehlung 5: Die Budgetierung einer Evaluation sollte eine ausreichend grosse Reserve für Unvorhergesehenes beinhalten, die in begründeten Fällen Kostenüberschreitungen auffangen kann.

Im Rahmen des Aushandlungsprozesses zwischen Auftraggeber und Evaluator kam es in den von uns untersuchten Studien verschiedentlich zu überhöhten Erwartungshaltungen gegenüber der Evaluation. Da dies nach Beginn der Evaluation nur noch sehr schwer bereinigt werden kann, sollte folgende Empfehlung beachtet werden:

Empfehlung 6: Es sollte - im Sinne einer offenen Informationspolitik - bei der Aushandlung des Evaluationskontrakts vermehrt darauf geachtet werden, dass Klarheit über die Möglichkeiten und Grenzen der Evaluation besteht, so dass die Erwartungen an die Studie nicht überhöht ausfallen.

Die aufgrund des Aushandlungsprozesses festgelegte Ausgestaltung der Evaluation sollte in einem für alle beteiligten Parteien verbindlichen Evaluationskontrakt detailliert festgehalten werden. Einige der hier analysierten Evaluationsstudien sind auf der Basis eines Evaluationskontraktes erarbeitet worden, der diesen Erfordernissen nicht ausreichend nachkommt. Sollten im Verlauf des Aushandlungsprozesses verschiedene Offerten, Projektskizzen und andere schriftliche Unterlagen entstanden sein, ist davon abzuraten, diese Unterlagen pauschal als Bestandteile des Evaluationskontraktes zu deklarieren. Sehr häufig enthalten diese Unterlagen nämlich Angaben, die zu Unklarheiten Anlass geben können oder direkte Widersprüche darstellen. Weiterhin ist darauf zu achten, dass alle beteiligten Akteure die im Kontrakt enthaltenen Angaben mittragen. Zudem ist der Evaluationskontrakt in einer möglichst präzisen Form abzufassen, da die vertragliche Bindung damit an Klarheit gewinnt und Missverständnisse vermieden werden können.

Empfehlung 7: Der Evaluationskontrakt sollte eine unumstrittene, eindeutige und ausreichend präzise Basis für die Evaluation bilden. Es ist zu vermeiden, dass Evaluationskontrakte zu pauschal ausfallen, dass sie Widersprüche aufweisen oder dass sie Regelungen enthalten, die nicht von allen Beteiligten getragen werden.

Die Projektbegleitung durch den Auftraggeber - in der Form von Begleitgruppen oder in informeller Weise - nimmt bei Evaluationen häufig eine zentrale Rolle ein. Die Projektbegleitung erhält dabei unterschiedliche Funktionen. In erster Linie geht es dabei darum, die Evaluation im Hinblick auf die Erfordernisse der Praxis abzustimmen. Zweitens nimmt die Begleitgruppe oft eine Kontrollfunktion wahr, indem sie den Projektverlauf verfolgt und gegebenenfalls interveniert. Drittens spielt oft auch der Aspekt eine Rolle, dass die beteiligten oder betroffenen Kreise mit der Projektbegleitung in den Evaluationsprozess eingebunden werden können. Oft bildet dabei die Begleitgruppe ein wichtiges Forum des Interessenausgleichs zwischen den Beteiligten. Viertens kann sie dazu dienen, den Informationszugang für die Evaluatoren zu erleichtern. Fünftens ist daran zu denken, dass die Begleitgruppe sehr oft auch das Medium ist, durch welches die Evaluation an einen weiteren Kreis von Beteiligten oder Betroffenen kommuniziert wird. Sie erhält damit auch eine wichtige Rolle bei der Umsetzung der Evaluation.

Aufgrund der in der empirischen Analyse konstatierten Schwachpunkte der Evaluationen im Bereich der Genauigkeit, sollte vermehrt auch evaluatives Know-how in begleitende Gremien integriert werden. Damit könnte das Gewicht methodischer Aspekte bei der Begleitung der Projekte verstärkt werden. Dabei besteht die Absicht aber keineswegs darin, eine methodische Kontrollinstanz zu generieren. Vielmehr soll die Begleitgruppe nicht ausschliesslich als Forum für die praxisrelevanten Aspekte dienen, sondern auch die wissenschaftlichen und vor allem methodischen Fragen thematisieren. Dadurch, dass seitens des Auftraggebers nicht nur die praxisrelevanten Aspekte vertreten sind, sondern auch Experten mit evaluations-spezifischen Kompetenzen, können die Begleitfunktionen auf die evaluations-technischen Themen ausgeweitet werden. Dadurch kann eine umfassende Projektbegleitung sichergestellt werden. Diese Überlegungen führen zur nachfolgenden Empfehlung:

***Empfehlung 8: Die Qualitätssicherung kann verbessert werden, indem Evaluations-
experten als Mitglieder von Begleitgruppen in den Evaluationsprozess integriert
werden.***

Hinsichtlich der Ausgestaltung der methodischen Berichterstattung zu Evaluationen - gemäss unserer Analyse eine der hauptsächlichen Schwächen der untersuchten Evaluationsstudien (vgl. Kapitel 6.2.4. und 6.3.2.) - drängen sich wesentliche Verbesserungen auf. Eine transparente und nachvollziehbare Darstellung der im Rahmen der Evaluation durchgeführten Analyseschritte kann verschiedene Funktionen erfüllen:

- Durch eine gelungene methodische Berichterstattung gewinnen die Folgerungen der Evaluation an Glaubwürdigkeit und Akzeptanz.
- Die Evaluation kann, wenn eine ausreichende methodische Berichterstattung vorliegt, besser vermittelt werden.
- Nur eine nachvollziehbare methodische Berichterstattung erlaubt eine fundierte Methodenkritik, die zur Qualitätskontrolle dienen kann. Eine fehlende oder nicht-ausreichende methodische Berichterstattung kann eine Evaluation vor (möglicherweise berechtigter) Kritik immunisieren.
- Ein nachvollziehbarer methodischer Bericht erlaubt es den nicht an der Evaluation beteiligten Personen (etwa anderen Evaluatoren), von den Erfahrungen aus der Evaluation zu lernen. Dies vermag auch einen Beitrag zur Weiterentwicklung der Evaluationsmethodologie zu leisten.

- Weiterhin bildet die sorgfältige Formulierung des Methodik-Berichts Gelegenheit zur Selbstreflexion (und Selbstkritik) und kann zur Vermeidung von Über- und Fehlinterpretationen beitragen.

Die Forderung nach einer Verbesserung der Berichterstattung im methodischen Bereich bedeutet keinesfalls, dass eine technische, für die potentiellen Adressaten oft schwer verständliche und damit praxisfremde Ausrichtung von Evaluationsberichten bevorzugt wird. Vielmehr besteht die Herausforderung darin, eine Berichterstattung zu gewährleisten, die den unterschiedlichen Bedürfnissen angepasst ist. Neben der Abfassung von adressatenspezifischen Berichten ist dabei auf die Möglichkeiten von aussagekräftigen Methodikteilen (evtl. im Anhang des Berichts) hinzuweisen.

Aus diesen Überlegungen ergibt sich die folgende Empfehlung:

Empfehlung 9: Die Berichterstattung von Evaluationen sollte vermehrt darauf ausgerichtet werden, dass die vorgenommenen Analysen in transparenter und nachvollziehbarer Form dargestellt werden.

Zur Verstärkung der Bedeutung der wissenschaftlichen Qualitätskriterien in Evaluationen eignet sich unseres Erachtens eine weitere Massnahme, die weiter unten (siehe Kapitel 7.3.) noch detaillierter diskutiert wird:

Empfehlung 10: Zur Qualitätssicherung in der Evaluationsforschung sind vermehrt Meta-Evaluationen durchzuführen.

Neben dem teilweise stark eingeschränkten Handlungsspielraum des Evaluators sehen wir eine weitere Ursache für die Schwächen im Bereich der 'Genauigkeit' darin, dass verschiedene Voraussetzungen nicht gegeben sind, um die hohen analytischen Ansprüche, die an die Evaluationsforschung gestellt werden, zu erfüllen.

Im Rahmen der empirischen Fallstudien sind verschiedentlich Mängel festgestellt worden, die durch eine speziell auf die Bedürfnisse von Evaluationen zugeschnittene Aus- oder Weiterbildung ausgeräumt werden können. Die akademische, disziplinar orientierte Ausbildung bildet zwar eine durchaus geeignete Grundlage für die Tätigkeit des Evaluators. Sie vermag den vielfältigen Ansprüchen jedoch keineswegs zu genügen. Es besteht zwar in Einzelfällen die Möglichkeit, durch eine zweckmässige (multidisziplinäre) Zusammenstellung eines Evaluatorenteams, die Nachteile einer uni-disziplinären Ausrich-

tung abzuschwächen. Vielfach ist dies jedoch aufgrund gegebener Restriktionen nicht oder nur eingeschränkt möglich.

Die nachfolgende Massnahme kann unseres Erachtens dazu beitragen, die Voraussetzungen massgeblich zu verbessern:

Empfehlung 11: Die Einrichtung interdisziplinär ausgerichteter, evaluationsspezifischer Lehrgänge in der Aus- und Weiterbildung leisten einen wichtigen Beitrag zur Verbreiterung und Vertiefung der Wissensbasis.

In den letzten Jahren ist der Begriff 'Evaluation' sehr häufig eingesetzt worden. Auch unter den von uns untersuchten zehn Studien befinden sich Projekte, die unseres Erachtens nicht als Evaluationen zu bezeichnen sind. Daraus ergibt sich die nachstehende Empfehlung.

Empfehlung 12: Der Begriff 'Evaluation' sollte nicht inflationär für verschiedenste Typen von Auftragsverhältnissen (Beratung, Organisationsentwicklung, Begleitforschung etc.) verwendet werden.

Im Laufe unserer Abklärungen zu den zehn Fallstudien sind wir verschiedentlich auf Situationen gestossen, bei denen Evaluationen bei der Informationsbeschaffung mit grossen Problemen konfrontiert waren. Neben dem - teilweise nur schwer zu erreichenden - Zugang zu amtlichen Dokumenten standen dabei einer effizienten Informationsbeschaffung vor allem die Verfügbarkeit und Qualität von Datenmaterialien der amtlichen Statistik oder der vollziehenden öffentlichen Stellen entgegen. Weiterhin erschwert wird die Datenbeschaffung auch durch das vielfach fehlende Öffentlichkeitsprinzip. In diesem Bereich besteht dringlicher Handlungsbedarf seitens der öffentlichen Hand. Von Verbesserungen in diesen Bereichen lassen sich auch anderweitig positive Auswirkungen erwarten. Neben der Verbesserung der Möglichkeiten zur internen Vollzugsüberprüfung sei etwa auf eine erhöhte Transparenz der öffentlichen Verwaltung und einer damit verbundenen Steigerung der Glaubwürdigkeit und Akzeptanz derselben hingewiesen.

Aus diesen Überlegungen ergibt sich die folgende, bereits vielfach geäusserte Empfehlung:

Empfehlung 13: Zur Erhöhung des Leistungspotentials von Evaluationen sind seitens der öffentlichen Hand Massnahmen zu treffen, die den Aufwand für Evaluationsprojekte massgeblich reduzieren. Im Vordergrund stehen dabei die Forderungen nach einer (weiteren) Verbesserung der Datenlage im Rahmen der amtlichen Statistik aber auch bei der Erhebung und Archivierung von Prozessdaten bei den vollziehenden Stellen sowie die generelle Etablierung des Öffentlichkeitsprinzips bei Bund, Kantonen und Gemeinden.

Im Hinblick auf die zukünftige schweizerische Evaluationspraxis sollte ein Forum geschaffen werden, das Impulse zur Weiterentwicklung der Evaluation, zur Aus- und Weiterbildung von Evaluatoren sowie zur Professionalisierung des Evaluationsbereichs geben kann. Dabei steht unseres Erachtens die Gründung einer Schweizer Evaluationsgesellschaft im Vordergrund. Eine solche Organisation ist in der Lage, eine massgebliche Verbesserung der Voraussetzungen der schweizerischen Evaluationsforschung herbeizuführen. Die Etablierung einer Evaluationsgesellschaft kann dazu beitragen, den interdisziplinären Austausch unter Evaluatoren zu verbessern. Eine solche Organisation erlaubt es zudem, eine breit abgestützte Formulierung von Qualitätsmassstäben zu realisieren, um so die Qualitätssicherung sowie -verbesserung zu erleichtern. Weiterhin sind von einer derartigen Organisation auch Anstösse zur Klärung, Abgrenzung und Verfestigung des Selbstverständnisses der 'Evaluationsgemeinde' zu erwarten. Zur Erreichung dieser Zielsetzung ist auf eine möglichst breite Abstützung der Vereinigung zu achten. Eine Zusammenarbeit mit anderen Evaluationsgesellschaften, wie der 'European Evaluation Society' (EES) oder der 'American Evaluation Association' (AEA) würde die Erreichung der oben genannten Ziele erleichtern.

