

**Nationales Forschungsprogramm 27
Wirksamkeit staatlicher Massnahmen
Serie "Schlussberichte"**

**Fallstudien
zur Meta-Evaluation
Teil I**

**'Umweltpolitik und technische Entwicklung'
(Studie Nr. 1)**

**Thomas Widmer
Institut für Politikwissenschaft, Universität Zürich**

Bern, 1996



Elektronische Neuauflage
Zürich, 2006

Im Rahmen des NFP 27-Projektes 'Meta-Evaluation: Kriterien zur Bewertung von Evaluationen' sind folgende Publikationen erschienen (Autor Thomas Widmer):

Hauptbericht:

- Meta-Evaluation: Kriterien zur Bewertung von Evaluationen. Bern: Haupt 1996.
Hauptbericht mit den Fallstudien 'Lohnleichheitspostulat des Heimarbeitsgesetzes' (Studie Nr. 3), 'Schneeräumung von Urner Passstrassen' (Studie Nr. 6) und 'Evaluation des BUWAL' (Studie Nr. 9).

Fallstudien:

- Fallstudien zur Meta-Evaluation, Teil I: 'Umweltpolitik und technische Entwicklung' (Studie Nr. 1), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil II: 'Unternehmerische Innovationsprozesse' (Studie Nr. 2), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil III: 'Wohneigentumsförderung durch den Bund' (Studie Nr. 4), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil IV: 'Evaluation der Zürcher Gemeinschaftszentren' (Studie Nr. 5), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil V: 'Evaluation Kehrrechtsackgebühr' (Studie Nr. 7), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil VI: 'Heizkostenabrechnung und erneuerbare Energien' (Studie Nr. 8), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)
- Fallstudien zur Meta-Evaluation, Teil VII: 'Evaluierung der EWR-Informationskampagne' (Studie Nr. 10), Reihe "Schlussberichte" NFP 27, Bern 1996. (Fr. 5.-)

Kurzbericht:

- Meta-Evaluation als Instrument zur Qualitätspflege in der Evaluationsforschung, Reihe "Kurzberichte" NFP 27, Bern 1996. (gratis)

Der Hauptbericht ist im Buchhandel erhältlich. Die übrigen Texte können bezogen werden über: Schweizerischer Nationalfonds, Abt. IV/NFP 27, Postfach 2338, 3001 Bern, Tel. 031 308 23 49.

INHALTSVERZEICHNIS

Einführung	1
<i>1. Vorbemerkungen</i>	<i>1</i>
<i>2. Fragestellung</i>	<i>1</i>
<i>3. Kriterien der Meta-Evaluation</i>	<i>2</i>
<i>4. Hinweise zum methodischen Vorgehen</i>	<i>5</i>
Fallstudie 'Umweltpolitik und technische Entwicklung' (Studie Nr. 1)	7
<i>1. Die Evaluation</i>	<i>7</i>
1.1. Das Thema	7
1.2. Die Fragestellung	7
1.3. Das Vorgehen	8
1.4. Der Auftraggeber	8
1.5. Die Forscher	9
<i>2. Beschreibung des Evaluationsprozesses</i>	<i>11</i>
<i>3. Bewertung entlang der Kriterien</i>	<i>14</i>
3.1. Nützlichkeit	14
3.1.1. Identifikation der beteiligten Gruppen	14
3.1.2. Glaubwürdigkeit des Evaluators	15
3.1.3. Informationsrahmen und -auswahl	16
3.1.4. Bewertende Interpretationen	20
3.1.5. Klarheit des Berichts	21
3.1.6. Verbreitung des Berichts	22
3.1.7. Rechtzeitigkeit des Berichts	22
3.1.8. Wirkung der Evaluation	24
3.2. Anwendbarkeit	26
3.2.1. Praktische Verfahren	26
3.2.2. Politische Überlebensfähigkeit	27
3.2.3. Kostenwirksamkeit	28

3.3. Korrektheit	29
3.3.1. Formale Verpflichtung	29
3.3.2. Deklaration von Interessenkonflikten	32
3.3.3. Vollständige und aufrichtige Offenlegung	33
3.3.4. Herstellung von Öffentlichkeit	33
3.3.5. Schutz individueller Menschenrechte	33
3.3.6. Menschliche Interaktionen	34
3.3.7. Ausgewogene Berichterstattung	34
3.3.8. Finanzielle Verantwortlichkeit	34
3.4. Genauigkeit	35
3.4.1. Identifikation des Gegenstandes	35
3.4.2. Kontextanalyse	37
3.4.3. Beschreibung von Zielen und Vorgehen	37
3.4.4. Verlässliche Informationsquellen	40
3.4.5. Valide Messung	41
3.4.6. Reliable Messung	47
3.4.7. Systematische Datenüberprüfung	49
3.4.8. Analyse quantitativer Informationen	49
3.4.9. Analyse qualitativer Informationen	56
3.4.10. Gerechtfertigte Folgerungen	58
3.4.11. Objektive Berichterstattung	62
4. Gesamteindruck	65
5. Anhang	66
5.1. Unterlagen zur Evaluationsstudie	66
5.2. Liste der kontaktierten Personen	66
5.3. Literaturverzeichnis	67
5.4. Abkürzungsverzeichnis	72
6. Stellungnahme der Evaluatoren	74

EINFÜHRUNG

1. Vorbemerkungen

Die vorliegende Publikation bildet einen Teil eines grösseren Forschungsprojektes, das sich mit der Evaluation von Evaluationen - Meta-Evaluation genannt - befasst. Dieses Projekt wurde unter der Leitung von Prof. Ulrich Klöti am Institut für Politikwissenschaft an der Universität Zürich erarbeitet und durch das Nationale Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen' (NFP 27) finanziell unterstützt (SNF-Kredit Nr. 4027-39858.93).

Diese Einführung verfolgt die Absicht, die Fragestellung, den theoretischen Kontext sowie die Vorgehensweise des Gesamtprojektes darzulegen, um dem Leser eine Einordnung der nachfolgenden Fallstudie zu ermöglichen. Diese kurze Einführung kann jedoch die ausführliche Darlegung in der Hauptstudie (Widmer 1996a) in keiner Weise ersetzen. Für eine vertiefte Beschäftigung mit der Thematik ist die Lektüre des Hauptberichtes unabdingbar. Neben der vorliegenden Fallstudie sind weitere sechs Fallstudien in analoger Weise publiziert (Widmer 1996b-g). Die übrigen drei Fallstudien sind im Hauptbericht enthalten (Widmer 1996a).

2. Fragestellung

Das Gesamtprojekt befasst sich mit einer dreiteiligen 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?

Die zur Qualitätsbewertung ausgewählten Kriterien werden wir nachfolgend (siehe Kapitel 1.3.) kurz darlegen. Eine ausführliche Behandlung dieser Thematik findet sich in Kapitel 3 des Hauptberichtes (Widmer 1996a).

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)

Die hier vorliegende Fallstudie dient der Bearbeitung dieser Fragestellung anhand einer bestimmten Evaluationsstudie. Die übrigen neun Fallstudien sind an anderer Stelle publiziert (Widmer 1996a-g). Einige kurze Hinweise zur methodischen Vorgehensweise finden sich im nachfolgenden Kapitel 1.4. Für eine ausführliche Diskussion der Methodik sei auf den Hauptbericht verwiesen (Widmer 1996a, Kapitel 4).

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. In diesem Syntheseteil werden 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 Hauptstudie behandelt (Widmer 1996a).

3. Kriterien der Meta-Evaluation

Die eingesetzten Evaluationskriterien sind abgeleitet aus den im US-amerikanischen Kontext vom 'Joint Committee on Standards for Educational Evaluation' entwickelten 'Standards for Evaluations of Educational Programs, Projects and Materials' (Joint Committee 1981, vgl. dazu auch die überarbeitete Fassung dieser Standards, siehe Joint Committee 1994). Die Kriterien bestehen aus insgesamt dreissig Einzelkriterien, die vier Gruppen (Nützlichkeit, Anwendbarkeit, Korrektheit und Genauigkeit) zugeordnet sind. Zur Ausweitung des methodologischen Anwendungsbereichs auf Studien mit alternativer epistemologischer Ausrichtung werden die in dieser Weise hergeleiteten Kriterien mit den durch Guba und Lincoln (1989: 236-243, für die Kriterien 4.5. b) und c), 4.6. b) und 4.11. b)) vorgeschlagenen Parallelkriterien ergänzt. Die Kriterien lauten damit folgendermassen:

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.

Eine ausführliche Diskussion des Kriterienrasters ist in Kapitel 3 des Hauptberichts enthalten (Widmer 1996a).

4. Hinweise zum methodischen Vorgehen

Im Rahmen des Gesamtprojektes sind insgesamt zehn Evaluationsstudien evaluiert worden. Diese zehn Evaluationsstudien verfügen über folgende Eigenschaften:

- Externe Evaluationen, die seit 1990 in der Schweiz entstanden sind, deren Ergebnisse in irgendeiner Form publiziert wurden und den Anspruch erheben, einen sozialwissenschaftlichen Ansatz zu verfolgen.
- Evaluationen, die (mit unterschiedlichen Steuerungsprinzipien ausgestattete) Einzelmassnahmen, Gesamtprogramme oder Organisationen auf kommunaler, kantonaler oder nationaler Ebene untersuchen.
- Evaluationen, die von einer oder mehreren Personen aus einem universitären oder privaten Kontext mit unterschiedlicher disziplinärer Herkunft in einer Zeit von 4 bis zu 39 Monaten für 40'000.- bis zu 254'000.- erarbeitet wurden und die dabei, ausgehend von vielfältigen er-

kenntnistheoretischen Standpunkten, eine breite Auswahl diverser Forschungsmethoden einsetzen.

- Evaluationen, die bei Bund, Kantonen oder Gemeinden durch die Exekutive, durch die Legislative oder durch Institutionen der Forschungsförderung in Auftrag gegeben wurden.

Zur empirischen Untersuchung der zehn Evaluationsstudien wurde reichhaltiges Datenmaterial beigezogen. Die Meta-Evaluation stützt sich zum einen auf eine detaillierte Dokumentenanalyse unterschiedlichster schriftlicher Unterlagen (Ausschreibungsunterlagen, Offerten und Projektskizzen, Arbeitspapiere, Sitzungsprotokolle, Briefwechsel, Kurzberichte, Pressemitteilungen, Medienberichterstattung, finanzielle Abrechnungen, Projektjournale, Zwischen- und Schlussberichte usw.), zum zweiten auf eine Reihe von Gesprächen, die mit involvierten Personen (Evaluatoren, Auftraggeber, weitere beteiligte oder betroffene Personen) geführt wurden und drittens auf offizielle Dokumente sowie auf die Sekundärliteratur zur interessierenden Thematik.

Nach Abschluss der Fallstudien wurden die Texte den jeweils verantwortlichen Evaluatoren zur Stellungnahme unterbreitet. Nach - zum Teil mehrmaligen Überarbeitungen der Texte - nutzten verschiedene betroffene Evaluatoren die Gelegenheit, eine von ihnen verfasste Stellungnahme zur Fallstudie als Anhang zur betreffenden Fallstudie zu publizieren.

Die fallvergleichende Analyse stützte sich auf die Erfahrungen aus den zehn Fallstudien ab. Die Analyse ist darauf ausgerichtet handlungsrelevante Empfehlungen zuhanden der Evaluationspraxis zu formulieren und zu überprüfen, ob der eingesetzte Kriterienraster zur Durchführung einer derartigen Meta-Evaluation geeignet ist und ob allenfalls - gestützt auf die empirischen Erkenntnisse - Anpassungen am Kriterienraster erforderlich sind. Eine ausführliche Darstellung des Fallvergleichs sowie seiner Resultate findet sich in den Kapiteln 6 und 7 des Hauptberichts (Widmer 1996a).

Die nachfolgende Fallstudie bildet somit nur ein Element im Rahmen des Gesamtprojektes. Um das Projekt als Ganzes und damit den Kontext dieser Fallstudie ausreichend gründlich erfassen zu können, sollte vor allem der Hauptbericht (Widmer 1996a), aber möglicherweise auch eine oder mehrere der anderen Fallstudien (Widmer 1996b-g) konsultiert werden.

FALLSTUDIE 'UMWELTPOLITIK UND TECHNISCHE ENTWICKLUNG' (STUDIE NR. 1)

1. Die Evaluation

1.1. Das Thema

Die Studie (Balthasar/Knöpfel 1994) konzentriert sich auf die Analyse der innovationsfördernden Auswirkungen eines Programms der schweizerischen Umweltpolitik, somit also um eine Nebenwirkung des politischen Programms. Sie untersucht luftreinhaltepolitische Massnahmen des technischen Umweltschutzes im Bereich der Hausfeuerungen. Konkret werden die Einzelmassnahmen 'Grenzwertsetzung', 'Typenprüfung' und 'Feuerungskontrolle' betrachtet. Dieses Massnahmenpaket wird auch als gesamtes Programm analysiert. Dabei werden Massnahmen im Bereich der Ölfeuerungen primär kleinerer und mittlerer Leistung betrachtet. Wie bereits angesprochen, werden die lufthygienischen und energetischen Auswirkungen der genannten Vorschriften explizit nicht berücksichtigt. Weiterhin ausgeklammert bleibt die Wahl der 'richtigen' luftreinhaltepolitischen Instrumente und auch gesellschaftspolitische Fragen im Kontext der Umweltpolitik, wie etwa die von Ulrich Beck thematisierte Frage um die Ethik von Grenzwerten (Beck 1986: 85-92), werden nicht untersucht. Umgesetzt wird die Fragestellung anhand eidgenössischer, kantonaler und kommunaler Politiken seit den 1960er-Jahren.

1.2. Die Fragestellung

Ausgehend von der Grundfrage nach den Auswirkungen der Umweltpolitik auf die technische Entwicklung, entwickeln die Autoren eine dreiteilige Fragestellung, welche sich auf die drei untersuchten Massnahmen bezieht. Das zugrundeliegende Wirkungsmodell geht davon aus, dass die Zusammenhänge zwischen Grenzwertsetzung und Invention, Typenprüfung und Innovation sowie zwischen Feuerungskontrolle und Diffusion von Bedeutung sind (Darstellung 2.2., Schlussbericht (SB): 20). Daraus ergeben sich für die Untersuchung der Einzelmassnahmen folgende Teilfragen:

Grenzwertsetzung <--> Invention

"In unserer Fallstudie fragen wir nach dem Wirkungszusammenhang zwischen der politischen Implementation dieses Grenzwertes und der technischen Invention zur Erreichung der geforderten Abgasqualität. Führte die Festlegung des neuen Grenzwertes erst zu den entsprechenden Entwicklungsbemühungen oder erlaubte umgekehrt die Entdeckung neuer Techniken zur Ver-

minderung schädlicher Abgase die Festschreibung der entsprechenden Emissionsvorschriften? Anders formuliert: Hinkt die Luftreinhaltepolitik im Bereich der Feuerungen dem technisch Machbaren hinten nach oder treibt erst sie die Unternehmen zu den notwendigen Forschungs- und Entwicklungsbemühungen?" (SB: 29) Kurz gefasst also: "Hat die politische Massnahme der Grenzwertsetzung den Prozess der Invention von Ölbrennern und Heizkesseln beeinflusst?" (SB: 29)

Typenprüfung <--> Innovation

"Welcher Wirkungszusammenhang ist "... zwischen der Typenprüfung für Ölbrenner und Heizkessel und der Vermarktung neuer Heizungen zu beobachten?" (SB: 58)

Feuerungskontrolle <--> Diffusion

"Wie beeinflusst die Feuerungskontrolle die Diffusion neuer Heizungen?" (SB: 84)

Weiterhin fragen die Autoren der Studie auch nach dem Gesamteffekt des - aus den drei Massnahmen bestehenden - politischen Programms auf die technische Entwicklung als Ganzes.

1.3. Das Vorgehen

Der oben dargestellten Unterteilung folgend, bearbeiten die Evaluatoren die Thematik mit unterschiedlichen Vorgehensweisen, die sie als "empirische Zugänge" bezeichnen und welche unterschiedlichen "Analysekonzepten" folgen (siehe SB: 18-23). Die Massnahme *Grenzwertsetzung* wird anhand einer Fallstudie untersucht. Diese baut auf einem Dokumentenstudium auf. Die Untersuchung der *Typenprüfung* erfolgt mit einer Marktanalyse, die eine "qualitative und quantitative Bearbeitung von Statistiken" (SB: 22) umfasst. Die *Feuerungskontrolle*, bei welcher als Erhebungstechnik eine Umfrage zum Zuge kommt, wird mit einem statistischen Verfahren, einer multiplen Regression, analysiert. Die Untersuchung des gesamten Programms beruht auf einer Zeitreihenuntersuchung mittels Statistiken. Zur Ergänzung sind bei allen Untersuchungsteilen Expertengespräche geführt worden.

1.4. Der Auftraggeber

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

1.5. Die Forscher

Die beiden Forscher, welche das Projekt durchgeführt haben, sind Andreas Balthasar (Jahrgang 1958) und Carlo Knöpfel (Jahrgang 1959). *Andreas Balthasar* hat an der Universität Bern Volkswirtschaft und Sozial- und Wirtschaftsgeschichte studiert und 1985 mit einer Arbeit über gewerkschaftliche Berufsverbände in der Zeit von 1880 und 1914 promoviert (Dr. rer. pol., siehe Balthasar 1988). In den Jahren 1985 und 1986 befasste sich Balthasar unter anderem mit dem Einsatz quantitativer Methoden in der historischen Forschung. Er hat zudem in verschiedenen sozialwissenschaftlichen Forschungsprojekten zur Forschungs- und Technologiepolitik in der Schweiz mitgewirkt, die mehrheitlich an der privaten Forschungsstelle für Wissenschaft und Politik (Bern) bearbeitet worden sind. Balthasar hat aber auch durch seine Tätigkeit als Projektleiter bei einem privaten Markt- und Meinungsforschungsinstitut weitere Erfahrungen erworben. Am Forschungszentrum für Schweizerische Politik an der Universität Bern, wo er bereits früher als Assistent tätig war, nahm er seit 1987 regelmässig Lehraufträge wahr. In diesem Zusammenhang ist auch ein Leitfaden für Evaluatoren entstanden (Balthasar 1993). Zu Beginn des hier zu evaluierenden Projektes hat er in Luzern das Institut für Politikstudien 'Interface' gegründet. Für diese Evaluation im Spannungsfeld zwischen Umweltpolitik und technischer Entwicklung sind drei Projekte, an denen Balthasar mitgewirkt hat, besonders erwähnenswert. Zum einen handelt es sich um die 'Evaluation der Förderung praxisorientierter Forschung' (Freiburghaus/Zimmermann/Balthasar 1990, siehe auch Freiburghaus/Balthasar/Zimmermann/Knöpfel 1991), zum zweiten ist dies die Studie mit dem Titel 'Wirtschaftliche Konsequenzen einer europäischen Harmonisierung im Umweltbereich' (Balthasar/Müller 1991) und drittens ist in diesem Zusammenhang die Evaluation der UVP im Kanton Bern zu erwähnen (Zimmermann/Balthasar/Neuenschwander 1991). Weiterhin nimmt Balthasar im Rahmen des Programms 'Energie 2000' evaluationsorientierte Beratungsaufträge wahr (vgl. dazu Balthasar 1992).

Carlo Knöpfel war vor seiner Promotion mit einer Arbeit zum 'Einfluss der politischen Stabilität auf die internationale Wettbewerbstätigkeit der Schweiz' (Dr. rer. pol., 1988, siehe Knöpfel 1988) als Assistent von Silvio Borner am Institut für Volkswirtschaft der Universität Basel tätig. Er hatte zuvor auch schon in Basel sein Ökonomie-Studium abgeschlossen. Nach der Promotion hat er am Wirtschaftswissenschaftlichen Zentrum der Universität Basel ein Projekt zu nichttarifären Handelshemmnissen zwischen der Schweiz und der EG geleitet (Knöpfel 1989). Dann hat er, ebenso wie Andreas Balthasar, an der Forschungsstelle für Wissenschaft und Politik mitgearbeitet (siehe etwa Freiburghaus/Balthasar/Zimmermann/Knöpfel 1991). Vor dem eigentlichen Projektbeginn hat er sich von der Forschungsstelle gelöst und war als selbständiger Wirtschaftswissenschaftler in Basel tätig. Heute bearbeitet Knöpfel Forschungsfragen für eine gemeinnützige Organisation im Angestelltenverhältnis.

Andreas Balthasar nahm formell die Funktion des Projektleiters wahr. Die beiden Evaluatoren haben bereits in früheren Forschungsprojekten zusammengearbeitet. Somit kann festgehalten werden, dass Balthasar und Knöpfel ein erfahrenes und auch eingespieltes Forschungsteam darstellen.

2. Beschreibung des Evaluationsprozesses

Im Januar 1990 haben die beiden Evaluatoren auf die Ausschreibung des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27) hin (SNF 1989) eine Forschungsskizze eingereicht. Die Skizze trägt den Titel 'Auswirkungen umweltpolitischer Normierungen auf die technologische Innovation - Am Beispiel der Hausfeuerungsanlagen' und bezog sich auf den im Ausführungsplan genannten Themenbereich der Innovationsförderung. Zu diesem Zeitpunkt haben beide Forscher noch an der Forschungsstelle für Wissenschaft und Politik mitgearbeitet. Mitte April 1990 fand eine Besprechung der Evaluatoren mit Werner Bussmann, dem Leiter des NFP 27, statt, an der über die Forschungsskizze diskutiert wurde. Am 8. Mai 1990 wurde die Forschergruppe aufgefordert bis zum 18. Juni 1990 ein Forschungsgesuch einzureichen, was sie dann am 12. Juni 1990 auch getan hat. Nach einer positiven Beurteilung des Forschungsgesuchs durch die Expertengruppe des NFP 27 wurde dieses durch den Nationalen Forschungsrat Ende August 1990 ohne Änderungen oder zusätzliche Auflagen bewilligt. Das Projekt mit einer Laufzeit von 24 Monaten wurde mit einem finanziellen Beitrag von 249'950.-sFr. ausgestattet. Auf Antrag der Projektbearbeiter wurde der Projektbeginn vom 1. Januar 1991 auf den 1. Oktober 1991 verschoben. Die Arbeitsteilung unter den beiden Forschern wurde von diesen so geregelt, dass Balthasar schwergewichtig an den politikwissenschaftlichen und Knöpfel an den wirtschaftspolitischen Fragen arbeitete. Die Verantwortung für die vier Untersuchungsteile wurde folgendermassen festgelegt: Knöpfel bearbeitet die Fallstudie 'Grenzwertsetzung' und die Marktanalyse zur 'Typenprüfung', Balthasar die statistische Analyse 'Feuerungskontrolle' und die Zeitreihenuntersuchung zu den Gesamtwirkungen des politischen Programms. Trotz der Arbeitsteilung verzichteten die Evaluatoren nicht auf einen regen Austausch.

In einer Besprechung mit dem Programmleiter des NFP 27, Werner Bussmann, anfangs Juni 1992, wurden kleinere Abweichungen zum im Forschungsgesuch enthaltenen Zeitplan dargelegt, Probleme im Zusammenhang mit dem Zugang zum empirischen Datenmaterial erläutert sowie ein erster Gedankenaustausch zur Form des Schlussberichts durchgeführt. Die Änderungen im zeitlichen Vorgehen betrafen die technische Fallstudie zur Grenzwertsetzung, welche erst später bearbeitet werden sollte. Im Juli 1992 lag ein Konzept zur Teilstudie Feuerungskontrolle vor. In einem Memorandum zum Untersuchungsteil 'technologische Fallstudien' vom September 1992 wurde festgehalten, dass entgegen den Angaben im Forschungsgesuch nicht zwei, sondern nur eine Fallstudie erarbeitet werden soll. Im gleichen Monat erarbeiteten die Evaluatoren auch erste Fassungen eines Konzepts zum Aufbau des Schlussberichts.

Vermutlich im Oktober 1992 (die Unterlagen sind nicht datiert) wurde der wissenschaftliche Zwischenbericht zum Projekt verfasst. Darin orientieren die Autoren über den Stand der Arbeiten und führen erste Ergebnisse auf. Die Untersuchungsteile 'Feuerungskontrolle' und 'Typenprü-

fung' waren bereits relativ weit fortgeschritten. Wie angekündigt hatte sich die Analyse der 'Grenzwertsetzung' etwas verschoben.