Die Empfehlung dazu lautet:

Empfehlung 14: Es sollte eine breit abgestützte und international vernetzte Schweizer Evaluationsgesellschaft gegründet werden.

7.3. Zur Durchführung von Meta-Evaluationen

In diesem Kapitel werden die bei der Durchführung der Meta-Evaluation gewonnenen Erfahrungen resümiert und im Hinblick auf die zukünftige Durchführung von Projekten dieser Art ausgewertet.

Die in Kapitel 4 erläuterte Vorgehensweise hat sich grundsätzlich bewährt. Die Meta-Evaluation führte zu Ergebnissen, die unseres Erachtens von wissenschaftlichem wie praktischem Wert sind. Die Aussagekraft ist zwar aufgrund der Vorgehensweise bei der Bestimmung der Stichprobe und wegen deren Umfang beschränkt. Wir sind jedoch der Ansicht, dass bei einer zurückhaltenden Interpretation, um die wir uns speziell bemüht haben, durchaus haltbare und nützliche Informationen gewonnen werden konnten.

Der Projektablauf konnte nicht in allen Teilen, wie dies ursprünglich vorgesehen war, realisiert werden. Es ergaben sich folgende Änderungen:

- Schon bald nach Beginn der empirischen Arbeiten an den Fallstudien zeigte sich, dass die aufgrund einer Analyse der schriftlichen Unterlagen erhobenen Informationen in keiner Weise ausreichen, um eine Bewertung entlang des Kriterienrasters vorzunehmen. Um den Informationsstand zu den einzelnen Evaluationen zu ergänzen, zu verdichten und besser abzustützen, wurde zu jeder Fallstudie eine Reihe von Gesprächen geführt (vgl. dazu ausführlich die Hinweise in Kapitel 4.3.).
- Die Erarbeitung der zehn Fallstudien hat sich deswegen und weil sich die Erhebung und Aufarbeitung des Materials zu den einzelnen Fallstudien als aufwendiger erwies, als dies zu Beginn angenommen worden war, stark ausgedehnt.
- Im Verlauf der Arbeiten an den empirischen Fallstudien wurde deutlich, dass es zweckmässig und wohl auch notwendig ist, die Autoren der untersuchten Evaluationen in den Meta-Evaluationsprozess zu integrieren, indem sie zu einer Stellungnahme zur Fallstudie, die sich mit ihrem Projekt befasst, eingeladen werden. Der Vernehmlassungsprozess erwies sich als ziemlich arbeitsintensiv. Zudem wurde nach dem Eingang der ersten Reaktionen deutlich, dass es unumgänglich sein wird, den Vernehmlassungsprozess (zumindest) zweistufig auszugestalten (vgl. dazu Kapitel 4.5.).
- Im Juli 1994 erschien die neue Version der als Grundlage eingesetzten 'Program Evaluation Standards' (vgl. dazu Kapitel 3.6.1.). Die neue Fassung konnte zwar nicht mehr für die Ableitung der in den empirischen Fallstudien eingesetzten Kriterien verwendet werden, weil die Projektarbeiten zu diesem Zeitpunkt schon zu weit vorangeschritten waren. Sie musste jedoch in die Untersuchung mit einfließen, da ansonsten unsere Arbeiten deutlich an Relevanz verloren hätten.

- Durch die genannten Punkte wurde der Aufwand deutlich erhöht. Um das Projekt trotzdem innert nützlicher Frist abschliessen zu können, musste ein Untersuchungsteil fallengelassen werden. Es war nicht möglich, die geplante Vertiefung der Fallstudien für zwei Evaluationen vorzunehmen. Auf diesen Untersuchungsteil musste vollständig verzichtet werden.

Selbstkritisch muss an dieser Stelle angeführt werden, dass die Projektplanung zu wenig auf die mit der Untersuchung verbundenen Risiken und Schwierigkeiten Rücksicht genommen hatte. Ein Teil der zusätzlichen Arbeiten konnte jedoch in keiner Weise vorausgesehen werden (etwa die Neuerscheinung der Standards).

Empfehlung 15: Die Planung von Meta-Evaluationen sollte (wie bei Evaluationen im allgemeinen) vorsichtig und sensitiv gegenüber potentiellen Schwierigkeiten vorgenommen werden.

Wie oben dargelegt, überstieg der entstehende Aufwand zur Durchführung der Meta-Evaluation die Erwartungen. Durch geeignete Massnahmen könnte der Aufwand reduziert werden. In Frage käme dafür beispielsweise ein erleichterter Datenzugang und eine verbesserte Datenverfügbarkeit für die Meta-Evaluation. Dies könnte beispielsweise dadurch erreicht werden, dass die im Evaluationsprozess anfallenden Informationen umgehend systematisch erfasst und aufgearbeitet werden (vgl. dazu die in Schwandt/Halpern 1988 vorgeschlagenen Verfahren). Die gesammelten Informationen sollten eine problemlose Rekonstruktion des Evaluationsprozesses erlauben. Dabei sind nicht nur die im Rahmen der Evaluation vorgenommenen Datenerhebungen und Datenauswertungen laufend zu dokumentieren, sondern auch die weiteren Aktivitäten im Zusammenhang mit der Evaluation (Aussenkontakte des Evaluationsteams, Begleitgruppensitzungen etc.). Dies erfordert natürlich zusätzliche Arbeiten seitens der Evaluatoren wie auch seitens der Auftraggeber. Da davon aber nicht nur die Durchführung von Meta-Evaluationen profitieren kann, sondern vor allem auch positive Auswirkungen im Bereich der methodischen Selbstreflexion sowie des Projektmanagements zu erwarten sind, erachten wir dieses Vorgehen als sinnvoll.

Empfehlung 16: Die im Evaluationsprozess anfallenden Informationen sollten umgehend systematisch erfasst und aufgearbeitet werden.

Zu einer erleichterten Durchführung von Meta-Evaluationen würden auch die in Empfehlung 13 formulierten Massnahmen (Verbesserung der Datengrundlage, Öffentlichkeitsprinzip) beitragen (siehe dazu Kapitel 7.2.).

Weiterhin ist es unseres Erachtens zweckmässig, in Evaluationskontrakten routinemässig eine Meta-Evaluationsklausel einzubauen. Diese würde den formellen Parteien einer Evaluation das Recht (nicht die Pflicht!) geben, die Evaluation (auf eigene Kosten) nachträglich einer externen Meta-Evaluation zu unterziehen. Diese Massnahme würde einerseits dem Auftraggeber ein Instrument in die Hand geben, um Qualitätsüberprüfungen vorzunehmen. Andererseits könnte der Auftraggeber auch davon ausgehen, dass sich die Qualität der Evaluation - auch ohne Realisation einer Meta-Evaluation - verbessern könnte. Die Position des Evaluators wiederum könnte so gestärkt werden, weil er die Möglichkeit erhalten würde, übermässig stark eingeschränkte Handlungsspielräume bei der Durchführung von Evaluationen offenzulegen. Die anfallenden Kosten würden die Zahl der durch diese Klausel ausgelösten Meta-Evaluationen stark beschränken, womit die Verhältnismässigkeit gewahrt sein sollte. Dies führt zu einer weiteren Empfehlung:

Empfehlung 17: Evaluationskontrakte sind routinemässig mit einer Meta-Evaluationsklausel zu versehen.

Meta-Evaluationen enthalten - wie alle Evaluationen - einen Ermessensspielraum bei der Beurteilung. Dies zeigte sich auch bei der Durchführung der vorliegenden Studie deutlich. Umso wichtiger wird deshalb, dass Meta-Evaluationen systematisch erfolgen. Wir haben uns bei dieser Studie eines Kriterienrasters bedient, um den Ermessensspielraum einzuschränken. Der Kriterienraster hat sich grundsätzlich bewährt (vgl. Kapitel 7.1.).

Empfehlung 18: In Meta-Evaluationen sollte eine systematische Vorgehensweise gewählt werden, um den Ermessensspielraum zu begrenzen. Der hier eingesetzte Kriterienraster stellt dazu eine geeignete - aber keinesfalls die einzig mögliche - Lösung dar.

Der von uns eingesetzte Kriterienraster stiess bei den beteiligten Evaluatoren nicht immer auf Zustimmung. So wurde kritisiert, dass der Kriterienraster (oder einzelne Kriterien) in bestimmten Kontexten nicht adäquat oder dass die einzelnen Kriterien je nach den bestehenden Rahmenbedingungen anders zu interpretieren oder zu gewichten seien. Weiterhin wurde die Kritik geäussert, dass der Kriterienraster zur Zeit der Evaluation nicht bekannt war und dass es schwierig sei, sich nach unbekanntem Zielsetzungen zu orientieren. Diese Kritik ist unseres Erachtens nur teilweise berechtigt. Währenddem der Einsatz der Kriterien sicherlich immer kontextspezifisch erfolgen sollte und es durchaus auch Situationen geben kann, in denen ein Kriterium nicht anwendbar ist, sollte nur in begründeten Fällen auf die Verwendung einzelner Kriterien verzichtet werden. Wir sind der Ansicht, dass es durchaus Situationen geben kann, in denen ein Kriterium (etwa aufgrund zu enger Re-

striktionen) nicht erfüllbar ist. Gerade diese Konstellationen zu erkennen, bildet aber ein wichtiges Ergebnis von Meta-Evaluationen. Wir haben die Auswahl des hier eingesetzten Kriterienrasters ausführlich begründet (vgl. Kapitel 3.2. und 3.3.) und sind der Ansicht, dass er für die Bearbeitung der vorliegenden Fragestellung geeignet ist. Die Problematik, dass der Kriterienraster zur Zeit der Planung und Durchführung der Evaluationen nicht bekannt war, sollte mit der weiteren Verbreitung dieser oder anderer Qualitätskriterien in Zukunft an Bedeutung verlieren. Bezüglich der Gewichtung der Einzelkriterien vertreten wir die Auffassung, dass eine solche nicht sachgerecht wäre. Aufgrund der stark unterschiedlichen Charakteristiken der einzelnen Kriterien, die sich auf deutlich voneinander zu unterscheidende Inhalte beziehen und im spezifischen Kontext einer Evaluation jeweils eine differierende Bedeutung erhalten können, ist eine allgemeingültige, übergreifende Festlegung ihrer Bedeutung kaum realisierbar und wohl auch nicht sachgerecht. Trotz diesen Überlegungen sind wir der Ansicht, dass es erstrebenswert ist, die zur Bewertung einzusetzenden Kriterien im Einvernehmen mit den Beteiligten zu definieren. Die Gründe dafür liegen nicht nur in den oben angebrachten Einwänden (oder Einwänden anderer Art), sondern auch im Umstand, dass es uns in einem Fall nicht gelungen ist, die erkenntnistheoretische Orientierung einer Evaluation zu bestimmen (siehe Studie Nr. 9, Kapitel 5.3.3.). Dadurch ergab sich die Schwierigkeit, dass der Einsatz der entsprechenden Kriterien nicht möglich war. Durch den Einbezug der Beteiligten in die Auswahl der relevanten Kriterien könnten derartige Probleme vermieden werden.

Empfehlung 19: Über die Auswahl, der in einer Meta-Evaluation einzusetzenden Kriterien, ist mit den Beteiligten im voraus eine konsensuale Entscheidung anzustreben.

Aufgrund der Ergebnisse der Meta-Evaluation stellt sich die Frage, ob bei der Auswahl der zehn Evaluationen eine geeignete Auswahl getroffen wurde, kommen wir doch in zwei Fällen zum Resultat, dass diese Studien aufgrund der empirischen Analyse keine Evaluationen darstellen. Wir haben uns bei der Auswahl der zehn Evaluationen diesbezüglich durch die Selbstdeklaration leiten lassen. In beiden Fällen handelt es sich um Studien, die im Titel als Evaluationen bezeichnet werden. Zudem wurden beide Studien explizit als Evaluationen in Auftrag gegeben. Da jede Definition immer auch eine Ausgrenzung des 'Nichtmitgemeinten' mit sich bringt, sind wir der Ansicht, dass unsere Wahl trotz des Ergebnisses richtig war. Selbstverständlich lässt sich darüber diskutieren, ob die hier verfolgte (und im Kriterienraster implizit enthaltene) Definition von Evaluation zweckmässig ist.