Nach Ablauf des ersten Projektjahres hatten die Forscher gemäss dem finanziellen Zwischenbericht vom 31. Oktober 1992 knapp die Hälfte der zur Verfügung stehenden Mittel aufgebraucht. Im November 1992 lag eine weitere Version eines Konzeptes zum Schlussbericht vor. An einer Tagung im Rahmen des NFP 27 am 14. Januar 1993 in Bern hielt Andreas Balthasar ein Referat mit dem Titel 'Wirkungserfassung durch Aktoren bzw. unabhängig von den Aktoren'. Darin legte er dar, welche Probleme sich bei der Erfassung der Wirkungen der politischen Massnahmen im Projektverlauf bisher gezeigt hatten. Gegen Ende Januar 1993 lag sodann ein weiterer Entwurf zur Konzeption des Schlussberichts vor. Der darin vorgeschlagene Aufbau entsprach nun bereits weitgehend der Form des endgültigen Schlussberichts. Im Februar 1993 hielten die Evaluatoren ein weiteres Referat, das in direktem Zusammenhang zum Evaluationsprojekt stand. Am Forschungszentrum für Schweizerische Politik referierten sie zum Thema 'Triangulation'. Nach einer Präsentation des Projekts erläuterten sie die verschiedenen methodischen Konzeptionen, welche in der Evaluationsforschung zur Anwendung gelangen. Danach gingen sie auf die Triangulation ein. Ende März 1993 hielten die Forscher in Kiel ein Referat zu den 'Methoden der Evaluation' und führten beim Umweltamt Hamburg sowie beim Umweltbundesamt in Berlin eine Projektpräsentation durch. Am 5. Juli 1993 legten die Evaluatoren der Programmleitung und der Expertengruppe des NFP 27 den Entwurf ihres Schlussberichts vor. Mitte August 1993 fand sodann in Bern eine Besprechung mit Mitgliedern der Expertengruppe NFP 27 (Ulrich Klöti, Gerhard Schmid), mit dem Programmleiter NFP 27 (Werner Bussmann) und den beiden Forschern statt. In dieser Sitzung wurde der Schlussbericht generell für gut befunden und akzeptiert. In einigen Punkten wünschten die Experten und der Programmleiter noch Retuschen am Schlussbericht.

Im September 1993 führten die beiden Evaluatoren ein Gespräch mit Vertretern des Amtes für Umweltschutz des Kantons Luzern, welches sich um die Resultate der Evaluation drehte. Am 3. November 1993 wurde das Projekt mit einem Brief Balthasars an die Abteilung NFP des Schweizerischen Nationalfonds formell abgeschlossen. In diesem Schreiben orientierten die Evaluatoren über ihre bisherigen Publikations- und Umsetzungstätigkeiten im Rahmen des Forschungsprojekts. Für den inhaltlichen Schlussbericht wurde auf das beim Verlag Helbing & Lichtenhahn erscheinende Buch verwiesen. Im Rahmen eines Weiterbildungskurses für Gemeindebeamte am Forschungszentrum für Schweizerische Politik an der Universität Bern, gestaltete Andreas Balthasar im Dezember 1993 einen ganztägigen Kurs zum Thema 'Der Umgang mit Experten'. Am 12. April 1994 veröffentlichte die NFP-Pressestelle eine Pressemitteilung im Zusammenhang mit der Publikation des nun erschienen Schlussberichts in Buchform. Im Laufe des Jahres 1994 führte Andreas Balthasar zusammen mit dem Programmleiter des NFP 27, Wer-

ner Bussmann, an der Universität Bern eine mehrtägige Weiterbildungsveranstaltung zur Evaluation durch.

3. Bewertung entlang der Kriterien

Da aus dem Schlussbericht der Evaluation von Andreas Balthasar und Carlo Knöpfel nicht klar hervorgeht, welche paradigmatische Position von den Autoren eingenommen wird, können wir nur indirekt darauf schliessen. Es finden sich in der Arbeit keinerlei Hinweise auf Texte oder Vertreter einer 'fourth generation evaluation'. Das Konzept der Triangulation verlangt nicht nach einem alternativen Paradigma, sondern ist auch im traditionellen Kontext üblich (so etwa Campbell/Fiske 1959, aber auch Denzin 1970, 1978a und 1978b, nicht aber 1989, zur Einordnung siehe Guba/Lincoln 1989: 162). Deshalb verzichten wir bei dieser Studie auf den Beizug der alternativen und stützen uns auf die konventionellen Kriterien ab.

3.1. Nützlichkeit

3.1.1. Identifikation der beteiligten Gruppen

Die Autoren gehen davon aus, dass das weite Feld der beteiligten Akteure in drei Gruppen unterteilt werden kann, nämlich in Wirtschaft, Wissenschaft und Staat (vgl. dazu und zum folgenden SB: 18-19, insbesondere Darstellung 2.1.).

Wichtigste Akteure aus dem Bereich der *Wirtschaft* sind die Produzenten von Heizungsanlagen (wobei die Hersteller von Brennern und die der Kessel deutlich zu unterscheiden sind), die Heizungsimporteure und -installateure, sowie die Hausbesitzer.

Bei der *Wissenschaft* betonen die Autoren zum einen die Rolle der Grundlagenforschung und zum zweiten jene Personen und Institutionen, welche sich im Bereich der Forschung und Entwicklung von Brennern und Kesseln (bzw. Kompositionen aus diesen) engagieren, wie auch die mit der Forschung und Entwicklung von entsprechenden Messgeräten betrauten.

Als wichtige *Institutionen im staatlichen Bereich* heben die Evaluatoren verschiedene Bundesämter (Bundesamt für Umwelt, Wald und Landschaft (BUWAL), Bundesamt für Energiewirtschaft (BEW) und das Eidgenössische Amt für Messwesen (EAM)), die Eidgenössische Materialprüfungs- und Forschungsanstalt (EMPA) sowie die Vielzahl von kantonalen und kommunalen Behörden hervor.

Eine Identifikation der beteiligten Gruppen ist somit also erfolgt. Die Studie baut nicht direkt auf dem Konzept der 'stakeholders' auf. Trotzdem erhalten in verschiedenen Teilen der Studie die spezifischen Interessenlagen der Beteiligten oder Betroffenen eine recht grosse Bedeutung. Besonders gilt dies für den Teil der Studie, in dem die Grenzwertsetzung mit einer Fallstudie analysiert wird (siehe Kapitel 3, SB: 28-56). Auf eine 'thick description' der Ereignisse erfolgt eine

Aktoranalyse, die nun jedoch das Untersuchungsfeld in vier Gruppen von Akteuren differenziert. Neben die Wissenschaft, die Wirtschaft und die Behörden treten nun zusätzlich die Interessenverbände, besonders die Vertreter der automobilen Interessen (Touring Club der Schweiz (TCS), Autopartei oder die City-Vereinigung Zürich) und die Umweltorganisationen. Letztere konnten jedoch kaum in den Entscheidungsprozess eingreifen (SB: 53-54). Die mit dem ökonomischen Kosten-Nutzen-Kalkül argumentierende Aktoranalyse ist aber nicht der einzige Ort in der Evaluation, an dem Interessen der 'stakeholder' diskutiert werden. Bis auf marginale Ausnahmen (Anmerkung 88, SB: 177) werden dabei aber die Interessen einer Gruppe von Betroffenen ausgeblendet. Die Anliegen der Hausbesitzer (bzw. der Mieter) nämlich, welche das ganze Programm zu finanzieren haben, werden weder im Aushandlungsprozess in der politischen Arena noch in der Evaluation berücksichtigt. Dem Kostenfaktor, etwa bei der Implementation der Feuerungskontrolle oder bei der Erhöhung der Anschaffungspreise für Low-NO_x-Heizanlagen, wird wenig Beachtung geschenkt. Die Evaluation negiert jedoch finanzielle Überlegungen nicht generell. Beispielsweise die Aufwendungen der Heizungsbranche für die Forschung und Entwicklung neuer Heizungstechnologien werden in der Studie behandelt. Da sich die Anlagenbetreiber dem Konsens der beteiligten Gruppen angeschlossen haben, kommt dieser Lücke jedoch keine allzu grosse inhaltliche Bedeutung zu.

Das Kriterium ist bis auf die erwähnten Lücken erfüllt.

3.1.2. Glaubwürdigkeit des Evaluators

Die Evaluatoren sind sowohl von ihrer Ausbildung wie auch von ihrem Erfahrungsschatz her für schweizerische Verhältnisse als kompetent zu bezeichnen. Sie können sich im Themengebiet über verschiedene Erfahrungen ausweisen. Die Förderung des Evaluationsprojekts durch den Nationalfonds verleiht den Evaluatoren zusätzliche Glaubwürdigkeit. Bei der Beschaffung von Datenmaterial zu Struktur und Entwicklung der Heizungsbranche für die Marktanalyse sind die Evaluatoren nicht besonders erfolgreich gewesen. Ob dabei eventuell die mangelnde Glaubwürdigkeit der Evaluatoren bei den Branchenvertretern von Bedeutung gewesen ist, kann aufgrund des vorliegenden Materials nicht beurteilt werden. Möglicherweise ist die Branche ja wirklich so schlecht dokumentiert gewesen (vgl. die Hinweise in Kapitel 3.1.3.). Soweit dies beurteilbar ist, verfügen die Evaluatoren über eine hohe Glaubwürdigkeit und Kompetenz. Das Kriterium ist erfüllt.

3.1.3. Informationsrahmen und -auswahl

Hier sind die verschiedenen Untersuchungsteile differenziert zu untersuchen.

a) Fallstudie 'Grenzwertsetzung'

Die Datenerhebung für die Fallstudie erfolgte mit einem Dokumentenstudium und Expertengesprächen. Die Evaluatoren konzentrierten sich auf die Nachzeichnung der Entstehungsgeschichte der Low-NO_x-Technologie beziehungsweise des NO_x-Emissionsgrenzwertes von 120 mg/m³ für Hausfeuerungen der untersuchten Leistungsklasse von Ölfeuerungen (die sogenannte Züri-Norm). Die aufgrund der Recherchen erfassten Informationen sind in eine Chronologie der Ereignisse mit einer Vorgeschichte und insgesamt 44 Sequenzen eingeflossen (SB: 34-47). Aufgrund dieser 'thick description' erfolgte die Aktoranalyse, auf die wir hier schon hingewiesen haben (siehe oben Kapitel 3.1.1.). Sowohl die Prozessbeschreibung wie auch die Aktoranalyse sind sehr informativ und für die Beantwortung der zur Debatte stehenden Fragen wertvoll. Trotzdem sind hierzu auch zwei kritische Anmerkungen anzuführen. Erstens wird aus dem Schlussbericht nicht ganz klar, weshalb sich die Evaluatoren gerade auf die Entwicklung der Züri-Norm konzentriert haben. Zudem wird im Forschungsgesuch von *zwei* verschiedenen Fallstudien gesprochen (siehe Kapitel 3.3.1.). Zweitens ist aufgrund der Angaben im Schlussbericht nicht ersichtlich, auf welche Quellen sich die Autoren stützen und wie sie bei der Erarbeitung der 'thick description' vorgegangen sind. Wir werden weiter unten nochmals ausführlicher auf diese Punkte eingehen (siehe Kapitel 3.4.3. und 3.4.10.).

b) Marktanalyse 'Typenprüfung'

Wie bereits erwähnt (vgl. oben Kapitel 3.1.2.), sind die Evaluatoren bei der Informationsbeschaffung zu diesem Untersuchungsteil auf unerwartet grosse Schwierigkeiten gestossen. Die für die Marktanalyse notwendigen Grundlagendaten konnten von den Evaluatoren nicht erhoben werden. Wie sie korrekterweise festhalten, ist deshalb die Teilstudie Marktanalyse eigentlich in grossen Teilen zu einer Fallstudie geworden (SB: 160). Da aus dem Schlussbericht nicht ganz klar wird, welche Datenzugangsmöglichkeiten abgeklärt worden sind, ist schwer zu beurteilen, ob noch zusätzliche Informationen zu beschaffen gewesen wären. Auf alle Fälle ist hier anzumerken, dass die vorhandenen Informationen nicht in allen Teilen übersichtlich aufbereitet wurden. Beispielsweise wird nicht klar, inwiefern sich die Ausgestaltung der Typenprüfung durch die drei hier relevanten Verordnungssetzungen LMFV, LRV 86 und LRV 92 verändert hat (SB: 64-68). Aufgrund der Darlegungen zur Marktstruktur fällt es dem Leser zum Teil schwer, diese zu überblicken (SB: 73-74). So folgt einem Hinweis auf die Übernahme der Firmen Cuénod Thermotechnique und Klöckner-Wärmetechnik durch Elco Looser (SB: 73) auf der nächsten

Seite des Berichts die Aussage, dass Elco Looser die Unternehmen Cuénod und *Brennwald* übernommen haben soll. Auch die Aussage: "In der Schweiz wird mit dem Verkauf von Heizanlagen ein von der Branche selber geschätzter Umsatz von 300 bis 500 Millionen Franken jährlich verdient" lässt Fragen offen: Umsatz oder Gewinn? Nur Handel oder auch Herstellung? Nur schweizerische Unternehmen oder auch ausländische Anbieter? (SB: 79)

Auf dem Hintergrund der Hinweise der Autoren, dass die Marktanalyse stark durch die schlechte Datenlage tangiert wurde (so SB: 158-160) überrascht, dass zugängliche Informationen nur teilweise genutzt wurden (weitere Informationen in Wüest & Partner 1992, vgl. SB: 182; Estermann/Feldmann 1990, vgl. FG: 3.1.). Die Auswirkungen der Qualität des Datenmaterials, auf dem die Analyse aufbaut, werden nur knapp angesprochen (SB: 159). Bei den von Wüest & Partner angeführten Daten handelt es sich beispielsweise zumeist um Hochrechnungen (vgl. Wüest & Partner 1992).