Eine grosse Schwäche der vorliegenden Untersuchung bildet der Umstand, dass die Auswirkungen der zehn Evaluationen nicht ausreichend untersucht werden konnten, um diesbezüglich eine fundierte Bewertung zu erlauben. Dies führen wir einerseits auf die Komplexität der Wirkungsprozesse zu sozialwissenschaftlichen Wissensbeständen zurück. Andererseits erwies sich die gewählte Untersuchungsanlage in dieser Hinsicht nur beschränkt als geeignet. Da insgesamt zehn Evaluationsstudien zu untersuchen waren, konnte die Analyse im Bereich der Wirkungen nicht ausreichend vertieft werden. Die zur Bearbeitung einer der Fallstudien zur Verfügung stehenden finanziellen wie zeitlichen Ressourcen reichten keineswegs aus, um die Wirkungsproblematik ausreichend fundiert zu erfassen (vgl. Kapitel 6.1.1.8. und 7.1.2.). Wir mussten uns in den Fallstudien zumeist auf die Diskussion einiger Anhaltspunkte direkter Wissensverwendung konzentrieren. Die indirekte Nutzung der Ergebnisse, wie auch die möglicherweise auftretenden Aufklärungswirkungen (Weiss 1977) längerfristiger Art, blieben uns jedoch verborgen.

Neben den grundsätzlichen Einwänden konzeptioneller Art, die zur Wirkungserfassung sozialwissenschaftlicher Untersuchung bestehen (vgl. die Hinweise in Kapitel 6.3.3.), erfordert die Wirkungsanalyse deutlich grössere Mittel. Für die Wirkungserfassung einer Evaluationsstudie sind Mittel zur Verfügung zu stellen, die sich in einem ähnlichen Rahmen bewegen, wie dies für Evaluationen eines politischen Programms erforderlich ist. Zudem müsste ein anderer konzeptioneller Untersuchungsansatz eingesetzt werden, als dies hier der Fall war (vgl. zu den verschiedenen Möglichkeiten Weiss 1981). Dies weist relativ deutlich darauf hin, dass Meta-Evaluationen, die einen vertretbaren Aufwand erfordern, zumeist nicht in der Lage sein werden, eine fundierte Abklärung der Evaluationswirkungen zu leisten.

Empfehlung 20: Meta-Evaluationen bedürfen - sofern dies überhaupt möglich ist - grösserer Mittel, um eine fundierte Bewertung des Wirkungsaspekts vornehmen zu können.

Die hier vorliegende Meta-Evaluation wurde als reine ex post-Analyse angelegt. Es wurden in der Stichprobe lediglich Evaluationen berücksichtigt, die bei Projektbeginn bereits abgeschlossen waren. Es stellt sich nun die Frage, ob es zweckmässiger wäre, eine Meta-Evaluation begleitend durchzuführen. Dies würde nicht nur den Vorteil mit sich bringen, dass die formative Komponente der Meta-Evaluation verstärkt (und eventuell auch gleich handlungsrelevant) werden könnte. Es ist auch davon auszugehen, dass dadurch auch der Datenzugang allgemein und im speziellen in bestimmten Aspekten (etwa hinsichtlich des Kriteriums 'Menschliche Interaktionen') erleichtert und verbessert werden würde. Weiterhin könnte sich dadurch auch die Akzeptanz der Meta-Evaluation bei den beteiligten

Akteuren erhöhen. Andererseits ist aber auch darauf hinzuweisen, dass eine begleitende Meta-Evaluation unweigerlich auch Nachteile mit sich bringt. An erster Stelle ist auf den Umstand hinzuweisen, dass sich in einer begleitenden Meta-Evaluation die Beschaffung von Informationen zum Wirkungsbereich (weiter) erschweren könnte (etwa bei den Kriterien 'Wirkung des Berichts', 'Politische Überlebensfähigkeit' und 'Kostenwirksamkeit'). Weiter ist damit zu rechnen, dass der Aufwand für eine begleitende Meta-Evaluation höher ausfallen würde. Zudem ist davon auszugehen, dass begleitende Meta-Evaluationen einen direkten Einfluss auf die Ausgestaltung der Evaluation ausüben. Dies ist einerseits von Vorteil, kann sich doch daraus eine Qualitätsverbesserung ergeben. Andererseits ist damit aber auch eine Veränderung der summativen Ergebnisse der Evaluation verbunden. Hieraus ergibt sich folgende Empfehlung:

Empfehlung 21: Sowohl ex post- wie auch begleitende Meta-Evaluationen verfügen über ihre spezifischen Vor- und Nachteile. Ex post-Meta-Evaluationen sind vor allem zur Bearbeitung summativer Fragestellungen geeignet. Begleitende Meta-Evaluationen sind dagegen bei formativen Fragestellungen vorzuziehen.

Die vorliegende Untersuchung ist an einem politikwissenschaftlichen Universitätsinstitut entstanden und befasst sich mit Evaluationen aus dem privatwirtschaftlichen und dem universitären Kontext. Sie evaluiert nicht nur Evaluationen, die von Evaluatoren mit politikwissenschaftlichem Hintergrund durchgeführt wurden, sondern auch Studien von Autoren mit einer anderen disziplinären Herkunft. Diese Konstellation traf nicht immer auf positive Reaktionen. Vor allem privatwirtschaftlich tätige Evaluatoren haben die Evaluation durch ein universitäres Institut nicht immer geschätzt. Auch Vorbehalte disziplinärer Art sind nicht auszuschliessen, auch wenn solche im Rahmen dieses Projektes nie explizit formuliert wurden. Um diese Situation zu verbessern ergibt sich für die zukünftige Durchführung vergleichbarer Studien die folgende Empfehlung:

Empfehlung 22: Die zur Durchführung von Meta-Evaluationen beizuziehenden Evaluatorenteams sollten über einen multidisziplinären, universitären wie privatwirtschaftlichen Hintergrund verfügen.

Meta-Evaluationen sollten aufgrund unserer Erfahrungen vermehrt eingesetzt werden. Sie verfügen über ein hohes Potential zur Sicherung und Verbesserung der Qualität von Evaluationen. Weiter geht von Meta-Evaluationen eine nicht unerhebliche präventive Wirkung aus. Verschiedenen Institutionen, die häufig Evaluationen in Auftrag geben (oder selber durchführen), steht mit der Meta-Evaluation ein geeignetes Instrument zur Qualitätskontrolle und Qualitätssicherung zur Verfügung. In den letzten Jahren sind in

der Schweiz deutlich mehr Evaluationen durchgeführt worden, als dies bisher üblich war. Auch wenn sich diese Evaluationen im staatlichen Bereich noch grösstenteils auf die Bundesebene bezogen, werden vermehrt auch auf kantonaler oder kommunaler Ebene Evaluationen durchgeführt. Für verschiedene Institutionen sind Evaluationen zu einem alltäglichen Instrument geworden. Anfängliche Vorbehalte konnten vielerorts abgebaut werden. Die Durchführung von Meta-Evaluationen ist dazu geeignet, diese Entwicklung zu unterstützen und kann einen Beitrag zu deren Konsolidierung erbringen.

Aus diesen Gründen kommen wir zur abschliessenden Empfehlung:

Empfehlung 23: Meta-Evaluation hat sich als geeignetes Instrument erwiesen, das zur weiteren Verwendung empfohlen wird.

8. ABKÜRZUNGSVERZEICHNIS

A	Anwendbarkeit
a.a.O.	am angegebenen Ort
Abb.	Abbildung
Abs.	Absatz
Abt.	Abteilung
AEA	American Evaluation Association
AFU	Amt für Umweltschutz
AG	Aktiengesellschaft
AGEVAL	Arbeitsgruppe 'Gesetzesevaluation'
AGV	Arbeitgeberverband
Amtl.	Amtlich
Anm.	Anmerkung
ANSI	American National Standards Institute
APS	Année Politique Suisse / Schweizerische Politik im Jahre ...
APSA	American Political Science Association
ARA	Abwasserreinigungsanlage
ArG	BG über die Arbeit in Industrie, Gewerbe und Handel, Arbeitsgesetz (SR 822.11)
Art.	Artikel
AS	Amtliche Sammlung des Bundesrechts
ASB	Bundesamt für Strassenbau
ASSP	Association suisse de science politique
BAP	Bundesamt für Polizeiwesen
BAV	Bundesamt für Verkehr
BAWI	Bundesamt für Aussenwirtschaft
BBl	Bundesblatt
BBW	Bundesamt für Bildung und Wissenschaft
bes.	besonders
BEW	Bundesamt für Energiewirtschaft
BfK	Bundesamt für Konjunkturfragen
BFL	Bundesamt für Forstwesen und Landschaftsschutz
BfS	Bundesamt für Statistik
BG	Bundesgesetz
BIGA	Bundesamt für Industrie, Gewerbe und Arbeit
BJ	Bundesamt für Justiz
BK	Schweiz. Bundeskanzlei
BLW	Bundesamt für Landwirtschaft
bspw.	beispielsweise
BUS	Bundesamt für Umweltschutz
BUWAL	Bundesamt für Umwelt, Wald und Landschaft
BV	Bundesverfassung der Schweizerischen Eidgenossenschaft (SR 101)
BVG	Gesetz über die berufliche Alters-, Hinterlassenen- und Invalidenvorsorge (SR 831.40)
bzw.	beziehungsweise
CES	Canadian Evaluation Society

CVP	Christlich-Demokratische Volkspartei
d.h.	das heisst
Diss.	Dissertation
div.	divers
éd(s).	éditeur(s)
ed(s).	editor(s)
EDA	Eidgenössisches Departement für auswärtige Angelegenheiten
EDI	Eidgenössisches Departement des Innern
EDMZ	Eidgenössische Druck- und Materialzentrale
EES	European Evaluation Society
ehem.	ehemalig
eidg.	eidgenössisch
EJPD	Eidg. Justiz- und Polizeidepartement
EK	Evaluationskontrakt
ENet	Evaluation Network
ERS	Evaluation Research Society
ESB	Entwurf zum Schlussbericht
ESRA	Evaluation Studies Review Annual
etc.	et cetera
ETH	Eidgenössische Technische Hochschule
EVD	Eidgenössisches Volkswirtschaftsdepartement
EVED	Eidgenössisches Verkehrs- und Energiewirtschaftsdepartement
EWI	Elektrowatt Ingenieurunternehmung
EWR	Europäischer Wirtschaftsraum
FAC	Forschungsanstalt für Agrikultur und Umwelthygiene
FDP	Freisinnig-Demokratische Partei
FG	Forschungsgesuch
Fig.	Figur
FSB	Finanzieller Schlussbericht
FWP	Forschungsstelle Wissenschaft und Politik
FZB	Finanzieller Zwischenbericht
G	Genauigkeit
GAO	United States General Accounting Office
GAV	Gesamtarbeitsvertrag
GES	Gesuch
GP	Grüne Partei
GPK	Geschäftsprüfungskommission
GS	Generalsekretariat
GVF	Dienst für Gesamtverkehrsfragen
HArG	BG über die Heimarbeit, Heimarbeitsgesetz (SR 822.31)
Hrsg.	Herausgeber
i.e.	id est
IDHEAP	Institut de hautes études en administration publique
inkl.	inklusive
insbes.	insbesondere
K	Korrektheit
KB	Kurzbericht