Der Informationsgehalt der Marktanalyse geht nicht weit über das hinaus, was bereits in anderen Studien erarbeitet wurde (vor allem Balthasar/Müller 1991: 8-45 und Wüest & Partner 1992).

c) Statistische Analyse 'Feuerungskontrolle'

Die Informationen zur Bearbeitung des dritten Untersuchungsteils wurden mittels Umfrage und Expertengesprächen erhoben. Dieser Teil umfasst einen deskriptiv-qualitativen Teil zu historischen und juristischen Aspekten der Feuerungskontrolle einerseits und eine quantitative Querschnittsanalyse auf der Ebene der Gemeinden andererseits.

Im deskriptiven ersten Teil (Kapitel 5.1. und 5.2., SB: 84-99) wird die Geschichte der Feuerungskontrolle aufgearbeitet und deren Vollzug in drei verschiedenen Kantonen beschrieben. Diese drei Kantone stehen gemäss den Autoren für drei verschiedene Vollzugsmodelle, welche als 'Luzerner Modell' (Kanton Luzern), 'Modell Romandie' (Kanton Waadt) und 'Basler Modell' (Kanton Basel-Stadt) bezeichnet werden. Anschliessend erfolgt eine Aufarbeitung entlang 'administrativer' und 'personeller Faktoren'. Dieser Teil erlaubt dem Leser eine Orientierung über die Feuerungskontrolle und deren Implementation. Der generell sehr positive Gesamteindruck ist durch zwei Anmerkungen zu ergänzen:

- *Ungenauigkeiten:* In die Beschreibung haben sich einige Ungenauigkeiten eingeschlichen. So trat die Verordnung über Luftreinemassnahmen bei Feuerungen (LMFV, AS 1984: 1516) nicht auf Mitte 1985 (SB: 84 und 86), sondern bereits auf Anfang 1985 in Kraft. Da damit erstmals auf Bundesebene eine Pflicht zur Durchführung einer Feuerungskontrolle verankert wurde, ist dieses Datum für die Thematik von grosser Bedeutung.

- *Punktuelle Lücken in der Informationsbeschaffung:* Die Feuerungskontrolle ist in einer Grosszahl von Gemeinden und auch in vielen Kantonen schon vor der Festlegung auf Bundesebene eingeführt worden. Die Stadt Zürich hat dabei sicherlich eine Vorreiterrolle gespielt. Darauf weisen die Autoren auch deutlich hin. Was sie aber unterlassen, und was für den Leser sicherlich von grossem Interesse gewesen wäre, ist eine Übersicht zu welchen Zeitpunkten in welchen Kantonen (bzw. Gemeinden) eine Feuerungskontrolle eingeführt wurde. Sicherlich ist es nur schon aus forschungsökonomischen Gründen legitim, sich für die detaillierte Analyse der Vollzugsdifferenzen auf drei typische Fälle (i.e. Kantone) zu konzentrieren.

Grundsätzlich ist auch hier anzumerken, dass das konkrete Vorgehen bei der Datengewinnung kaum transparent gemacht wird. So beschränken sich die Angaben zur Umfrage darauf, dass eine solche durchgeführt worden sei. Es bleibt im Dunkeln, wer in dieser Umfrage befragt wurde und welche Informationen man damit beschaffen wollte bzw. beschafft hat. Es wird auch nicht klar, in welcher Form diese Umfrage stattgefunden hat. Die eigentliche Vorgehensweise präsentiert sich als 'black box'.

Die quantitative Analyse (Kapitel 5.3., SB: 99-114) beschränkt sich auf das sogenannte 'Luzerner Modell', das anhand der Kantone Luzern, Bern und Basel-Land, untersucht wurde. Die Autoren erreichen damit zweierlei:

- Erstens wird der Aufwand zur Datenbeschaffung auf ein bewältigbares Mass reduziert. In Kantonen, welche nach diesem Modell vollziehen, besteht eine kantonale Stelle, die über die notwendigen Informationen verfügt.
- Zweitens sind die drei Untersuchungskantone aufgrund des gleichen Vollzugsmodells besser miteinander vergleichbar, als dies bei einem Vergleich zwischen den verschiedenen Vollzugsmodellen möglich gewesen wäre. Wir stimmen mit den Autoren aber überein, dass ein Vergleich der Modelle sicherlich eine Fragestellung darstellt, deren Bearbeitung lohnenswert wäre (SB: 161).

Die Auswertung des Datenmaterials zu insgesamt 419 der 592 Gemeinden aus den drei ausgewählten Kantonen erfolgt mit einer multiplen Querschnittsregression. Die darin integrierten Variablen sind aufgrund der qualitativen Vorarbeiten ausgewählt worden. Ausgehend von ad-hoc-theoretischen Überlegungen werden Wirkungshypothesen generiert, aufgrund derer ein Regressionsmodell formuliert wird. Der Modernisierungsgrad des Heizanlagenbestandes in einer Gemeinde bildet die abhängige Variable. Die relevanten Variablen sind zuerst univariat analysiert worden. Die gewonnenen Daten erscheinen - soweit dies beurteilt werden kann - ausreichend zu sein. Die getroffene Variablenauswahl verfügt über eine hohe Plausibilität. Auf der Ebene der Gemeinden standen nicht für alle Gemeinden der drei ausgewählten Kantone Daten zur Verfü-

gung. Im Kanton Bern sind von den insgesamt 412 Gemeinden lediglich 247 in der Analyse berücksichtigt. Die Autoren geben an, dass es sich dabei "um eine weitgehend repräsentative Auswahl" (SB: 103) handelt, ohne dies weiter zu begründen. Auf die Auswertung werden wir zu einem späteren Zeitpunkt noch genauer zu sprechen kommen (siehe unten, Kapitel 3.4.8.).

d) Zeitreihenuntersuchung 'Gesamtes Programm'

Das Gesamtprogramm wird mit zwei unterschiedlichen Perspektiven analysiert. Einerseits erfolgt eine echte Längsschnittanalyse am Beispiel der Entwicklung in der Stadt Zürich im Zeitraum von 1964 und 1991. Andererseits wird für die Stadt Luzern eine Analyse des Heizanlagenbestandes im Jahre 1991 durchgeführt, bei der dem Alter der Heizanlagen besondere Beachtung geschenkt wird. Warum gerade diese beiden Städte untersucht werden, bleibt unklar. Es ist anzunehmen, dass die Auswahl durch die Datenlage bestimmt wurde.

Das beschaffte Datenmaterial aus der *Stadt Zürich* weist verschiedene Lücken auf. Die fehlenden Daten scheinen nicht mehr zu existieren, oder der Aufwand für deren Erhebung wäre unverhältnismässig gewesen. Das vorhandene Datenmaterial wird im Schlussbericht nicht ausreichend detailliert beschrieben. Da die Feuerungskontrolle in der Stadt Zürich vor allem im Winter erfolgt und die Autoren jeweils nur ein Jahr angeben, bleibt offen, ob das Datum '1964' nun für die Kontrollperiode 1963/64 oder 1964/65 steht. Auch die Umstände, unter denen die Daten zustande gekommen sind, bleiben im Dunkeln: Sind nur Routinekontrollen berücksichtigt? Oder auch Nachkontrollen? Und Kontrollen aufgrund eingegangener Klagen? Punktuell finden sich auch Widersprüche. So geht aus einer chronologischen Übersicht (Darstellung 6.1., SB: 117) hervor, dass 1982 in der Stadt Zürich ein Obligatorium für die Wirkungsgradermittlung eingeführt worden sei. Zwei Seiten später wird diese auf das Jahr 1981 datiert, was aufgrund weiterer Quellen korrekt ist (vgl. SVG, Nr. 84: 3).

Das Datenmaterial aus der *Stadt Luzern* ist vollständig verfügbar. Es handelt sich dabei um Angaben zur Kontrollperiode des Jahres 1991 (1990/91, 1991/92 oder effektiv das Kalenderjahr 1991?). Dieses Datenmaterial wird aufgeschlüsselt nach den Jahrgängen der Heizanlagen. Dabei fällt auf, dass eine Grundannahme, von der die Autoren ausgehen, nicht generell zutrifft. Die Prämisse lautet: "Technische Entwicklungen im Bereich der Heizungen haben zur Folge, dass neuere Anlagen geringere Schadstoffemissionen aufweisen als ältere und dass sie weniger Energie verbrauchen." (SB: 115) Betrachtet man das Datenmaterial aus der Stadt Luzern, wird deutlich, dass diese Feststellung zumindest für die NO_x-Emissionen nicht in allen Fällen zutrifft. Die Kesseljahrgänge 1960, 1963, 1965 und 1968 weisen durchschnittlich tiefere NO_x-Emissionen auf, als die Jahrgänge 1990 und 1991. Bei den Abgasverlusten stellen sich ähnliche Probleme. Der Jahrgang 1991 (der jüngste Jahrgang in der Auswertung) weist höhere Abgasverluste auf, als

die Jahrgänge 1987 bis 1990. Da das Alter der Heizanlagen in der Studie häufig als Indikator Verwendung findet, stellt das Luzerner Datenmaterial Teile der quantitativen Auswertung in Frage. Auf alle Fälle kann das Alter nicht uneingeschränkt als valider Indikator für die energetische oder lufthygienische Qualität einer Heizung angesehen werden (vgl. auch Kapitel 3.4.5.). Das Kapitel zur Teilstudie 'Gesamtes Programm' schliesst mit einer illustrativen qualitativen Beschreibung des Kommunikationsprozesses im Rahmen des diskutierten Gesamtprogramms.

Betrachtet man die gesamte Studie, ist es den Evaluatoren sicherlich geglückt, eine Vielfalt von Informationen zur Bearbeitung ihrer Fragestellung zu erheben. Dazu hat sicherlich der eingesetzte Methodenmix massgeblich beigetragen. Der Informationsgehalt ist nicht in allen Untersuchungsteilen gleich hoch. Während die Untersuchungsteile 'Fallstudie Grenzwertsetzung', 'Statistische Analyse Feuerungskontrolle' und 'Zeitreihenuntersuchung Gesamtes Programm' in dieser Dimension sehr gut abschliessen, ist dies bei der Teilstudie 'Marktanalyse Typenprüfung' nur beschränkt der Fall. Hier stiessen die Evaluatoren auf Probleme der mangelnden Datenverfügbarkeit, die den zu erreichenden Informationsgehalt massgeblich einschränkten. In einigen wenigen Punkten zeigen sich bei den dargelegten Informationen gewisse Ungenauigkeiten. Aufgrund der Reichhaltigkeit, welche die Studie mittels einer Kombination unterschiedlicher Ansätze erreichen konnte, wird das Kriterium des Informationsrahmens und der Informationsauswahl aber klar erfüllt.

3.1.4. Bewertende Interpretationen

Die Interpretationen der Evaluation werden durch die Forscher klar dargelegt und begründet. Mehrheitlich wird dieses Kriterium erfüllt. Punktuell vermögen jedoch die Interpretationen nicht voll zu überzeugen. Besonders in jenen Punkten nicht, an denen die Autoren über ihre eigentliche Fragestellung hinaus Folgerungen formulieren. So fehlt diesen Folgerungen teilweise eine empirische Grundlage. Dass eine Dynamisierung der Grenzwerte, also eine "periodische Verschärfung der Auflage um einen im voraus festgelegten Prozentsatz" (SB: 146) zu einer verstärkten Entwicklungstätigkeit der Heizungsbranche führt und das Auftragsvolumen dieser Branche sicherstellt, ist durch die empirische Analyse gestützt. Dass damit eine "Verringerung der Abgase und eine effektivere Nutzung knapper Energie" (SB: 147) erreicht werden kann, ist aufgrund dieser Analyse empirisch nicht zu stützen. Technische Verbesserungen führen nicht automatisch zu einer Verringerung der Abgase und einer effektiveren Energienutzung. Für diesbezügliche Aussagen ist der ganze Produktions-, Nutzungs- und Entsorgungsprozess sowie der Schadstoffausstoss zu betrachten (Stichworte 'sinkender Grenznutzen' und 'graue Energie'). Dies tun die Evaluatoren jedoch nicht.