KIGA	Kantonales Amt für Industrie, Gewerbe und Arbeit
lit.	litera
LPS	Liberale Partei der Schweiz
LRV	Luftreinhalte-Verordnung (SR 814.318.142.1)
MDS	Multidimensionale Skalierung
Mio.	Million(en)
N	Nützlichkeit
NADEL	Nachdiplomstudiengang Entwicklungsländer
NFP	Nationales Forschungsprogramm
NFP 5	Nationales Forschungsprogramm 'Regionalprobleme'
NFP 22	Nationales Forschungsprogramm 'Boden'
NFP 25	Nationales Forschungsprogramm 'Stadt und Verkehr'
NFP 27	Nationales Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen'
NFP 31	Nationales Forschungsprogramm 'Klimaänderungen und Naturkatastrophen'
No.	number
NR	Nationalrat
Nr.	Nummer
NZZ	Neue Zürcher Zeitung
N°	nombre
OF	Offerte
OLS	Ordinary-least-squares
PC	Personal Computer
PL	Programmleitung
PVK	Parlamentarische Verwaltungskontrollstelle
PVS	Politische Vierteljahresschrift
Q	Beziehungsdichte
RERO	Réseau des bibliothèques romandes et tessinoise
Rz.	Randziffer
s.	siehe
SB	Schlussbericht
sFr.	Schweizer Franken
SGS	Schweizerische Gesellschaft für Soziologie
SIDOS	Schweizerischer Informations- und Daten-Archivdienst für die Sozialwissenschaften
SKUB	Schweizerische Kommission für Umweltbeobachtung
SNF	Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung
sog.	sogenannt
spez.	speziell
SPS	Sozialdemokratische Partei der Schweiz
SR	Systematische Sammlung des Bundesrechts
Stv., stv.	Stellvertreter(in), stellvertretend(er)
SVP	Schweizerische Volkspartei
SVPW	Schweiz. Vereinigung für Politische Wissenschaft
TA	Tages-Anzeiger
Tab.	Tabelle
TCS	Touring Club der Schweiz
USG	Bundesgesetz über den Umweltschutz (Umweltschutzgesetz) (SR 814.01)

usw.	und so weiter
UVP	Umweltverträglichkeitsprüfung
UVPV	Verordnung über die Umweltverträglichkeitsprüfung (UVP) (SR 814.011)
vgl.	vergleiche
Vol.	Volume
WWF	World Wildlife Found
z.B.	zum Beispiel
ZB	Zwischenbericht
ZGZ	Zürcher Gemeinschaftszentren
Ziff.	Ziffer
zit. n.	zitiert nach
zw.	zwischen

9. LITERATURVERZEICHNIS

Die Literaturhinweise zu den empirischen Fallstudien finden sich in den Kapiteln 5.1.5.3., 5.2.5.3. und 5.3.5.3.

- ABERBACH, Joel D. (1990): Keeping a watchful eye. The politics of congressional oversight. Washington D.C.: The Brookings Institution.
- AGEVAL (1991): Die Wirkungen staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge. Arbeitsgruppe 'Gesetzesevaluation'. Schlussbericht an das EJPD. Bern: EDMZ.
- ALBAEK, Erik (1995): "Between knowledge and power: Utilization of social science in public policy making." In: Policy Sciences, Vol. 28, No. 1: 79-100.
- ALDENDERFER, Mark S./BLASHFIELD, Roger K. (1984): Cluster analysis. Beverly Hills: Sage.
- ALKIN, Marvin C. (ed.) (1990): Debates on evaluation. Newbury Park: Sage.
- AMERICAN EVALUATION ASSOCIATION (AEA) (1994): "Guiding principles for evaluators." In: American Evaluation Association: Evaluation Practice News, June 1994.
- AMY, Douglas J. (1984): "Why policy analysis and ethics are incompatible." In: Journal of Policy Analysis and Management, Vol. 3, No. 4: 573-591.
- AMY, Douglas J. (1987): "Can policy analysis be ethical?" In: Fischer, Frank/Forester, John (eds.): Confronting values in policy analysis. The politics of criteria. Newbury Park: Sage: 45-67.
- ANDERSON, Scarvia B./BALL, Samuel (1978): The profession and practice of program evaluation. San Francisco. Jossey-Bass.
- BACKHAUS, Klaus/ERICHSON, Bernd/PLINKE, Wulff/WEIBER, Rolf (1994): Multivariate Analysemethoden. 7. Auflage. Berlin: Springer.
- BADURA, Bernhard (Hg.)(1976): Seminar: Angewandte Sozialforschung. Frankfurt: Suhrkamp.
- BALTHASAR, Andreas/KNÖPFEL, Carlo (1994): Umweltpolitik und technische Entwicklung. Eine politikwissenschaftliche Evaluation am Beispiel der Heizungen. Basel: Helbing & Lichtenhahn.
- BANK, Adrienne (1985): "Private thoughts and public actions: A commentary on articles dealing with ethical issues in evaluation". In: Evaluation and Program Planning, Vol. 8: 67-72.
- BARBER, Bernhard (1990): Social studies of science. New Brunswick: Transaction.
- BARKER, Anthony/PETERS, B. Guy (1993): The politics of expert advice: Creating, using and manipulating scientific knowledge for public policy. Edinburgh: Edinburgh University Press.
- BECK, Ulrich (1974): Objektivität und Normativität - Die Theorie-Praxis-Debatte in der modernen deutschen und amerikanischen Soziologie. Reinbek: Rowohlt.
- BECK, Ulrich/BONSS, Wolfgang (1984): "Soziologie und Modernisierung. Zur gesellschaftlichen Ortsbestimmung der Verwendungsforschung". In: Soziale Welt 35, Heft 4: 381-405.
- BECK, Ulrich/BONSS, Wolfgang (Hg.) (1989): Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens. Frankfurt: Suhrkamp.
- BENSON, Jeri/MICHAEL, William B. (1987): "Designing evaluation studies: A twenty-year perspective." In: International Journal of Educational Research, Vol. 11, No. 1: 43-56.
- BERG, Gerda M. van den (1995): Evaluation-quality audit. The Hague: Netherlands Court of Audit. (mimeo)

- BERGER, Peter L./LUCKMANN, Thomas (1977): Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie. Frankfurt: Fischer.
- BERK, Richard A./ROSSI, Peter H. (1990): Thinking about program evaluation. Newbury Park: Sage.
- BEYWL, Wolfgang (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Frankfurt: Lang.
- BICKMAN, Leonard (ed.) (1987): Using program theory in evaluation. New Directions for Program Evaluation, No. 33. San Francisco: Jossey-Bass.
- BIERTER, Willy/BINDER, Hans-Martin (1993): Unternehmerische Innovationsprozesse. Die Wirksamkeit staatlicher Förderungsmaßnahmen - eine Evaluation. Chur: Rügger.
- BLACK, Thomas R. (1993): Evaluating social science research. An introduction. London: Sage.
- BOGGS, James P. (1992): "Implicit models of social knowledge use." In: Knowledge: Creation, Diffusion, Utilization, Vol. 14, No. 1: 29-62.
- BORGATTI, Steve/EVERETT, Martin/FREEMAN, Lin (1992): UCINET IV. Version 1.0. Columbia: Analytic Technologies.
- BOZEMAN, Barry (1986): "The credibility of policy analysis: Between method and use." In: Policy Studies Journal, Vol. 14, 4 (June): 518-539.
- BOZEMAN, Barry/LANDSBERGEN, David (1989): "Truth and credibility in sincere policy analysis. Alternative approaches for the production of policy-relevant knowledge." In: Evaluation Review, Vol. 13, No. 4: 355-379.
- BRAYBROOKE, David/LINDBLOM, Charles E. (1969 (1963)): A strategy of decision: Policy evaluation as a social process. 3rd printing. New York: Free Press.
- BROWN, Robert D./BRASKAMP, Larry A./NEWMAN, Dianna L. (1978): "Evaluator credibility as a function of report style. Do jargon and data make a difference?" In: Evaluation Quarterly, Vol. 2, No. 2: 331-341.
- BROWN, Robert D./NEWMAN, Dianna L. (1992): "Ethical principles and evaluation standards. Do they match?" In: Evaluation Review, Vol. 16, No. 6: 650-663.
- BROWN, Robert D./NEWMAN, Dianna L./RIVERS, Linda S. (1984): "A decisionmaking context model for enhancing evaluation utilization." In: Educational Evaluation and Policy Analysis, Vol. 6: 383-400. Auch in: ESRA, Vol. 11 (1986): 224-231.
- BROWNE, Angela/WILDAVSKY, Aaron (1984): "What should evaluation mean to implementation?" In: Pressman, Jeffrey L./Wildavsky, Aaron: Implementation. 3rd ed. Berkeley: University of California Press: 181-205. Auch in: Palumbo, Dennis James (ed.) (1987): The politics of program evaluation. Newbury Park: Sage: 146-172.
- BRYK, Anthony S. (ed.) (1983): Stakeholder-based evaluation. New Directions for Program Evaluation, No. 17. San Francisco: Jossey-Bass.
- BRYMAN, Alan (1988): Quality and quantity in social research. London: Unwin Hyman.
- BULMER, Martin (1982): The uses of social research. Social investigation in public policy-making. London: Allen & Unwin.
- BUNDA, Mary Anne (1985): "Alternative systems of ethics and their application to education and evaluation." In: Evaluation and Program Planning, Vol. 8: 25-36.
- BUSSMANN, Werner (1995): Evaluations and intensive democratic participation: Complement or replacement. Evaluation '95. Vancouver, Canada, November 2, 1995. (mimeo)
- BUSSMANN, Werner (Hrsg.) (1994): Lernen in Verwaltungen und Policy-Netzwerken. Chur: Rügger.
- CAMPBELL, Donald T. (1957): "Factors relevant to the validity of experiments in social settings." In: Psychological Bulletin 54 (4): 297-312.

- CAMPBELL, Donald T. (1984): "Can we be scientific in applied social science?" In: *Evaluation Studies Review Annual*, Vol. 9: 26-48.
- CAMPBELL, Donald T. (1988): *Methodology and epistemology for social science. Selected papers*. Edited by E. Samuel Overman. Chicago: University of Chicago Press.
- CANADIAN EVALUATION SOCIETY STANDARDS DEVELOPMENT COMMITTEE (1992): "Standards for program evaluation in Canada: A discussion paper." In: *Canadian Journal of Program Evaluation*, Vol. 7, No. 1: 157-170.
- CARACELLI, Valerie J./RIGGIN, Leslie J.C. (1994): "Mixed-method evaluation: Developing quality criteria through concept mapping." In: *Evaluation Practice*, Vol. 15, No. 2: 139-152.
- CHELIMSKY, Eleanor (1987): "Linking program evaluation to users needs." In: Palumbo, Dennis James (ed.): *The politics of program evaluation*. Newbury Park: Sage: 72-99.
- CHEN, Huey-Tsyh (1990): *Theory driven evaluations*. Newbury Park: Sage.
- CHEN, Huey-Tsyh/ROSSI, Peter H. (1980): "The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science." In: *Social Forces* 59 (1): 106-122. Auch in: *ESRA*, Vol. 6 (1981): 38-54.
- CHEN, Huey-Tsyh/ROSSI, Peter H. (1983): "Evaluating with sense: The theory-driven approach to evaluation." In: *Evaluation Review* 7(3): 283-302. Auch in: *ESRA*, Vol. 9 (1984): 337-356.
- CHEN, Huey-Tsyh/ROSSI, Peter H. (eds.) (1992): *Using theory to improve program and policy evaluations*. New York: Greenwood Press.
- CIARLO, J.A. (Hg.) (1981): *Utilizing evaluation. Concepts and measurement techniques*. Beverly Hills: Sage.
- COMMISSARIAT GENERAL DU PLAN (ed.) (1991): *Outils, pratiques, institutions pour évaluer les politiques publiques. Actes du séminaire PLAN-ENA, Avril-Juillet 1990. Textes rassemblés par Bernard Perret*. Paris: La Documentation Française.
- COMPTROLLER GENERAL OF THE UNITED STATES (1988): *Government auditing standards. Standards for audit of governmental organizations, programs, activities, and functions*. Washington, D.C.: General Accounting Office.
- CONNER, Ross F./HENDRICKS, Michael (eds.) (1989): *International innovations in evaluation methodology. New Directions for Program Evaluation*, Vol. 42. San Francisco: Jossey-Bass.
- COOK, Thomas D. (1985): "Postpositivist critical multiplism." In: Shotland, Robert Lance/Mark, Melvin M. (eds.): *Social science and social policy*. Beverly Hills: Sage: 21-62.
- COOK, Thomas D./CAMPBELL, Donald T. (1979): *Quasi-experimentation. Design and analysis for field-settings*. Boston: Houghton Mufflin.
- COOK, Thomas D./COOPER, Harris/CORDRAY, David S./HARTMANN, H./HEDGES, Larry V./LIGHT, Richard J./LOUIS, T./MOSTELLER, F. (eds.) (1992): *Meta-analysis for explanation: A case book*. New York: Russell Sage Foundation.
- COOK, Thomas D./GRUDER, C.L. (1978): "Metaevaluation research." In: *Evaluation Quarterly*, Vol. 2: 5-51.
- COOK, Thomas D./REICHHARDT, Charles S. (eds.) (1979): *Qualitative and quantitative methods in evaluation research*. Beverly Hills: Sage.
- COOK, Thomas D./SHADISH, William R. (1982): "Metaevaluation: An assessment of the congressionally mandated evaluation system for community mental health centers." In: Stahler, Gerald J./Tash, William R. (eds.): *Innovative approaches to mental health evaluation*. New York: Academic Press: 221-253.
- COOPER, Harris/HEDGES, Larry V. (eds.) (1994): *The handbook of research synthesis*. New York: Russell Sage Foundation.

- CORDRAY, David S. (1992): "Theory-driven meta-analysis: Practices and prospects." In: Chen, Huey-Tsyh/Rossi, Peter H. (eds.): Using theory to improve program and policy evaluations. New York: Greenwood Press: 85-95.
- COVERT, Robert W. (1995): "A twenty-year veteran's reflections on the guiding principles for evaluators." In: Shadish, William R./Newman, Dianna L./Scheirer, Mary Ann/Wye, Christopher (eds.): Guiding principles for evaluators. New Directions for Program Evaluation, No. 66 (Summer 1995). San Francisco: Jossey-Bass: 35-45.
- COXON, A.P.M. (1982): The user's guide to multidimensional scaling. London: Heinemann.
- CRONBACH, Lee J. (1982): "In praise of uncertainty." In: Rossi, Peter H. (ed.): Standards for evaluation practice. New Directions for Program Evaluation, No. 15. San Francisco: Jossey-Bass: 49-58. Auch in: ESRA, Vol. 9 (1984): 693-702.
- CRONBACH, Lee J. (1989): Designing evaluations of educational and social programs. 4th printing. San Francisco: Jossey-Bass.
- CRONBACH, Lee J./AMBRON, S.R./DORNBUSCH, Sanford M./HESS, R.D./HORNICK, R.C./PHILLIPS, D.C./WALKER, D.F./WEINER, S.S. (1980): Toward reform of program evaluation. San Francisco: Jossey-Bass.
- DAILLAK, Richard H. (1982): "What is evaluation utilization?" In: Studies in Educational Evaluation, Vol. 8: 157-162.
- DATTA, Lois-Ellin/PERLOFF, Robert (eds.) (1979): Improving evaluations. Beverly Hills: Sage.
- DAVIS, Dwight F. (1990): "Do you want a performance audit or a program evaluation?" In: Public Administration Review, Vol. 50, No. 1: 35-41.
- DEGRACIE, James S./MERRILL, Beverly/ZAHARIS, James K. (1994): Implications of using the revised program evaluation standards in local education agencies." In: The Journal of Experimental Education, Vol. 63, No. 1 (Fall 1994): 45-53.
- DELEAU, Michel/NIOCHE, Jean-Pierre/PENZ, Philippe/POINSARD, Robert (1986): Évaluer les politiques publiques. Méthodes, déontologie, organisation. Paris: La Documentation Française.
- DeLEON, Peter (1988): Advice and consent. The development of the policy sciences. New York: Russell Sage Foundation.
- DENZIN, Norman K. (1989): The research act. 3rd ed. Englewood Cliffs: Prentice Hall.
- DEWE, Bernd (1988): Wissensverwendung in der Fort- und Weiterbildung. Zur Transformation wissenschaftlicher Informationen in Praxisdeutungen. Baden-Baden: Nomos.
- DEWE, Bernd (1991): Beratende Wissenschaft. Unmittelbare Kommunikation zwischen Sozialwissenschaftlern und Praktikern. Göttingen: Schwartz.
- DUFFY, Barbara Poitras (1994): "Use and abuse of internal evaluation." In: New Directions for Program Evaluation, No. 64 (Winter 1994). San Francisco: Jossey-Bass: 25-32.
- DUNN, William N. (1991-92): "Assessing the impact of policy analysis: The functions of usable ignorance." In: Knowledge and Policy, Vol. 4, No. 4: 36-55.
- DUNN, William N. (1994): Public policy analysis: An introduction. Second edition. Englewood Cliffs: Prentice-Hall.
- DUNN, William N./KELLY, Rita Mae (eds.) (1992): Advances in policy studies since 1950. Policy Studies Review Annual, Vol. 10. New Brunswick: Transaction.
- EDLUND, Claes/HERMEREN, Göran (eds.) (1984): Evaluation research. Proceedings from the MURA symposium. Lund: Studentlitteratur.
- EGERBERG, Morten (1994): "Bridging the gap between theory and practice: The case of administrative policy." In: Governance, Vol. 7, No. 1: 83-98.

- EISNER, Elliot W. (1988): "Educational connoisseurship and criticism: Their form and functions in educational evaluation." In: Fetterman, David M. (ed.): *Qualitative approaches to evaluation in education. The silent scientific revolution.* New York: Praeger: 138-150.
- ERS STANDARDS COMMITTEE (1982): "Evaluation Research Society standards for program evaluation." In: Rossi, Peter H. (ed.): *Standards of evaluation practice. New Directions for Program Evaluation*, No. 15. San Francisco: Jossey-Bass. Auch in: *ESRA*, Vol. 9 (1984): 680-692.
- FETTERMAN, David M./KAFTARIAN, Shakeh J./WANDERSMAN, Abraham (eds.) (1996): *Empowerment evaluation. Knowledge and tools for self-assessment & accountability.* Thousand Oaks: Sage.
- FISCHER, Frank/FORESTER, John (eds.) (1993): *The argumentative turn in policy analysis and planning.* London: UCL Press.
- FORESTER, John (1989): *Planning in the face of power.* Berkeley: University of California Press.
- FORMAINI, Robert (1990): *The myth of scientific public policy.* New Brunswick: Transaction Publishers.
- FORSCHUNGSSTELLE WISSENSCHAFT UND POLITIK (1993): *Umweltforschungskatalog der Schweiz, 1990-1992.* 5. Auflage. Schriftenreihe Umwelt, Nr. 201. Bern: BUWAL.
- FREIBURGHAUS, Dieter (1988): *Methoden der Evaluation von Verwaltungshandeln aus der Sicht der Politik- und Verwaltungswissenschaft.* Cahiers de l'IDHEAP, No. 46.
- FREIBURGHAUS, Dieter (1993): "De l'efficacité des évaluations. Un discours en trois étapes." Dans: Morand, Charles-Albert (ed.): *Evaluation législative et lois expérimentales.* Aix-en-Provence: Presses universitaires d'Aix-Marseille: 47-60.
- FREIBURGHAUS, Dieter/ZIMMERMANN, Willi (1985): *Wie wird Forschung politisch relevant. Erfahrungen in und mit den Schweizerischen Nationalen Forschungsprogrammen.* Bern: Haupt.
- FRIEDMANN, John (1987): *Planning in the public domain: From knowledge to action.* Princeton: Princeton University Press.
- FUNNELL, Sue (1995): *Program theory for program design and evaluation: An Australian approach using program logic.* Paper, International Evaluation Conference, Vancouver, 1995. (mimeo)
- GALLEGOS, Arnold (1994): "Meta-evaluation of school evaluation models." In: *Studies in Educational Evaluation*, Vol. 20, No. 1: 41-54.
- GANZ, Carole (1984): "Plädoyer für eine Evaluierung des Evaluierungsforschungsprozesses." In: Hellstern, Gerd-Michael/Wollmann, Hellmut (Hg.): *Handbuch zur Evaluierungsforschung*, Band 1. Opladen: Westdeutscher Verlag: 623-636.
- GEORGHIU, Luke (1995): "Assessing the framework programmes. A meta-evaluation." In: *Evaluation*, Vol. 1, No. 2. (October 1995): 171-188.
- GERHEUSER, Frohmüt W./SCHMID, André (1993): *Bedingt wirksam: Das Lohnungleichheitspostulat des Heimarbeitsgesetzes.* Serie "Schlussberichte" NFP 27. Bern: NFP 27.
- GLASER, Renita R./SHADISH, William R. (1995): *Coding design in meta-analysis.* Paper presented at the annual convention of the American Evaluation Association, Vancouver, B.C. Canada, November 2, 1995. (mimeo)
- GLASS, Gene V./McGAW, B./SMITH, Mary Lee (1981): *Meta-analysis in social research.* Beverly Hills: Sage.
- GOWIN, D. Bob (1981): "Philosophic analysis." In: Nick L. Smith (ed.): *New techniques for evaluation.* Beverly Hills: Sage: 299-309.

- GREENBERG, David H./MANDELL, Marvin B. (1991): "Research utilization in policy making: A tale of two series (of social experiments)." In: *Journal of Policy Analysis and Management* 10: 633-656 (fall).
- GREENE, David/DAVID, Jane L. (1984): "A research design for generalizing from multiple case studies." In: *Evaluation and Program Planning*, Vol. 7: 73-85.
- GREENE, Jennifer C. (1990): "Technical quality versus user responsiveness in evaluation practice." In: *Evaluation and Program Planning*, Vol. 13: 267-274.
- GREENE, Jennifer C. (1992): "A case study of evaluation auditing as metaevaluation." In: *Evaluation and Program Planning*, Vol. 15: 71-74.
- GUBA, Egon Gotthold (ed.) (1990): *The paradigm dialog*. Newbury Park: Sage.
- GUBA, Egon Gotthold/LINCOLN, Yvonna S. (1981): *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.
- GUBA, Egon Gotthold/LINCOLN, Yvonna S. (1989): *Fourth generation evaluation*. Newbury Park: Sage.
- GUBA, Egon Gotthold/LINCOLN, Yvonna S. (1994): "Competing paradigms in qualitative research." In: Denzin, Norman K./Lincoln, Yvonna S. (eds.): *Handbook of qualitative research*. Thousand Oaks: Sage: 105-117.
- HAERING, Barbara et al. (1990): *Evaluation Kehrrechtsackgebühr. Expertenbericht von Infrac*. November 1990. Bern: BUWAL.
- HAKIM, Catherine (1992): *Research design. Strategies and choices in the design of social research*. London: Routledge.
- HANSEN, Joe B. (1994): "Book review: The Joint Committee on Standards for Educational Evaluation (1994): The program evaluation standards: How to assess evaluation of educational programs. Thousand Oaks: Sage." In: *Educational and Psychological Measurement*, Vol. 54, No. 2 (Summer 1994): 550-561.
- HEILMAN, J.G. (1980): "Paradigmatic choices in evaluation methodology." In: *Evaluation Review* 4 (5): 693-712.
- HELLSTERN, Gerd-Michael/WOLLMANN, Hellmut (1983): *Evaluierungsforschung. Ansätze und Methoden - dargestellt am Beispiel des Städtebaus*. Basel: Birkhäuser Verlag.
- HELLSTERN, Gerd-Michael/WOLLMANN, Hellmut (Hg.) (1984): *Handbuch zur Evaluierungsforschung. Band 1*. Opladen: Westdeutscher Verlag.
- HENDRICKS, Michael/CONNER, Ross F. (1995): "International perspectives on the guiding principles." In: Shadish, William R./Newman, Dianna L./Scheirer, Mary Ann/Wye, Christopher (eds.): *Guiding principles for evaluators*. *New Directions for Program Evaluation*, No. 66 (Summer 1995): 77-90.
- HÉRITIER, Adrienne (Hrsg.) (1993): *Policy-Analyse. Kritik und Neuorientierung*. PVS Sonderheft 24/1993. Opladen: Westdeutscher Verlag.
- HILL, James K./ANDREWS, D.A./HOGE, Robert D. (1991): "Meta-analysis of treatment programs for young offenders: The effect of clinically relevant treatment on recidivism, with controls for various methodological variables." In: *Canadian Journal of Program Evaluation*, Vol. 6, No. 1: 97-109.
- HITZLER, Ronald/HONER, Anne/MAEDER, Christoph (Hrsg.) (1994): *Expertenwissen. Die institutionalisierte Kompetenz zur Konstruktion von Wirklichkeit*. Opladen: Westdeutscher Verlag.
- HORBER-PAPAZIAN, Katia (1990): *Evaluation des politiques publiques en Suisse. Pourquoi? Pour qui? Comment?* Lausanne: Presses polytechniques et universitaires romandes.