Insgesamt betrachtet werden die Grundlagen für die Werturteile jedoch klar dargelegt. Deshalb kommen wir zum Schluss, dass die Evaluation das Kriterium mehrheitlich erfüllt.

3.1.5. Klarheit des Berichts

Der Schlussbericht der hier diskutierten Evaluationsstudie von Andreas Balthasar und Carlo Knöpfel besticht durch seine Klarheit. Das gilt sowohl für den Aufbau der Studie, wie auch für die Sprache. Aber auch die Hinweise, die den Leser durch die Studie begleiten, sind lobenswert zu erwähnen. Bis auf einige Details ist der Text sauber redigiert. Ein Abkürzungsverzeichnis, ein Verzeichnis wichtiger Rechtsquellen und ein Literaturverzeichnis erleichtern die Orientierung. Tabellen und Abbildungen sind von hoher Qualität und werden zweckmässig eingesetzt.

Wie wurde der Schlussbericht strukturiert? Im Gegensatz zu anderen Schlussberichten, welche im Rahmen des NFP 27 entstanden sind, wird der Bericht nicht strikte in einen inhaltlichen und einen methodischen Teil aufgetrennt. Der Aufbau der Studie gestaltet sich folgendermassen:

Erster Teil: Ausgangspunkte (SB: 1-23)

1. Technische Entwicklung als Nebenwirkung der Umweltpolitik? (SB: 4-15)
2. Wirksamkeit staatlicher Massnahmen (SB: 16-23)

Zweiter Teil: Empirische Zugänge (SB: 25-131)

3. Grenzwertsetzung oder: Die Politik und die Entwicklung neuer Technologien - Eine Fallstudie (SB: 28-56)
4. Typenprüfung oder: Die Politik und die Vermarktung neuer Technologien - Eine Marktanalyse (SB: 57-83)
5. Feuerungskontrolle oder: Die Politik und die Verbreitung neuer Technologien - Eine quantitative Analyse (SB: 84-114)
6. Grenzwertsetzung, Typenprüfung und Feuerungskontrolle als politisches Programm (SB: 115-131)

Dritter Teil: Interpretationen (SB: 133-171)

7. Wirtschaftspolitische Schlussfolgerungen: Die Wirkung umweltpolitischer Massnahmen auf die technische Entwicklung (SB: 136-154)
8. Evaluation - Ein Lernprozess (SB: 155-171)

Anhänge (SB: 173-184)

Sowohl der erste wie auch der dritte Teil des Berichts enthalten Überlegungen zur Evaluationsmethodik. Den Autoren gelingt mit dem gewählten Aufbau eine klare Strukturierung ihrer doch recht umfangreichen Studie. Die einzelnen Kapitel werden jeweils mit einem Überblick über den Inhalt des Kapitels eingeleitet. Was im Schlussbericht selbst fehlt, ist eine Zusammenfassung der wichtigsten Aussagen der Studie. Eine solche ist jedoch in der Reihe 'Kurzberichte'

des NFP 27 erschienen (siehe KB), sodass sich Praktiker, welche ja zumeist als Adressaten solcher Kurzzusammenfassungen gelten, in dieser Form innerhalb kurzer Zeit über die Grundzüge des Projekts informieren können.

Trotz dem sehr positiven Gesamteindruck, möchten wir noch exemplarisch auf einige Details zu sprechen kommen.

- Fundstellen offizieller Texte: Die beispielsweise im Rechtsquellenverzeichnis angegebenen Fundstellen führen nicht immer zum gesuchten Text. Botschaften oder Richtlinien etwa werden üblicherweise nicht in der Amtlichen Sammlung (AS) veröffentlicht, sondern im Bundesblatt (BBl). (vgl. SB: 183)
- Unklarheiten bei der Schilderung der Angebotsstruktur und der Marktanteile (SB: 73-76, vgl. oben Kapitel 3.1.3.).
- Was ist mit einer qualitativen Bearbeitung von Statistiken gemeint? Versteht man darunter eine Interpretation in Worten? (SB: 22)

Da diese Aspekte geringfügiger Art sind, der Bericht generell aber über eine hohe Klarheit verfügt, sehr übersichtlich angelegt ist und sauber redigiert wurde, kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt.

3.1.6. Verbreitung des Berichts

Der Schlussbericht zu diesem Projekt ist in *Buchform* beim Verlag Helbing & Lichtenhahn erschienen und kann im Buchhandel zu einem Preis von 59.- sFr. erworben werden. Die Autoren haben dem Verlag den Text im Oktober 1993 übergeben. Das Buch ist im April 1994 erschienen. Neben dem ausführlichen Schlussbericht haben die Autoren eine *Zusammenfassung* des Berichts erstellt, welche im Rahmen der 'Kurzberichte' im Eigenverlag des NFP 27 publiziert wurde und die gratis an Interessenten abgegeben wird. Die Evaluatoren haben in Zusammenarbeit mit der NFP-Pressestelle eine *Pressemitteilung* verfasst, um auf die Herausgabe des Schlussberichts in Buchform hinzuweisen (zu den Reaktionen, vgl. Kapitel 3.1.8.). Die Evaluation erfüllt dieses Kriterium vollständig.

3.1.7. Rechtzeitigkeit des Berichts

Da diese Evaluationsstudie im Auftrag des Schweizerischen Nationalfonds (SNF) erstellt wurde, kann die Rechtzeitigkeit des Berichts in zweierlei Hinsicht beurteilt werden, nämlich bezüglich des SNF, bzw. der Expertengruppe des Nationalen Forschungsprogramms 27 (NFP 27) oder hinsichtlich des politischen Umfelds in dem sich die Studie bewegt.

Die aktuelle Debatte in der Umweltpolitik ist geprägt durch das Ringen um die Einführung von marktwirtschaftlich orientierten Massnahmen wie Lenkungsabgaben, welche die Gesellschaft mit finanziellen Anreizen zu einem umweltverträglicheren Verhalten führen sollen. Hinzu kommt, dass auch die Umweltpolitik dem Druck einer Deregulierungstendenz ausgesetzt ist, wie er in praktisch allen politischen Bereichen festzustellen ist. Aus diesen Debatten können zwei Hauptkritikpunkte an einer regulativen Umweltpolitik herauskristallisiert werden:

1. Traditionelle Regelungen sind im ökonomischen Sinn nicht effizient, das heisst sie sorgen nicht dafür, dass die Umweltausgaben dort erfolgen, wo sie am meisten zur Verbesserung der Umweltqualität beitragen könnten. Deshalb soll die regulative Politik durch ein marktkonformes Instrumentarium abgelöst werden.
2. Regulative Politiken verschlechtern die wirtschaftlichen Rahmenbedingungen einer Volkswirtschaft, da die unternehmerischen Freiheiten der Wirtschaftssubjekte eingeschränkt werden. Deshalb sollen die Regulative zum Schutz der Umwelt gelockert oder aufgehoben werden.

Die traditionellen, regulativ angelegten und zudem noch technisch ausgerichteten Massnahmen im Umweltbereich, wie sie von Balthasar und Knöpfel untersucht worden sind, sind somit harter Kritik ausgesetzt. Die zurzeit gültige Version der Luftreinhalteverordnung (LRV, SR 814.318.142.1), in der die in der Studie prominent behandelte sogenannte Züri-Norm festgehalten ist, ist seit dem 1. Februar 1992 in Kraft (AS 1992 124). Im Gegensatz zum Umweltschutzgesetz (USG, SR 814.01), dessen Revision zurzeit im Parlament beraten wird (NZZ, 3. Juni 1994: 13 und 17, TA, 3. Juni 1994: 7), sind unseres Wissens keine Änderungen der LRV vorgesehen, soweit sie nicht durch eine in Zukunft gültige Fassung des USG notwendig werden. Die Studie von Balthasar und Knöpfel verfügt aber auch darüber hinaus über praktische Relevanz. Erstens sind hier die formativen Elemente der Studie zu nennen, die zu einer Verbesserung der Implementationsqualität einen Beitrag leisten könnten. Andererseits weist die Studie auf die Bedeutung traditioneller Politiken im Umweltbereich hin und stellt die heute quasi 'common sense' darstellende Aussage in Frage, dass regulative Politiken innovationshemmende Wirkungen entfalten und der positiven Entwicklung einer Volkswirtschaft abträglich sind.

Bezüglich des politischen Kontextes, möchten wir noch kurz auf einen zweiten Punkt zu sprechen kommen. Die vorliegende Untersuchung ist eine ex post-Analyse. In der Studie zeigt sich nun verschiedentlich, dass die 1992er Revision der LRV im Prinzip noch nicht reif für eine ex post-Evaluierung war. Das erhobene Datenmaterial konnte nur selten Auskunft über die wirtschaftlichen Auswirkungen der LRV 92 geben. Dieselbe Problematik stellte sich den Evaluatoren auch bei den durch Gemeinden (Städte) oder Kantone aufgrund der Massnahmenpläne erlassenen Vorschriften zur Sanierung von älteren Heizungsanlagen. Die Evaluatoren haben sich

in jenen Untersuchungsteilen, in denen sich dies aufdrängte, darauf konzentriert, Vorläuferregelungen genauer zu untersuchen. Das Problem der Evaluierbarkeit sich wandelnder politischer Programme kann sicherlich nicht den Evaluatoren angelastet werden.

Hinsichtlich des Auftraggebers, also des SNF, vertreten durch die Expertengruppe des NFP 27, kann hier festgehalten werden, dass die Berichterstattung sicherlich rechtzeitig erfolgt ist. Die Mitglieder der Expertengruppe und die Programmleitung verfügten bereits im Juli 1993 über den Schlussbericht in Form eines Entwurfs. Die Evaluatoren haben dem Buchverlag das endgültige Manuskript im Oktober 1993 übergeben. Da das Projekt bis Ende September 1993 lief - Verlängerungen der Laufzeit gab es keine - ist der Schlussbericht rechtzeitig fertiggestellt worden.

Zusammenfassend lässt sich festhalten, dass die Evaluation dem Kriterium der Rechtzeitigkeit nachgekommen ist.

3.1.8. Wirkung der Evaluation

Die Wirkung dieser Evaluationsstudie zu erheben, ist eine äusserst schwierige Angelegenheit. Wir wollen uns hier darauf beschränken zu untersuchen, inwiefern sich die Autoren bei der Gestaltung der Evaluation und des Schlussberichts darum bemüht haben, das Wirksamkeitspotential ihrer Studie zu steigern und welche direkten und beobachtbaren Wirkungen die Studie verursachte. Da sich die Evaluation nicht auf einen partizipativen Ansatz beruft, fällt die Umsetzung im Rahmen der eigentlichen Projekterarbeitung ausser Betracht.

Wirksamkeitspotential:

Die Studie ist von ihrer Ausgestaltung her bestens geeignet, um von potentiellen Adressaten auch beachtet zu werden. Die Studie kann durch ihre klare Darstellung, ihre sprachliche Klarheit, aber auch durch ihren Aufbau überzeugen. Die Autoren ermöglichen es dem Leser, trotz der verschiedenen Ansätze und Untersuchungsteile, die Übersicht über die Studie zu wahren. Gleichzeitig wird aber auch auf die vielfältigen Interdependenzen und Schnittstellen zwischen den verschiedenen Untersuchungsteilen und Massnahmen hingewiesen. Dies gelingt unter anderem aber auch durch Kompromisse, die bei der Darlegung der empirischen Vorgehensweise eingegangen worden sind. Konkretere Angaben zum methodischen Vorgehen (Fallauswahl, Durchführung der Expertengespräche, genauere Angaben zu verwendeten Erhebungs- und Auswertungsverfahren im allgemeinen, Datenqualität etc.) fehlen fast durchwegs. Im Schlussbericht werden auch alle Irrwege und Sackgassen, auf die man sich im Verlaufe des Forschungsprozesses begeben hat, systematisch ausgespart. Dies erhöht die Klarheit und Geradlinigkeit der Berichterstattung. Die Studie büsst dadurch aber auch an Transparenz ein. Die Evaluatoren haben sich bei der Gestaltung des Berichts dafür entschieden, der Verständlichkeit und Zugänglichkeit

für ein weiteres Publikum grösseres Gewicht beizumessen, als der detailgetreuen Berichterstattung zur Methodik.

Direkte und beobachtbare Wirkungen:

Dazu können wir nur gerade zwei Feststellungen machen. Zum einen betrifft dies das Echo, das der Schlussbericht in der Presse hervorrief. Aufgrund der Pressemitteilung der NFP-Pressestelle des SNF sind in verschiedenen Tageszeitungen kurze Agenturmeldungen zur Publikation der Evaluationsstudie publiziert worden (so etwa Der Landbote vom 14. 4. 1994, Luzerner Neueste Nachrichten (LNN) vom 14. 4. 1994, Neue Zürcher Zeitung (NZZ) vom 16. 4. 1994: 24). In einigen Fachorganen sind ebenfalls kürzere Artikel erschienen (Schweizer Baublatt vom 15. 4. 1994, Chemische Rundschau vom 21. 4. 1994 und La Revue polytechnique vom 24. 5. 1994). Längere Berichte zur Studie erschienen unseres Wissens bisher in zwei Zeitungen (Tages-Anzeiger (TA) vom 20. 4. 1994: 84 und Die Weltwoche (WeWo) vom 16. 6. 1994: 35).

Zweitens haben wir uns bei einigen Personen, welche sich intensiv mit dem Untersuchungsgegenstand befassen, nach Einschätzungen zur Studie erkundigt. Dabei ist ein sehr positiver Grundtenor festzustellen. Die angefragten Personen haben kaum negative Bemerkungen zur Studie geäußert. Die wenigen kritischen Anmerkungen betreffen etwa das Fehlen von Hinweisen auf die Problematik schwankender - und scheinbar auch tendenziell abnehmender - Qualitäten der Brennstoffe oder die bestehenden Defizite beim Vollzug der Züri-Norm, welche bei Stichprobenmessungen festzustellen sind und unter anderem im Zusammenhang mit billiger produzierten Heizungsanlagen zu stehen scheinen. In den Augen der Praktiker, welche ja mit der Sachlage sehr genau vertraut sind, erscheint der Bericht etwas zu lang ausgefallen zu sein, wobei gleichzeitig aber auch darauf hingewiesen wird, welche Bedeutung der Bericht hat, indem er einem breiteren Leserkreis Einblick in die Materie verschafft. Den Autoren ist es gelungen, einer sozialwissenschaftlich angelegten Untersuchung in einem traditionell eher technisch geprägten Umfeld zu einer hohen Akzeptanz und breiter Anerkennung zu verhelfen.

Im Kanton Luzern hat der Bericht bereits auch Eingang in die praktische Verwaltungstätigkeit gefunden, indem er zur Erarbeitung von Antworten zu parlamentarischen Vorstössen hilfreiche Argumentationshilfen lieferte. Andere Vollzugspraktiker haben darauf hingewiesen, dass sie die Studie im Zusammenhang mit der Weiterbildung von Feuerungskontrolleuren als Grundlage verwenden. In den Unternehmen der Heizungsbranche ist der Bericht (bisher) nur wenig bekannt.

Inwiefern der Bericht hinsichtlich seiner methodischen Komponenten Wirkungen erzeugt hat oder erzeugen wird, kann zum jetzigen Zeitpunkt (drei Monate nach der Publikation des Schlussberichts) noch nicht beurteilt werden.

Für eine tiefere Abklärung der Wirksamkeit der Evaluationsstudie reichen die Informationen aus der Dokumentenanalyse nicht aus. Wir müssen uns hier auf diese knappen Hinweise beschränken.

3.2. Anwendbarkeit

3.2.1. Praktische Verfahren

Die Evaluatoren haben mit verschiedenen Verfahren gearbeitet, nämlich mit Dokumentenstudium, Expertengesprächen, Sekundärstatistik und einer Umfrage. Die im Schlussbericht enthaltenen Informationen zur Durchführung der Datenerhebung und Datenauswertung erlauben nur teilweise einen Einblick in die Vorgehensweise. Deshalb sind wir bei der nachfolgenden Betrachtung darauf angewiesen, uns am produzierten 'output' zu orientieren.

Generell lässt sich sagen, dass die Evaluatoren mit ihrem Methodenset eine grosse Menge von Informationen beschaffen konnten. Mehrheitlich wurde die Zielsetzung erreicht. Trotzdem gab es jedoch einige Probleme bei der Datenerhebung. Hier sind vor allem zwei Punkte hervorzuheben (vgl. dazu auch Kapitel 3.1.3.):

- Das quantitative Datenmaterial zur Durchführung der Marktanalyse konnte nicht ausreichend differenziert erhoben werden, wobei unklar ist, ob die gesuchten Daten schlicht nicht vorhanden sind oder ob aus unbekanntem Gründen darauf verzichtet wurde, diese für die Evaluation zugänglich zu machen. Da der Schlussbericht keine genaueren Angaben dazu macht, ist schwierig abzuschätzen, inwiefern die Evaluatoren die Möglichkeiten zur Datenbeschaffung auch ausgeschöpft haben.
- Die quantitativen Zeitreihendaten für die Stadt Zürich konnten nur unvollständig erhoben werden. Für einige Jahre bestehen Lücken, welche die Interpretation der Datenreihen erschweren.

Ob nun diese Mängel bei einer anderen Vorgehensweise hätten ausgeräumt werden können, ist fraglich.

Durch die Verwendung verschiedener Verfahren in den thematisch eng miteinander verbundenen Untersuchungsteilen, erreichen die Evaluatoren, dass ihre Studie als ganzes weniger störungsanfällig wird. Diese Vorgehensweise kann im Prinzip mit der Diversifizierung eines Wertschriften-Portefeuilles verglichen werden. Risiken eines völligen Scheiterns werden damit minimiert. Auch wenn in einzelnen Untersuchungsteilen Probleme entstanden sind, hat dies in

keiner Weise die Praktikabilität der Gesamtanalyse in Frage stellen können. Das Kriterium wird erfüllt.

3.2.2. Politische Überlebensfähigkeit

Grundsätzlich bewegen sich die Evaluatoren in einem Politikfeld, das vergleichsweise wenig konfliktiv strukturiert ist. Dass Hausfeuerungen in ihren Emissionen beschränkt werden müssen, dass eine Typenprüfung die lufthygienische, energetische und feuerpolizeiliche Qualität von neuen Produkten kontrollieren soll und dass schlussendlich eine periodische Kontrolle die Überwachung des Anlagenbestandes sicherstellen soll, ist in der politischen Diskussion unbestritten. Dass technisch orientierte Umweltschutzmassnahmen generell eine höhere Akzeptanz aufweisen, verstärkt den Konsens noch. Die Diskussionen in diesem Themenbereich sind stark von Experten und Technikern geprägt und finden praktisch unter Ausschluss der Öffentlichkeit statt.

Die These der Evaluatoren aber, dass eine *regulative* Umweltpolitik die wirtschaftliche Entwicklung fördert, ist hingegen hart umstritten. Den Evaluatoren gelingt es empirisch zu belegen, dass im untersuchten Fall die staatliche Regulierung zur Förderung technischer Modernisierungen beitragen konnte. Diese Erkenntnis kann durch die empirische Untersuchung gestützt werden. Da die Resultate zudem zugunsten des status quo sprechen, kann ihre politische Überlebensfähigkeit als hoch eingestuft werden.

Die Evaluatoren beschränken sich jedoch nicht nur auf diese Resultate, sondern verallgemeinern diese in zweierlei Richtung:

- Die Auswirkungen des Programms zeige nicht nur positive Auswirkungen bezüglich der Nebenwirkung 'Förderung der wirtschaftlichen Entwicklung', sondern führe auch zu lufthygienischen und energetischen Verbesserungen (SB: 146-147). Wir sind schon weiter oben darauf zu sprechen gekommen, dass - gemäss den von Balthasar und Knöpfel erhobenen Daten - neue Heizungen nicht zwangsläufig auch energetisch *und* lufthygienisch bessere Heizungen sein müssen (vgl. oben Kapitel 3.1.3.).
- Die Ergebnisse der Evaluationsstudie sind - so die Aussage der Evaluatoren - auch auf andere Bereiche der Umweltpolitik übertragbar (siehe SB: 149-154). So wäre unter bestimmten Voraussetzungen ein Massnahmenset bestehend aus einem Grenzwert, einer Soll-Kontrolle und einer Ist-Kontrolle auch für andere Anlagen, Geräte oder Fahrzeuge geeignet, zur Förderung wirtschaftlicher Modernisierung beizutragen. Die angeführten Voraussetzungen sind ökologische (hoher Anteil lokal produzierter Emissionen, Möglichkeit der Zuordnung der Emissionen zu Emittenten), ökonomische (Innovationspotential, ausreichend hoher Umsatzanteil auf

dem Binnenmarkt) und politische Rahmenbedingungen (durchsetzungsfähige Verhandlungspartner).

Für beide Aussagen lässt sich sagen, dass sie aufgrund der vorliegenden Untersuchung legitimerweise nur in der Form von Thesen zulässig erscheinen (vgl. Kapitel 3.4.10.).

Da sich die angeführten Mängel nur gerade auf zwei spezifische Aspekte beziehen, ansonsten die politische Überlebensfähigkeit aber als hoch eingestuft wird, kommen wir zum Schluss, dass die Evaluation das Kriterium grösstenteils erfüllt.

3.2.3. Kostenwirksamkeit

Die Evaluation von Balthasar und Knöpfel gehört zu den teuersten der hier untersuchten Studien. Die Evaluation hat Kosten im Umfang von rund 254'000.- sFr. verursacht.

Die Evaluatoren haben mit der Triangulation für die Untersuchung der Fragestellung auf den ersten Blick eine sehr aufwendige Vorgehensweise gewählt. Da, wie dies die Autoren in der Studie auch deutlich hervorheben, es sich bei der Vorgehensweise nicht um eine Triangulation im eigentlichen Sinne handelt, fällt dies nicht besonders ins Gewicht. Die drei untersuchten Massnahmen werden nämlich nicht jeweils parallel mit unterschiedlichen methodischen Ansätzen analysiert, sondern es wird zu Untersuchung jeder Massnahme ein Ansatz verwendet (siehe SB: 164-168). Dadurch steigt der notwendige Aufwand nicht so stark, wie wenn jede Massnahme mit unterschiedlichen Ansätzen bearbeitet worden wäre. Da die untersuchten Massnahmen viele gegenseitige Berührungspunkte aufweisen, können die Evaluatoren von Synergien Gebrauch machen, was bei unabhängig voneinander erarbeiteten Einzelstudien nicht möglich gewesen wäre.

Wie gross der Wert der produzierten Informationen ist, kann nur schwer abgeschätzt werden. Aufgrund des Echos aus der Praxis scheint der Wert der Studie als vergleichsweise hoch eingestuft zu werden. Auch die methodisch orientierten Ergebnisse der Studie haben für die im Bereich der Evaluation tätigen Personen einen hohen Wert. Dieser wird jedoch etwas reduziert, da die Autoren auf eine detailliertere Beschreibung ihrer Vorgehensweise weitgehend verzichten.

Zusammenfassend lässt sich festhalten, dass sowohl die Kosten, wie auch der Wert der produzierten Informationen, als hoch einzustufen sind. Der Wert der produzierten Informationen rechtfertigt die Kosten der Studie. Deshalb wird die Kostenwirksamkeit als zufriedenstellend eingestuft.

3.3. Korrektheit

3.3.1. Formale Verpflichtung

Da die Evaluationsstudie im Auftrag des Schweizerischen Nationalfonds (NFP 27) erstellt worden ist, tritt das von den Evaluatoren verfasste Forschungsgesuch an die Stelle des sonst üblichen Evaluationskontrakts.

Finanzen: Im Forschungsgesuch wurde von den Evaluatoren ein Betrag von rund 250'000.- sFr. für die Durchführung des Projekts veranschlagt. Dieser Betrag ist durch Expertengruppe und Forschungsrat bewilligt worden. Es wurden keine Zusatzkredite gesprochen. Das Projekt hat Kosten von rund 254'000.- sFr. verursacht, wobei die Überschreitung durch Einnahmen abgedeckt worden ist. Die finanziellen Verpflichtungen sind von den Evaluatoren voll und ganz eingehalten worden.