- HORVATH, Lester (1995): Structured case study methodology in special education: Metaevaluation using The Program Evaluation Standards. Paper presented at the International Evaluation Conference, Vancouver, British Columbia, Canada. November 4, 1995. (mimeo)
- HOUSE, Ernest R. (1993): Professional evaluation. Social impact and political consequences. Newbury Park: Sage.
- HOWLETT, Michael/RAMESH, M. (1995): Studying public policy: Policy cycles and policy subsystems. Toronto: Oxford University Press.
- HUBERMAN, Michael/COX, Pat (1990): "Evaluation utilization: Building links between action and reflection." In: Studies in Educational Evaluation, Vol. 16: 157-179.
- HUDSON, Joe/McROBERTS, Hugh A. (1984): "Auditing evaluation activities." In: Rutman, Leonard (ed.): Evaluation research methods: A basic guide. 2nd Edition. Newbury Park: Sage: 219-236.
- HUNTER, John E./SCHMIDT, Frank L. (1989): Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park: Sage.
- JERRELL, S. Lee/JERRELL, Jeanette M. (1985): "Road signs in ethical quicksand." In: Evaluation and Program Planning, Vol. 8: 73-76.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981a): Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981b): Principles and by-laws. Western Michigan University Evaluation Center.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1988): The personnel evaluation standards. How to assess systems for evaluating educators. Newbury Park: Sage.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994): The program evaluation standards. How to assess evaluations of educational programs. 2nd edition. Thousand Oaks: Sage.
- KARLSSON, Ove (1995): The political dimension of the stakeholder model: How can the interaction between evaluation and politics be tackled? Paper presented at Evaluation '95. The Canadian Evaluation Society, and the American Evaluation Association Annual Meeting, November 1995. Vancouver, Canada. (mimeo)
- KAUFMANN, Franz-Xaver/MAJONE, Giandomenico/OSTROM, Vincent (eds.) (1986): Guidance, control, and evaluation in the public sector. The Bielefeld interdisciplinary project. Berlin: Walter de Gruyter.
- KAUFMANN, Franz-Xaver/ROSEWITZ, Bernd (1983): "Typisierung und Klassifikation politischer Massnahmen." In: Mayntz, Renate (Hrsg.): Implementation politischer Programme II. Ansätze zur Theoriebildung. Opladen: Westdeutscher Verlag: 25-49.
- KEITH, D. (1989): "Refining concept maps." In: Evaluation and Program Planning, Vol. 12, No. 1: 75-80.
- KELLY; Marisa/MAYNARD-MOODY, Steven (1994): "Policy analysis in the postpositivist era: Engaging stakeholders in evaluating the economic development district's program." In: White, Jay D./Adams, Guy B. (eds.): Research in public administration. Reflections on theory and practice. Thousand Oaks: Sage.
- KING, Jean A. (1995): Whose theory of use? The potential of participatory evaluation approaches for creating a theory of use. Handout prepared for presentation at Evaluation '95, Vancouver, British Columbia, November 2-5, 1995. (mimeo)

- KIRK, Jerome/MILLER, Marc L. (1986): Reliability and validity in qualitative research. Beverly Hills: Sage.
- KLAGES, Helmut (Hg.) (1985): Arbeitsperspektiven angewandter Sozialforschung. Opladen: Westdeutscher Verlag.
- KLÖTI, Ulrich/SCHNEIDER, Gerald (1989): Die Informationsbeschaffung des Gesetzgebers. Eine vergleichende Evaluationsstudie zur Bedeutung von Wirkungsanalysen im Schweizer Legiferierungsprozess. Grösch: Rüegger.
- KNOEPFEL, Peter et al. (1991): Evaluation des BUWAL. Expertenbericht zur Evaluation der Luftreinhaltepolitik, des ländlichen Gewässerschutzes und der UVP des Bundes. Bern: EDMZ.
- KOCH, Uwe/WITTMANN, Werner W. (Hrsg.) (1990): Evaluationsforschung. Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen. Berlin: Springer.
- KRUSKAL, Joseph B./WISH, Myron (1978): Multidimensional scaling. Beverly Hills: Sage.
- LAMNEK, Siegfried (1988): Qualitative Sozialforschung. Band 1: Methodologie. München: Psychologie Verlags Union.
- LAMNEK, Siegfried (1989): Qualitative Sozialforschung. Band 2: Methoden und Techniken. München: Psychologie Verlags Union.
- LAMPINEN, Osmo (1992): The utilization of social science research in public policy. Helsinki: Vapk-Kustannus.
- LASSEGARD, E. (1992): Assessing the reliability of the concept map process. Paper presented at the annual conference of the American Evaluation Association, Seattle WA. (mimeo)
- LASSWELL, Harold Dwight (1951): "The policy orientation." In: Lerner, Daniel/Lasswell, Harold Dwight (eds.): The policy sciences. Stanford: Stanford University Press.
- LeCOMPTE, Margaret D. (1994): "Sensible matchmaking: Qualitative research design and the Program Evaluation Standards." In: The Journal of Experimental Education, Vol. 63, No 1 (fall 1994): 29-43.
- LeCOMPTE, Margaret D./PREISSLE, Judith (1993): Ethnography and qualitative design in educational research. 2nd edition. San Diego: Academic Press.
- LEEUW, Frans L. (1991): "Policy theories, knowledge utilization, and evaluation." In: Knowledge and Policy, Vol. 4, No. 3: 73-91.
- LEGGE, Karen (1984): Evaluating planned organizational change. London: Academic Press.
- LESTER, James P. (1992): "The utilization of policy analysis by state agency officials." In: Knowledge: Creation, Diffusion, Utilization, Vol. 14, No. 3: 267-290.
- LESTER, James P./WILDS, Leah J. (1990): "The utilization of public policy analysis." In: Evaluation and Program Planning, Vol. 13: 313-319.
- LEVITON, Laura C./HUGHES, E. F. X. (1981): "Research on the utilization of evaluations: A review and synthesis." In: Evaluation Review, Vol. 5, No. 4: 525-548.
- LEWY, Arie (1984): "Evaluation standards: Comments from Israel." In: Evaluation and Program Planning, Vol. 7: 377-381.
- LIGHT, Richard J. (1984): "Six evaluation issues that synthesis can resolve better than single studies." In: Yeaton, W.H./Wortman, P.M. (eds.): Issues in data synthesis. New Directions for Program Evaluation, No. 24. San Francisco: Jossey-Bass: 57-73. Auch in: ESRA, Vol. 11 (1986): 703-719.
- LIGHT, Richard J./PILLEMER, David B. (1984): Summing up: The science of reviewing research. Cambridge, Mass.: Harvard University Press.
- LINCOLN, Yvonna S. (1985): The ERS standards for program evaluation: Guidelines for a fledging profession. In: Evaluation and Program Planning 8 (3): 251-253.

- LINCOLN, Yvonna S./GUBA, Egon Gotthold (1985): *Naturalistic inquiry*. Beverly Hills: Sage.
- LINDBLOM, Charles E. (1990): *Inquiry and change. The troubled attempt to understand and shape society*. New Haven: Yale University Press.
- LINDER, Stephen H./PETERS, B. Guy (1989): *Integration of policy studies through policy design*. Paper for the 1989 annual meeting of the APSA, Atlanta, September 1989. (mimeo)
- LINDER, Wolf et al. (1990): *Verbrauchsabhängige Heizkostenabrechnung, Förderung erneuerbarer Energien. Evaluation von Prozess und Wirkung in drei Kantonen*. März 1990. Bern: FSP.
- LIPSEY, Mark W. (1992): "Meta-analysis in evaluation research: Moving from description to explanation." In: Chen, Huey-Tsyh/Rossi, Peter H. (eds.): *Using theory to improve program and policy evaluations*. New York: Greenwood Press: 229-241.
- LIPTON, Douglas S. (1992): "How to maximize utilization of evaluation research by policymakers." In: *Annals of the American Academy of Political and Social Science* 521: 175-188.
- LONGCHAMP, Claude (1993): *Evaluierung der bundeseigenen EWR-Informationskampagne. Schlussbericht*. Bern: GfS.
- LORR, Maurice (1983): *Cluster analysis for social scientists*. San Francisco: Jossey-Bass.
- LOVE, Arnold J. (1991): *Internal evaluation. Building organizations from within*. Newbury Park: Sage.
- LOVE, Arnold J. (1993): "Internal evaluation: An essential tool for human organizations." In: *Canadian Journal of Program Evaluation*, Vol. 8, No. 2: 1-15.
- LOVE, Arnold J. (1994): "Should evaluators be certified?" In: *New Directions for Program Evaluation*, No. 62. San Francisco: Jossey-Bass: 29-40.
- LOVE, Arnold J. (ed.) (1983): *Developing effective internal evaluation*. *New Directions for Program Evaluation*, No. 20. San Francisco: Jossey-Bass.
- LOVELL, Robert G. (1995): "Ethics and internal evaluators." In: Shadish, William R./Newman, Dianna L./Scheirer, Mary Ann/Wye, Christopher (eds.): *Guiding principles for evaluators*. *New Directions for Program Evaluation*, No. 66 (Summer 1995): 61-67.
- LOWI, Theodore J. (1964): "American business, public policy, case-studies, and political theory." In: *World Politics*, Vol. 16, No. 1: 1-7.
- LOWI, Theodore J. (1972): "Four systems of policy, politics and choice." In: *Public Administration Review*, Vol. 32: 298-310.
- LUHMANN, Niklas (1971): *Politische Planung*. Opladen: Westdeutscher Verlag.
- MacDONALD, Barry (1993): "A political classification of evaluation studies in education." In: Hammersley, Martyn (ed.): *Social research. Philosophy, politics, and practice*. London: Sage: 105-108.
- MacRAE, Duncan Jr. (1991): "Policy analysis and knowledge use." In: *Knowledge and Policy*, Vol. 4, No. 3: 27-40.
- MADAUS, George F./SCRIVEN, Michael S./STUFFLEBEAM, Daniel L. (eds.) (1983): *Evaluation models. Viewpoints on educational and human services evaluations*. Boston: Kluwer-Nijhoff.
- MADER, Luzius (1985): *L'évaluation législative pour une analyse empirique des effets de la législation*. Lausanne: Payot.
- MAJONE, Giandomenico (1989): *Evidence, argument, and persuasion in the policy process*. New Haven: Yale University Press.
- MARKLUND, Sixten (1984): "Applicability of standards for evaluation of educational programs, projects and materials in an international setting." In: *Evaluation and Program Planning*, Vol. 7: 355-362.

- MATHISON, Sandra (1991): "Role conflicts for internal evaluators." In: *Evaluation and Program Planning*, Vol. 14: 173-179.
- MAYNE, J./BEMELMANS-VIDEC, M.L./HUDSON, J./CONNER, Ross F. (eds.) (1992): *Advancing public policy evaluation*. New York: North-Holland.
- MAYNTZ, Renate (Hrsg.) (1980): *Implementation politischer Programme*. Empirische Forschungsberichte. Königstein: Hain.
- MAYNTZ, Renate (Hrsg.) (1983): *Implementation politischer Programme II. Ansätze zur Theoriebildung*. Opladen: Westdeutscher Verlag.
- MAYRING, Philipp (1993): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 4., erweiterte Auflage. Weinheim: Deutscher Studien Verlag.
- McKILLIP, J./GARBERG, R. (1986): "Demands of the Joint Committee's standards for educational evaluation." In: *Evaluation and Program Planning*, Vol. 9: 325-333.
- McLAUGHLIN, John A./WEBER, Larry J./COVERT, Robert W./INGLE, Robert B. (eds.) (1988): *Evaluation utilization*. *New Directions for Program Evaluation*, Vol. 39, Fall 1988. San Francisco: Jossey-Bass.
- MERTENS, Donna M. (1990): "Practical evidence of the feasibility of the utilization-focused approach to evaluation." In: *Studies in Educational Evaluation*, Vol. 16: 181-194.
- MERTENS, Donna M. (1994): "Training evaluators: Unique skills and knowledge." In: *New Directions for Program Evaluation*, No. 62 (Summer 1994): 17-27.
- MILLMAN, Jason (1981): "A checklist procedure." In: Nick L. Smith (ed.): *New techniques for evaluation*. Beverly Hills: Sage: 309-320.
- MOHR, Lawrence B. (1995a): *Impact analysis for program evaluation*. Second edition. Thousand Oaks: Sage.
- MOHR, Lawrence B. (1995b): *The qualitative method of impact analysis*. Paper prepared for delivery at the annual meeting of the American Evaluation Association, Vancouver, November 1-5, 1995. (mimeo)
- MONNIER, Eric (1987): *Evaluations de l'action des pouvoirs publics. Du projet au bilan*. Paris: Economica.
- MORAND, Charles-Albert (ed.) (1993): *Evaluation législative et lois expérimentales*. Aix-en-Provence: Presses universitaires d'Aix-Marseille.
- MORRIS, Michael/COHN, Robin (1993): "Program evaluators and ethical challenges. A national survey." In: *Evaluation Review*, Vol. 17, No. 6 (December): 621-642.
- NÄF, Andreas et al. (1993): *Evaluation der Zürcher Gemeinschaftszentren (ZGZ). Vom Januar bis Mai 1993*. Im Auftrag des Sozialamtes der Stadt Zürich. Bern: Price Waterhouse.
- NEVO, David (1984): "Applying the evaluation standards in a different social context." In: *Evaluation and Program Planning*, Vol. 7: 383-386.
- NEWMAN, Dianna L. (1995): "The future of ethics in evaluation: Developing the dialogue." In: Shadish, William R./Newman, Dianna L./Scheirer, Mary Ann/Wye, Christopher (eds.): *Guiding principles for evaluators*. *New Directions for Program Evaluation*, No. 66 (Summer 1995): 99-110.
- NEWMAN, Dianna L./BROWN, Robert D. (1992): "Violations of evaluation standards. Frequency and seriousness of occurrence." In: *Evaluation Review*, Vol. 16, No. 3: 219-234.
- NEWMAN, Dianna L./BROWN, Robert D. (1996): *Applied ethics for program evaluation*. Thousand Oaks: Sage.
- NEWMAN, Frederick L./SMUKLER, Michael/GRIFFIN, Brian P./FISHMAN, Daniel B. (1992): "Conditions influencing the impact of state level program evaluation." In: *Evaluation and Program Planning*, Vol. 15: 345-355.

- NOWAKOWSKI, Jeri (1990): "Exploring the role of professional standards in evaluations: Areas of needed research." In: *Studies in Educational Evaluation*, Vol. 16: 271-296.
- NOWOTNY, Helga (1990): *In search of usable knowledge: Utilization contexts and the application of knowledge*. Frankfurt: Campus.
- OFFE, Claus (1977): *Strukturprobleme des kapitalistischen Staates*. 4. Auflage. Frankfurt: Suhrkamp.
- OFFE, Claus (1982): "Sozialwissenschaften zwischen Auftragsforschung und sozialer Bewegung." In: Beck, Ulrich (Hrsg.): *Soziologie und Praxis. Erfahrungen, Konflikte, Perspektiven. Soziale Welt, Sonderband 1*: 107-113.
- ORATA, P.T. (1940): "Evaluating evaluation." In: *Journal of Educational Research* 33: 641.
- PALUMBO, Dennis James (ed.) (1987): *The politics of program evaluation*. Newbury Park: Sage.
- PALUMBO, Dennis James/GREGWARE, Peter R. (1992): "Evaluating New Dawn and Pegasus using the Chen and Rossi multigoal, theory-driven approach." In: Chen, Huey-tsyh/Rossi, Peter H (eds.): *Using theory to improve program and policy evaluations*. New York: Greenwood Press: 145-163.
- PALUMBO, Dennis James/HALLET, Michael A. (1993): "Conflict versus consensus models in policy evaluation and implementation." In: *Evaluation and Program Planning*, Vol. 16: 11-23.
- PALUMBO, Dennis James/NACHMIAS, David (1983): "The preconditions for successful evaluation: Is there an ideal paradigm?" In: *Policy Sciences*, Vol. 16: 67-79.
- PAPPI, Franz Urban (Hrsg.) (1987): *Methoden der Netzwerkanalyse*. München: Oldenbourg.
- PARSONS, Wayne (1995): *Public policy. An introduction to the theory and practice of policy analysis*. Aldershot: Edward Elgar.
- PATTON, Michael Quinn (1978): *Utilization focused evaluation*. Beverly Hills: Sage.
- PATTON, Michael Quinn (1990): *Qualitative evaluation and research methods*. 2nd edition. Newbury Park: Sage.
- PATTON, Michael Quinn (1994): "Book review: The Joint Committee on Standards for Educational Evaluation (1994): The program evaluation standards: How to assess evaluations of educational programs. Thousand Oaks: Sage." In: *Educational and Psychological Measurement*, Vol. 54, No. 2 (Summer 1994): 561-567.
- PATTON, Michael Quinn (1994): "The program evaluation standards: How to assess evaluations of educational programs, by The Joint Committee on Standards for Educational Evaluations, Newbury Park, CA: Sage, (In Press)." Review in: *Evaluation Practice*, Vol. 15, No. 2: 193-199.
- PETERS, B. Guy/BARKER, Anthony (1993): *Advising west European governments: Inquiries, expertise and public policy*. Edinburgh: Edinburgh University Press.
- POPPER, Karl R. (1989): *Logik der Forschung*. Neunte, verbesserte Auflage. Tübingen: Mohr.
- POSAVAC, Emil J./CAREY, Raymond G. (1989): *Program evaluation. Methods and case studies*. Third Edition. Englewood Cliffs: Prentice Hall.
- PRÖPPER, Igno M.A.M. (1993): "Argumentation and power in evaluation-research and in its utilization in the policy-making process." In: Schomberg, Rene von (ed.): *Science, politics, and morality: Scientific uncertainty and decision making*. Dordrecht: Kluwer: 127-158.
- PUTNAM, L. (ed.) (1977): *Evaluation research methods*. Beverly Hills: Sage.
- RAIZEN, Senta A./ROSSI, Peter H. (1981): *Program evaluation in education: When? How? To what ends?* Washington: National Academy Press.

- RAVEN, John (1984): "Some limitations of the standards." In: *Evaluation and Program Planning*, Vol. 7: 363-370.
- RIDINGS, S.M./STUFFLEBEAM, Daniel L. (1981): "The project to develop standards for educational evaluation: Its past and future." In: *Studies in Educational Evaluation*, Vol. 7: 3-16.
- RIST, Ray C. (1989/90): "On the application of program evaluation designs: Sorting out their use and abuse." In: *Knowledge in Society*, Winter 1989-90, Vol. 2, No. 4: 74-96.
- RIST, Ray C. (1994): "Influencing the policy process with qualitative research." In: Denzin, Norman K./Lincoln, Yvonna S. (eds.): *Handbook of qualitative research*. Thousand Oaks: Sage: 545-557.
- ROSENTHAL, Robert (1991): *Meta-analytic procedures for social research*. Rev. edition. Newbury Park: Sage.
- ROSSI, Peter H. (ed.) (1982): *Standards of evaluation practice*. New Directions for Program Evaluation, Vol. 15. San Francisco: Jossey-Bass.
- ROSSI, Peter H./FREEMAN, Howard Edgar (1989): *Evaluation. A systematic approach*. Fourth Edition. Newbury Park: Sage.
- RUTMAN, Leonard (1980): *Planning useful evaluations. Evaluability assessment*. Beverly Hills: Sage.
- RUTMAN, Leonard (1982): "Dimensions of utilization and types of evaluation approaches." In: *Studies in Educational Evaluation*, Vol. 8: 163-168.
- RUTMAN, Leonard (ed.) (1984): *Evaluation research methods: A basic guide*. 2nd edition. Newbury Park: Sage.
- SABATIER, Paul A./JENKINS-SMITH, Hank C. (eds.) (1993): *Policy change and learning. An advocacy coalition approach*. Boulder: Westview Press.
- SANDEFUR, Gary D./ROSSI, Peter H./FREEMAN, Howard Edgar (1989): *Workbook for evaluation*. 4th edition. Newbury Park: Sage.
- SANDERS, James R. (1994): "The process of developing national standards that meet ANSI guidelines." In: *The Journal of Experimental Education*, Vol. 63, No. 1 (Fall 1994): 5-12
- SANDERS, James R. (1995): "Standards and principles." In: Shadish, William R./Newman, Dianna L./Scheirer, Mary Ann/Wye, Christopher (eds.): *Guiding principles for evaluators*. New Directions for Program Evaluation, No. 66 (Summer 1995): 47-52.
- SCHULZ, Hans-Rudolf/MUGGLI, Christoph/HÜBSCHLE, Jörg (1993): *Wohneigentumsförderung durch den Bund. Die Wirksamkeit des Wohnbau- und Eigentumsförderungsgesetzes (WEG)*. Schriftenreihe Wohnungswesen, Band 55. Bern: BWO/EDMZ.
- SCHWANDT, Thomas A. (1990): "Defining quality in evaluation." In: *Evaluation and Program Planning*, Vol. 13: 177-188.
- SCHWANDT, Thomas A. (1995): *Justifying claims in evaluation: Something old, something new?* Paper prepared for the first International Evaluation Conference, 4 November 1995, Vancouver, British Columbia. (mimeo)
- SCHWANDT, Thomas A./HALPERN, Edward S. (1988): *Linking auditing and metaevaluation: Enhancing quality in applied research*. Newbury Park: Sage.
- SCHWEIZERISCHE GESELLSCHAFT FÜR SOZIOLOGIE (SGS) (div. Jahre): *Laufende Soziologische Forschungen*. Zürich: SGS/Seismo.
- SCHWEIZERISCHER NATIONALFONDS ZUR FÖRDERUNG DER WISSENSCHAFTLICHEN FORSCHUNG (SNF) (diverse Jahre): *Jahresbericht*. Bern: SNF.
- SCOTT, John (1991): *Social network analysis. A handbook*. London: Sage.
- SCRIVEN, Michael (1969): *An introduction to meta-evaluation*. Educational Product Report, Vol. 2, No. 5: 36-38.

- SCRIVEN, Michael (1983): "Costs in evaluation. Concept and practice." In: Alkin, Marvin C./Solmon, Lewis C. (eds.): The costs of evaluation. Beverly Hills: Sage: 27-44.
- SCRIVEN, Michael (1984): "Evaluation ideologies." In: Madaus, George F./Scriven, Michael S./Stufflebeam, Daniel L. (eds.) (1983): Evaluation models. Viewpoints on educational and human services evaluations. Boston: Kluwer-Nijhoff. Auch In: ESRA, Vol. 9 (1984): 49-80.
- SCRIVEN, Michael (1991): Evaluation thesaurus. 4th edition. Newbury Park: Sage.
- SHADISH, William R. (1989): "The perception and evaluation of quality in science." In: Gholson, Barry/Shadish, William R./Neimeyer, Robert A./Houts, Arthur C. (eds.): The psychology of science: Contributions to metascience. Cambridge: Cambridge University Press: 383-426.
- SHADISH, William R. (1992): "Theory-driven metaevaluation." In: Chen, Huey-tysh/Rossi, Peter H. (eds.): Using theory to improve program and policy evaluations. New York: Greenwood Press: 29-45.
- SHADISH, William R. (1995): Of boundaries and politics: The case of consequential validity. Paper presented at the annual meeting of the American Evaluation Association, Vancouver, Canada, November 2, 1995. (mimeo)
- SHADISH, William R./COOK, Thomas D./LEVITON, Laura C. (1991): Foundations of program evaluation. Newbury Park: Sage.
- SHADISH, William R./NEWMAN, Dianna L./SCHEIRER, Mary Ann/WYE, Christopher (eds.) (1995): Guiding principles for evaluators. New Directions for Program Evaluation, No. 66 (Summer 1995). San Francisco: Jossey-Bass.
- SHULA, Lyn M./COUSINS, J. Bradley (1995): Utilization and social justice: Interconnections of meta-evaluation frameworks. A paper presentation for Evaluation '95. An International Evaluation Conference, Vancouver, British Columbia, November, 1995. (mimeo)
- SIDOS (1993): Sozialwissenschaftliche Forschung 1993. Zürich: Seismo.
- SMITH, Denise M./SMITH, Nick L. (1981): "Writing effective evaluation reports." In: Studies in Educational Evaluation, Vol. 7: 33-41.
- SMITH, Nick L. (1981a): "Criticism and meta-evaluation." In: Nick L. Smith (ed.): New techniques for evaluation. Beverly Hills: Sage: 266-274.
- SMITH, Nick L. (1981b): "Evaluating evaluation methods." In: Studies in Educational Evaluation, Vol. 7: 173-181.
- SMITH, Nick L. (1990): "Cautions on the use of investigative case studies in meta-evaluation." In: Evaluation and Program Planning, Vol. 13: 373-378.
- SMITH, Nick L. (1994): "Clarifying and expanding the application of program theory-driven evaluations." In: Evaluation Practice, Vol. 15, No. 1: 83-87.
- SMITH, Nick L. (1995): Designing investigative evaluations: Problem solution versus client responsiveness. Paper presented at the joint meeting of the Canadian Evaluation Society and the American Evaluation Association, Vancouver, B.C. Canada, November, 1995. (mimeo)
- SMITH, Nick L. (ed.) (1981c): New techniques for evaluation. Beverly Hills: Sage.
- SMITH, Nick L./CHIRCOP, Saviour/MUKHERJEE, Prachee (1993): "Considerations on the development of culturally relevant standards." In: Studies in Educational Evaluation, Vol. 19, No. 1: 3-13.
- SOMMER, Heini/SUTER, Stefan (1993): Schneeräumung von Urner Passstrassen. Evaluation im Auftrag der Baudirektion des Kantons Uri, mitfinanziert durch NFP 27. Schlussbericht. September 1993. Altdorf: Ecoplan.