Zeit: Das Forschungsgesuch sah eine Projektlaufzeit von zwei Jahren vor. Diese Projektdauer ist nicht verlängert worden. Lediglich eine zeitliche Verschiebung des Projektbeginns ist festzustellen. Das Projekt wurde zeitgerecht abgeschlossen. Die Evaluatoren sind den zeitlichen Verpflichtungen nachgekommen.

Personen: Andreas Balthasar und Carlo Knöpfel gehörten von Beginn weg und bis zum Abschluss der Studie dem Forschungsteam an. Die beiden Forscher haben sich nach Eingabe des Forschungsgesuchs - aber noch vor dem Projektbeginn - von der Forschungsstelle Wissenschaft und Politik in Bern gelöst und in Luzern beziehungsweise in Basel eigene Büros eröffnet. Da die Forscher aber auch schon in Bern als selbständig-erwerbende Mitglieder der Forschungsstelle tätig waren und da die Kontakte mit der Forschungsstelle nach wie vor vorhanden waren, haben diese räumlichen Veränderungen kaum weitergehende Konsequenzen nach sich gezogen. Die Zusammenarbeit zwischen den beiden Evaluatoren ist - soweit dies von aussen beurteilt werden kann - harmonisch verlaufen. Die festgelegte Arbeitsteilung hat sich bewährt. Die räumliche Trennung hat sich kaum negativ ausgewirkt.

Das Forschungsteam ist von verschiedenen externen Experten während der Projektbearbeitung unterstützt worden. Peter Knoepfel (IDHEAP, Lausanne) und Willi Zimmermann (Forschungsstelle Wissenschaft und Politik, Bern) haben die Evaluatoren in konzeptionellen Fragen beraten und Beratungshonorare im Umfang von 2'975.-- sFr. erhalten. Dies war bereits im Forschungsgesuch enthalten (FG: 5.4). Für technische Fragen ist im Verlaufe des Projektes ein Ingenieurbüro beigezogen worden (Waldhauser Haustechnik Ingenieurbüro, Münchenstein BL). In methodischen Fragen hat das Evaluatorenteam von Hans-Peter Kriesi (Universität Genf) und Thomas Widmer (Universität Zürich) Hinweise erhalten.

Die personellen Verpflichtungen sind durch die Evaluation eingehalten worden.

Inhalt: Bezüglich der inhaltlichen Dimension drängen sich verschiedene Hinweise auf:

- In der Marktanalyse sollte das "Verhalten der Nachfrager" ein "besonderes Gewicht erlangen" (FG: 3.4., siehe auch FG: 3.6.) Im Schlussbericht ist dieser Anspruch nicht eingelöst worden. Die Bemerkungen zum Nachfrageverhalten sind in der Marktanalyse sehr knapp ausgefallen. Wir haben dazu gerade zwei Textstellen eruiert. Bei der Beschreibung der Nachfragestruktur im Umfang von einer halben Seite handelt es sich mehrheitlich um eine Übernahme von Angaben aus einer Studie, welche im Auftrag des Bundesamtes für Umwelt, Wald und Landschaft (BUWAL) und des Bundesamtes für Energiewirtschaft (BEW) erstellt wurde (SB: 76, vgl. Wüest & Partner 1992). Die zweite Stelle im Schlussbericht befasst sich mit den Reaktionen der Nachfrager und hat einen Umfang von knapp einer halben Seite (SB: 79). Hier findet sich zum einen der Satz: "Den schweizerischen Nachfragern nach Heizanlagen wird eine grosse Kundentreue nachgesagt." Neben dieser eher allgemeinen Aussage beschränken sich die übrigen Hinweise auf das Aufkommen von Kompaktanlagen, sogenannten Unit-Lösungen, bei denen Brenner und Kessel integriert sind. Insgesamt gesehen ist dies doch etwas dürftig.
- Im Forschungsgesuch findet sich folgender Hinweis: "Als Grundlage" ... "werden mit einer Policy-Analyse zuerst die Prozesse der Normengenesse und -Durchsetzung auf der Ebene des Bundes, der Kantone und der Gemeinden aufzuarbeiten sein." (FG: 3.4.) Diese sehr weite Zielsetzung konnte im Rahmen der Evaluation nicht vollständig behandelt werden. Dies führen wir aber in erster Linie auf eine zu umfangreiche Zielfestlegung zurück.
- Untersucht man die im Forschungsgesuch formulierten Fragen zur Marktuntersuchung (FG: 3.5-3.6), muss festgestellt werden, dass diese oft nicht bearbeitet wurden. Dies hängt natürlich auch mit der schwierigen Datensituation zusammen, auf die wir bereits hingewiesen haben (vgl. Kapitel 3.1.3.). Die Angaben im Forschungsgesuch lassen aber deutlich mehr erwarten, als realisiert werden konnte.
- Im Zusammenhang mit der Bedeutung des Projekts für die Praxis machen die Autoren im Forschungsgesuch folgenden Hinweis: "Es sollen Überlegungen angestellt werden, ob Verfahren, wie sie im Teilbereich Feuerungskontrolle dieses Forschungsprojektes zu[r, Ergänzung TW] Anwendung kommen, als Monitoring standardisiert und regelmässig in die Vollzugskontrolle eingebaut werden könnten." (FG: 3.8.) Im Schlussbericht haben wir keine derartigen Überlegungen finden können.
- Im Gegensatz zu den angeführten Lücken wurde der Untersuchungsbereich im Vergleich zum Forschungsgesuch in manchen Teilen auch ausgeweitet. Dies gilt insbesondere für jenen Untersuchungsteil, der sich mit dem Gesamtprogramm auseinandersetzt. Das Forschungsge-

such hatte hier nur eine verhältnismässig knappe Synthese angekündigt (FG: 3.7.), der Schlussbericht enthält dazu einen zusätzlichen, empirischen Untersuchungsteil (SB: 115-131).

Methodik/Vorgehen: Auch hier sind verschiedene Aspekte zu erwähnen:

- Im Forschungsgesuch werden zwei technologische Fallstudien angekündigt (FG: 3.4.). Im Schlussbericht findet sich jedoch lediglich eine solche. Dafür hat die Methodik an anderen Stellen eine Erweiterung erfahren. Dies gilt insbesondere für den Untersuchungsteil zum Gesamtprogramm (SB: 115-131; vgl. FG: 3.7.).
- Das Untersuchungsdesign, wie es im Forschungsgesuch dargelegt ist, musste im Projektverlauf teilweise stark geändert werden. Das betrifft besonders die quantitative Studie zur Feuerungskontrolle. (vgl. FG: 3.6.)
- Da die Evaluatoren mit Angaben zur konkreten Vorgehensweise im Schlussbericht sehr zurückhaltend sind, kann nicht beurteilt werden, inwiefern sie den im Forschungsgesuch festgehaltenen methodischen Angaben gefolgt sind. Es wird dort beispielsweise festgehalten, dass im Verlauf des Projekts für die verschiedenen Untersuchungsteile insgesamt 35 Expertengespräche geführt werden sollen. Der Schlussbericht enthält dazu die Angabe, dass mit "rund 50 Expertinnen und Experten" (SB: VI) Gespräche geführt wurden. Eine Liste der kontaktierten Personen fehlt im Schlussbericht.

Zusammenfassend kann festgehalten werden, dass die Evaluatoren die finanziellen, zeitlichen und personellen Verpflichtungen vollständig eingehalten haben. Bezüglich der inhaltlichen und der methodischen Aspekte ist dies nur eingeschränkt der Fall. Andererseits enthält die Studie in diesen Bereichen aber auch einige Teile, die über das im Forschungsgesuch enthaltene hinausgehen. Inhaltliche Ausrichtung und die Untersuchungsanlage mussten hinsichtlich einiger Punkte im Verlauf des Projektes angepasst werden. Dies ist damit zu erklären, dass ein Forschungsprojekt, wie das hier diskutierte, verschiedenen Einflussfaktoren ausgesetzt ist, welche vom Evaluationsteam nicht vorhergesehen werden können und auf die es auch wenig Einfluss geltend machen kann. Unter Umständen sind beispielsweise Änderungen in der Untersuchungsanlage nicht nur legitim, sondern stellen - im Vergleich zu einem starren Festhalten am geplanten Vorgehen - auch eine bessere Lösung dar. Die Autoren verzichteten darauf, diese Unwegsamkeiten (und Lernprozesse) im Forschungsablauf im Schlussbericht zu thematisieren.

Unter Berücksichtigung dieser Rahmenbedingungen kommen wir zum Schluss, dass die Evaluation das Kriterium erfüllt.

3.3.2. Deklaration von Interessenkonflikten

Interessenkonflikte sind im Zusammenhang mit dieser Studie kaum auszumachen. Wir haben lediglich folgende expliziten oder impliziten Hinweise eruieren können:

- In einer Anmerkung (Anm. 88, SB: 177) wird darauf hingewiesen, dass die Anlagenkäufer - also die Anlagenbetreiber oder weitere Personen (etwa Mieter), wenn die Kosten überwältzt werden können - die Kosten für die Modernisierung der Heizungsanlagen zu tragen hätten. Der Konflikt zwischen den Interessen der am Prozess Beteiligten und den Nicht-Beteiligten aber sehr wohl Betroffenen, wird im Schlussbericht kaum thematisiert (vgl. oben, Kapitel 3.1.1.).
- Im Laufe der Datenerhebung hatten die Evaluatoren Schwierigkeiten, Informationen für die Marktanalyse zu erheben. Die Branche scheint nicht sonderlich kooperativ gewesen zu sein, wie man aufgrund der Hinweise im Schlussbericht vermuten kann (SB: 159-160). Möglicherweise bestanden beim Zugang zu Daten bei öffentlichen Stellen ebenfalls Probleme. Der Schlussbericht enthält keine diesbezüglichen Hinweise. Möglicherweise war die Zugänglichkeit zum Datenmaterial ein wichtiger Faktor bei der Auswahl der untersuchten Fälle.
- Die Evaluatoren machen in der Prozessbeschreibung zur Entstehung der Züri-Norm einen Hinweis auf personelle Wechsel bei zwei Firmen der Heizungsbranche. Die Ursachen dafür werden nicht explizit gemacht. Es wird auch nicht klar, ob diese Veränderungen in direktem Zusammenhang mit den Entwicklungen hin zur Züri-Norm stehen.
- Ein Konflikt, der bereits in der Fragestellung der Untersuchung angelegt ist, betrifft das Verhältnis zwischen umweltpolitischer Intervention und der Wirtschaft. Den Autoren gelingt es gut, die Abläufe darzustellen, welche dazu führten, dass die Wirtschaft und der Staat sich auf eine Verschärfung der staatlichen Normierung einigten. Was unseres Erachtens jedoch eine problematische Auslassung durch die Evaluatoren darstellt, sind die ökologischen Auswirkungen einer forcierten Modernisierung des Anlagenparks. Die Autoren haben in ihrer Studie die ökologischen Auswirkungen bewusst *nicht* untersucht. Die Fragestellung der Untersuchung beschränkt sich ganz explizit auf eine Nebenwirkung - die Förderung technischer Entwicklung. Es ist durchaus legitim, die Fragestellung einer Evaluation auf bestimmte Auswirkungen des Evaluandums auszurichten, solange dies auch deutlich dargelegt wird (vgl. dazu Kapitel 3.4.10.).

Es sind uns keinerlei Informationen bekannt, die darauf hindeuten würden, dass aufgrund von Interessenkonflikten eine Kompromittierung des Prozesses oder der Resultate der Evaluation erfolgt wäre. Deshalb sind wir der Ansicht, dass die Evaluation das Kriterium erfüllt.

3.3.3. Vollständige und aufrichtige Offenlegung

Soweit dies aufgrund der vorliegenden Informationen beurteilt werden kann, halten die Evaluatoren dieses Kriterium ein. Auf die Grenzen der verwendeten Ansätze und Methoden wird deutlich hingewiesen - mit Ausnahme der vorgängig diskutierten Überinterpretationen (vgl. dazu ausführlich Kapitel 3.4.10.).

Punktuell verwenden die Autoren direkte Zitate, ohne diese deutlich zu kennzeichnen. So finden sich einige Textübernahmen aus Freiburghaus (1988: 15-16) in den einleitenden Bemerkungen zur Evaluationsforschung (SB: 17). In der Marktanalyse finden sich direkte Zitate, die ebenfalls nicht als solche ausgewiesen werden (SB: 71-73 und 76). Die Textpassagen wurden einer Studie entnommen, welche durch das BUWAL und das BEW in Auftrag gegeben wurde (Wüest & Partner 1992: 13, 15-16).