- SONNICHSEN, Richard C. (1994): "Effective internal evaluation: An approach to organizational learning." In: Leeuw, Frans L./Rist, Ray C./Sonnichsen, Richard C. (eds.): *Can governments learn? Comparative perspectives on evaluation and organizational learning*. New Brunswick: Transaction: 125-141.
- SPÖHRING, Walter (1989): *Qualitative Sozialforschung*. Stuttgart: Teubner.
- STAKE, Robert E. (1981): "Setting standards for educational evaluation." In: *Evaluation News*, Vol. 2, No. 1: 148-152.
- STAKE, Robert E. (1986): *Quieting reform*. Urbana: University of Illinois Press.
- STAKE, Robert E. (1995): *The art of case study research*. Thousand Oaks: Sage.
- STEHR, Nico (1991): *Praktische Erkenntnis*. Frankfurt: Suhrkamp.
- STENZEL, Norman (1991): "Use and abuse of power: Evaluation in a state education agency." In: *Evaluation and Program Planning*, Vol. 14: 181-187.
- STEVENS, Carla J./DIAL, Micah (1994b): "What constitute misuse?" In: *New Directions for Program Evaluation*, No. 64 (Winter 1994). San Francisco: Jossey-Bass: 3-13.
- STEVENS, Carla J./DIAL, Micah (eds.) (1994a): *Preventing the misuse of evaluation*. *New Directions for Program Evaluation*, No. 64 (Winter 1994). San Francisco: Jossey-Bass.
- STOCKDILL, Stacey Hueftle (1986): *Evaluation standards: A study of their appropriateness in business & education*. Ph.D. University of Minnesota. *Dissertation Abstracts International (DAI)*, Vol. 47-08A: 2967.
- STRATON, Ralph G. (1995): *Enhancing information use: Implications of research on utilization for evaluation practice*. Paper presented at the International Evaluation Conference, Vancouver, November, 1995. (mimeo)
- STRUBE, Michael J./HARTMANN, Donald P. (1984): "Meta-analysis. techniques, applications, and functions." In: *Evaluation Studies Review Annual*, Vol. 9: 286-299.
- STRUENING, Elmer L./GUTTENTAG, Marcia (Hrsg.) (1975): *Handbook of evaluation research*. 2 vols. Beverly Hills: Sage.
- STUFFLEBEAM, Daniel L. (1978): "Meta evaluation: An overview." In: *Evaluation & the Health Professions* 1: 17-43.
- STUFFLEBEAM, Daniel L. (1981): "Metaevaluation: Concepts, standards, and uses." In: Berk, Richard A. (ed.): *Educational evaluation methodology: The state of the art*. Baltimore: Johns Hopkins University Press: 146-163.
- STUFFLEBEAM, Daniel L. (1983): "Reflections on the movement to promote effective educational evaluations through the use of professional standards." In: *Studies in Educational Evaluation*, Vol. 9: 119-124.
- STUFFLEBEAM, Daniel L. (1987): "Professional standards for assuring the quality of educational program and personnel evaluations." In: *International Journal of Educational Research*, Vol. 11, No. 1: 125-143.
- STUFFLEBEAM, Daniel L. (1994): "Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go." In: *Evaluation Practice*, Vol. 15, No. 3: 321-338.
- STUFFLEBEAM, Daniel L./SHINKFIELD, Anthony J. (1990): *Systematic evaluation: A self-instructional guide to theory and practice*. 4th printing. Boston: Kluwer-Nijhoff.
- THOMPSON, Bruce (1994): "The revised program evaluation standards and their correlation with the evaluation use literature." In: *The Journal of Experimental Education*, Vol. 63, No. 1 (Fall 1994): 54-81

- THURSTONE, Paul W./ORY, John C./MAYBERRY, Paul W./BRASKAMP, Larry A. (1984): "Legal and professional standards in program evaluation." In: Educational Evaluation and Policy Analysis, Vol. 6: 15-26. Auch in: ESRA, Vol. 11 (1986): 262-273.
- TOMLINSON, Carol/BLAND, Lori/MOON, Tonya/CALLAHAN, Carolyn (1994): "Case studies of evaluation utilization in gifted education." In: Evaluation Practice, Vol. 15, No. 2: 153-168.
- TORRES, Rosalie T. (1991): "Improving the quality of internal evaluation: The evaluator as consultant mediator." In: Evaluation and Program Planning, Vol. 14: 189-198.
- TROCHIM, William M.K. (1989a): "An introduction to concept mapping for planning and evaluation" In: Evaluation and Program Planning, Vol. 12: 1-16.
- TROCHIM, William M.K. (1989b): "Concept mapping: Soft science or hard art?" In: Evaluation and Program Planning, Vol. 12: 87-110.
- TROCHIM, William M.K. (1993): "The reliability of concept mapping. Paper presented at the annual conference of the American Evaluation Association, Dallas, Texas, November 1993. (mimeo)
- UHR, John (ed.) (1991): Program evaluation. Decision making in Australian government. Canberra: Federalism Research Centre.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1990): Prospective evaluation methods: The prospective evaluation synthesis. November 1990. Transfer Paper 10.1.10. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1991a): Designing evaluations. May 1991. GAO/PEMD-10.1.4. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1991b): Using structured interviewing techniques. July 1991. GAO/PEMD-10.1.5. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1992a): Quantitative data analysis: An introduction. June 1992. GAO/PEMD-10.1.11. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1992b): Case study evaluations. November 1990. Transfer Paper 10.1.9. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1992c): The Evaluation synthesis. Revised March 1992. GAO/PEMD-10.1.2. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- UNITED STATES GENERAL ACCOUNTING OFFICE (GAO) (ed.) (1992d): Using statistical sampling. Revised May 1992. GAO/PEMD-10.1.6. Washington D.C.: General Accounting Office, Program Evaluation and Methodology Division.
- VALL, Marc van de (1993): Angewandte Sozialforschung. Begleitung, Evaluierung und Verbesserung sozialpolitischer Massnahmen. Weinheim: Juventa.
- WAGNER, Peter/WEISS, Carol Hirschon/WITTROCK, Björn/WOLLMANN, Hellmut (eds.) (1991): Social sciences and modern states. National experiences and theoretical crossroads. Cambridge: Cambridge University Press.
- WALKER, Wallace Earl (1985): "The conduct of program evaluation reviews in the General Accounting Office." In: Evaluation and Program Planning, Vol. 8: 271-280.

- WALLER, Mike (1992): "Evaluating policy advice." In: Australian Journal of Public Administration, Vol. 51, No. 4: 440-446.
- WASSERMAN, Stanley/FAUST, Katherine (1994): Social network analysis: Methods and applications. Cambridge: Cambridge University Press.
- WASSERMAN, Stanley/GALASKIEWICZ, Joseph (eds.) (1994): Advances in social network analysis. Research in the social and behavioural sciences. Thousand Oaks: Sage.
- WEISS, Carol Hirschon (1977b): "Research for policy's sake: The enlightenment function of social science research." In: Policy Analysis, Vol. 3, No. 4 (1977): 531-545.
- WEISS, Carol Hirschon (1981): "Measuring the use of evaluation." In: Ciarlo, James A. (ed.): Utilizing evaluation: Concepts and measurement techniques. Beverly Hills: Sage: 17-33. Auch in: ESRA, Vol. 7 (1982): 129-145.
- WEISS, Carol Hirschon (1983): "Toward the future of stakeholder approaches in evaluation." In: New Directions for Program Evaluation, No. 17: 83-96. San Francisco: Jossey-Bass. Auch in: ESRA, Vol. 9 (1984): 255-268.
- WEISS, Carol Hirschon (1993): "Mega-lessons from U.S. evaluation experience." In: Politica, 25 (1): 64-76.
- WEISS, Carol Hirschon (ed.) (1977a): Using social research in public policy making. Lexington: Lexington Books.
- WHOLEY, Joseph S./NEWCOMER, Kathryn E. (1989): Improving government performance. Evaluation strategies for strengthening public agencies and programs. San Francisco: Jossey-Bass Publishers.
- WHOLEY, Joseph S./SCANLON, John W./DUFFY, Hugh, G./FUKUMOTO, James S./VOGT, Leona M. (1976): Federal evaluation policy. Analyzing the effects of public programs. Washington, D.C.: The Urban Institute.
- WIDMER, Thomas (1987): "Innerstaatliche Konfliktivität bilateraler Zusammenarbeit der Schweiz." In: Frei, Daniel/Ruloff, Dieter (Hrsg.): Internationale Zusammenarbeit. Kleine Studien zur Politischen Wissenschaft, Nr. 232. Zürich: Forschungsstelle für Politische Wissenschaft.
- WIDMER, Thomas (1991): Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz. Chur: Rüegger.
- WIDMER, Thomas (1996a): Fallstudien zur Meta-Evaluation, Teil I: 'Umweltpolitik und technische Entwicklung' (Studie Nr. 1), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996b): Fallstudien zur Meta-Evaluation, Teil II: 'Unternehmerische Innovationsprozesse' (Studie Nr. 2), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996c): Fallstudien zur Meta-Evaluation, Teil III: 'Wohneigentumsförderung durch den Bund' (Studie Nr. 4), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996d): Fallstudien zur Meta-Evaluation, Teil IV: 'Evaluation der Zürcher Gemeinschaftszentren' (Studie Nr. 5), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996e): Fallstudien zur Meta-Evaluation, Teil V: 'Evaluation Kehrichtsackgebühr' (Studie Nr. 7), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996f): Fallstudien zur Meta-Evaluation, Teil VI: 'Heizkostenabrechnung und erneuerbare Energien' (Studie Nr. 8), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WIDMER, Thomas (1996g): Fallstudien zur Meta-Evaluation, Teil VII: 'Evaluierung der EWR-Informationenkampagne' (Studie Nr. 10), Serie "Schlussberichte" NFP 27, Bern: NFP 27.
- WILDAVSKY, Aaron (1979): Speaking truth to power. The art and craft of policy analysis. Boston: Little, Brown and Co.

- WILDEMUTH, Barbara M. (1981): A bibliography to accompany the Joint Committee's standards on educational evaluation. Princeton, NJ: ERIC Clearinghouse on Tests, Measurement, & Evaluation, Educational Testing Service.
- WILLIAMS, Janice E. (1989): "A numerical developed taxonomy of evaluation theory and practice." In: *Evaluation Review*, Vol. 13, No. 1: 18-31.
- WINBERG, Alan (1991): "Maximizing the contribution of internal evaluation units." In: *Evaluation and Program Planning*, Vol. 14: 167-172.
- WINDHOFF-HÉRITIER, Adrienne (1987): *Policy-Analyse. Eine Einführung*. Frankfurt: Campus.
- WINGENS, Matthias (1988): *Soziologisches Wissen und politische Praxis. Neuere theoretische Entwicklungen der Verwendungsforschung*. Frankfurt: Campus.
- WINGENS, Matthias (1990): "Toward a general utilization theory. A systems theory reformulation of the two-communities metaphor." In: *Knowledge: Creation, Diffusion, Utilization*, Vol. 12, No. 1 (September): 27-42.
- WITTMANN, Werner W. (1985): *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Berlin: Springer.
- WOLF, Fredric (1986): *Meta-analysis*. Newbury Park: Sage.
- WORTHEN, Blaine R. (1994): "Is evaluation a mature profession that warrants the preparation of evaluation professionals?" In: *New Directions for Program Evaluation*, No. 62 (Summer 1994): 3-15.
- WOTTAWA, Heinrich/THIERAU, Heike (1990): *Lehrbuch Evaluation*. Bern: Huber.
- WULF, Christoph (Hrsg.) (1972): *Evaluation: Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München: Piper.
- ZIMMERMANN, Willi/KOLLY, Doris/MAHON, Pascal (1989): *Die wissenschaftliche Evaluation von Bundespolitik*. Bern: AGEVAL/EJPD.