Soweit sich dies beurteilen lässt, ist die Evaluation in der Darlegung ihrer Ergebnisse und Grenzen offen, direkt und ehrlich. Davon ausgenommen werden müssen die im vorangehenden Kapitel genannten und an anderer Stelle (siehe Kapitel 3.4.10.) ausführlich diskutierten Anregungen, die über den eigentlichen Gegenstandsbereich der Evaluation hinausreichen. Hier hätte deutlicher deklariert werden müssen, welcher Stellenwert diesen Ausführungen zukommt. Die Evaluation erfüllt das Kriterium mit Einschränkungen.

3.3.4. Herstellung von Öffentlichkeit

Das Recht der Öffentlichkeit ist von der Evaluation beachtet worden. Durch den klar gestalteten Bericht gelingt es den Evaluatoren, die Thematik auch Laien näherzubringen. Der Schlussbericht wurde nach Fertigstellung dem Verlag zur Buchproduktion übergeben. Die Evaluatoren haben sich darum bemüht, dass die Studie in den Medien ein grosses Echo erhält, was teilweise gelungen ist. Verschiedene Personen, welche mit der untersuchten Thematik beruflich zu tun haben, sind aufgrund der Berichterstattung der Presse auf die Studie aufmerksam geworden. Dieses Kriterium ist eingehalten worden.

3.3.5. Schutz individueller Menschenrechte

Die Rechte und das Wohlergehen der Menschen sind - soweit wir dies beurteilen können - von der Evaluation beachtet worden. Das Kriterium ist erfüllt.

3.3.6. Menschliche Interaktionen

Aufgrund des Echos verschiedener Personen, welche den Evaluatoren als Auskunftspersonen gedient hatten, ist dieses Kriterium eingehalten worden. Die Hilfeleistungen verschiedener Akteure, welche den Evaluatoren Informationen zur Verfügung stellten, werden in Anmerkungen zumeist verdankt. Im übrigen geht aus dem Schlussbericht nicht hervor, mit welchen Experten die Evaluatoren im Rahmen der Expertengespräche Kontakt aufgenommen hatten (siehe Kapitel 3.4.3.). Deshalb fällt es schwer, nur aufgrund einer Dokumentenanalyse eine Einschätzung abzugeben. Soweit unsere Informationen reichen, sind jedoch keinerlei Verstösse hinsichtlich dieses Kriteriums zu beobachten.

3.3.7. Ausgewogene Berichterstattung

Grundsätzlich liefern die Evaluatoren eine ausgewogene Präsentation der Stärken und Schwächen des Evaluandums. Einerseits wird die Stärke des vorliegenden Instrumentariums ausführlich dargelegt, andererseits scheuen sich die Autoren aber auch nicht, Kritik anzubringen. Sie geben auch Empfehlungen für eine verbesserte Ausgestaltung des evaluierten Programms. Wir können feststellen, dass die Evaluation das Kriterium in ausreichendem Masse berücksichtigt hat.

3.3.8. Finanzielle Verantwortlichkeit

Die Evaluatoren haben für die Beantwortung der Fragestellung ein recht aufwendiges Design gewählt. Da sich die verschiedenen Ansätze jeweils mit unterschiedlichen Massnahmen befassen und sich die Untersuchungsteile sinnvoll ergänzen, hat sich dies als kluge Lösung erwiesen. Da die Studie zu den teuersten der hier untersuchten gehört, darf bezüglich der einzelnen Untersuchungsteile kritisch angemerkt werden, dass teilweise etwas wenig an Informationen in den Schlussbericht eingeflossen ist. Dabei muss jedoch deutlich festgehalten werden, dass die Evaluatoren bei der Auswahl der 'schlussberichtswürdigen' Informationen selektiv vorgegangen sind.

Die Evaluatoren haben einen Betrag von knapp 3'000.- sFr. für zwei externe Berater (Peter Knoepfel und Willi Zimmermann) ausgegeben. Um diese Ausgaben bewerten zu können, fehlen uns die entsprechenden Angaben. Die ausgewählten Berater sind sicherlich bestens ausgewiesen und für die vorliegende Thematik äusserst kompetent.

Die Evaluatoren haben beim Schweizerischen Nationalfonds um einen Druckkostenbeitrag im Umfang von 13'870.- sFr. für die Buchpublikation ersucht und diesen auch zugesprochen erhal-

ten. Aufgrund der Qualitäten der Publikation war dies sicherlich gerechtfertigt. Das Kriterium wird erfüllt.

3.4. Genauigkeit

3.4.1. Identifikation des Gegenstandes

Die Evaluatoren behandeln in ihrer Untersuchung "die Auswirkungen der Grenzwertsetzung, der Typenprüfung und der Feuerungskontrolle auf die technische Entwicklung von Hausheizungen." (SB: V) Sie wählen damit drei Massnahmen, welche einen technisch orientierten Ansatz der Umweltpolitik verfolgen. Die Gründe für die Auswahl gerade dieser Massnahmen, erläutern die Evaluatoren nicht. Sie untersuchen nicht die Hauptwirkungen dieser Massnahmen (Luftreinhaltung, effektiver Energieeinsatz und Sicherheit), sondern einen vermuteten Nebeneffekt, die Zusammenhänge zur wirtschaftlich-technologischen Entwicklung bei den Hausfeuerungen. Dabei entscheiden sich die Autoren, sich auf kleinere und mittlere Heizungen (bis 350 kW Kesselleistung) zu konzentrieren. Zudem beschränken sie sich auf Anlagen, welche mit Öl betrieben werden. Die Evaluatoren grenzen somit den Untersuchungsgegenstand recht stark ein. Insbesondere verzichten sie auf die Untersuchung der ökologischen Hauptwirkungen und auf die Diskussion gesellschaftspolitischer Fragen. Ebenso wird die Wahl der 'richtigen' Massnahme nicht diskutiert. (siehe dazu Kapitel 1.1., SB: 7-8)

Durch die präzise Fassung und die Wahl eines stark eingegrenzten Untersuchungsgegenstandes schaffen die Evaluatoren die Voraussetzungen für eine mit den vorgegebenen Randbedingungen bewältigbare Untersuchung. Leider wird die zu Beginn der Studie dargelegte thematische Definition nicht bis zum Schluss der Studie konsequent beibehalten. Insbesondere bei der Darstellung der Ergebnisse im Kapitel "Wirtschaftspolitische Lehren" kommen die Autoren gerade auf Aspekte zu sprechen, die zuvor explizit ausgeschlossen wurden. Zum einen betrifft dies die Ausführungen zu den ökologischen Auswirkungen des Programms, auf die wir bereits hingewiesen haben (SB: 146-147, vgl. oben Kapitel 3.3.2. und unten 3.4.10.). Zweitens diskutieren die Evaluatoren im erwähnten Abschnitt auch die 'richtige' Wahl der Instrumente (SB: 146-152). Auch dies eine Frage, welche zu Beginn der Studie explizit ausgeschlossen wurde (SB: 8). Die Frage, wie die Evaluatoren zu den Aussagen zu diesen Fragen kommen, ohne diese untersucht zu haben, ist berechtigt.

Die Evaluation beschränkt sich in der quantitativen Analyse der Feuerungskontrolle auf Kantone, welche nach dem sogenannten 'Luzerner Modell' vollziehen (SB: 99-114). Dies ist enttäuschend, weil der Untersuchungsgegenstand hier nochmals eingeschränkt wird und zwar besonders deshalb, weil, diesem Abschnitt vorangestellt, ein qualitativer Vergleich der drei

Vollzugsmodelle (Luzerner Modell, Modell Romandie und Basler Modell, siehe SB: 89-99) erfolgt ist. Dies hat beim Leser natürlich zur Erwartung geführt, dass nun ein quantitativer Modellvergleich erfolgt. Aus pragmatischen Gründen, die explizit gemacht sind und legitim sind, wird aber darauf verzichtet. Eine Erweiterung auf die drei Modelle hätte den Rahmen der Evaluation gesprengt.

An einigen Stellen der Studie finden sich ungenaue oder widersprüchliche Angaben:

- Das *Umweltschutzgesetz* ist anfangs 1985 in Kraft getreten (AS 1984 1143), und nicht 1986 (SB: 57).
- Die *Typenprüfung* ist durch die LMFV per 1. Januar 1985 obligatorisch erklärt worden, wobei eine Übergangsfrist bis Ende Juni 1985 eingeräumt wurde (Art. 13 und 14, LMFV, AS 1984 1518). Der Schlussbericht spricht davon, dass die Typenprüfung ab 1986 obligatorisch geworden sei (SB: 57).
- Die *Typenprüfungen* nach LMFV (AS 1984 1516) und jene nach LRV 86 (AS 1986 208) sind nicht identisch. Mit der LRV 92 wurde auch die Typenprüfung für Anlagen mit einer Leistung von über 70 kW obligatorisch. Diese Differenzierungen fehlen bei den Erläuterungen zur Ausgestaltung der Typenprüfung.
- Die *Feuerungskontrolle* ist seit dem 1. Januar 1985 obligatorisch (AS 1984 1516), und nicht seit Mitte 1985 (SB: 84, 86)
- Die *Feuerungskontrolle* im 'Luzerner Modell' wird nicht "ausschliesslich im Winterhalbjahr angesetzt" (SB: 96), sondern auch im Sommer vorgenommen, sofern die Heizung - zum Beispiel für die Warmwasseraufbereitung - ganzjährig in Betrieb ist (telefonische Auskunft, Herr Wintsch, Gesundheits- und Wirtschaftsamt der Stadt Zürich).
- Die *Zunahme des Gebäudebestandes* in den untersuchten Gemeinden beträgt durchschnittlich 17,4 Prozent (Darstellung 5.5., SB: 106) und nicht 8 Prozent (SB: 107).
- "Im Mittelpunkt unserer Abklärungen steht der Anteil '*moderner Heizungen*' am kommunalen Anlagebestand in den Kantonen Luzern, Bern und Basel-Land"...". Darunter verstehen wir jene Heizungen, welche mit einem Kessel mit Jahrgang 1986 oder jünger bestückt sind. Das sind die Jahrgänge welche gemäss den gesetzlichen Vorschriften typengeprüft sein müssen." (SB: 100, Hervorhebung TW) Diese Bemerkungen sind in zweierlei Hinsicht problematisch. Erstens ist eine Typenprüfung erstmals 1985 eingeführt worden. Zweitens bezieht sich dieses Obligatorium bis zum Inkrafttreten der LRV 92 lediglich auf Heizungen in der Leistungsklasse bis 70 kW. Da die Autoren im Anschluss an die oben zitierte Feststellung mit Daten arbeiten, welche auch Heizungen mit Leistungen von 70 kW bis 1MW umfassen, ist diese Aussage irreführend.

- Die Definition, wonach als '*moderne Heizungen*' jene Anlagen bezeichnet werden, "die mit einem Kessel mit Jahrgang 1986 oder jünger ausgerüstet sind" (SB: 120), ist wohl für eine Betrachtung einer Zeitreihe von 1964 bis 1991 (Darstellung 6.4., SB: 120) nicht sinnvoll. Hier sind Alterskohorten und nicht Jahrgangskohorten zu verwenden.

Zusammenfassend lässt sich feststellen, dass die Evaluation doch an einigen Stellen ungenau ist. Mehrheitlich ist jedoch der Untersuchungsgegenstand ausreichend genau untersucht worden. Der Verzicht auf eine quantitative Analyse der drei Implementationsmodelle bei der Feuerungskontrolle lässt Fragen offen. Trotzdem erreichen die Evaluatoren - unterstützt durch die enge thematische Eingrenzung und die breit angelegte Methodik - eine gute Abdeckung des Untersuchungsgegenstandes. Das Kriterium wird grösstenteils erfüllt.

3.4.2. Kontextanalyse

Die Ausführungen zum Kontext des evaluierten Programms finden sich vor allem im ersten Teil der Studie (SB: 1-23). Besonders ist dabei die Darstellung der Problemwechsel zwischen lufthygienischen und energetischen Aspekten hervorzuheben (SB: 8-15). Auch die Debatte um die wirtschaftlichen Auswirkungen umweltpolitischer Normierung wird angesprochen (SB: 4-7). Diese Bemerkungen erlauben es dem Leser, den Untersuchungsgegenstand in sein Umfeld einzuordnen und erleichtern ihm den Zugang zur Thematik massgeblich. Die Vorgehensweise der Evaluation - verschiedene Massnahmen eines Programms mit je einem unterschiedlichen Ansatz zu untersuchen - bietet den Vorteil, dass die einzelnen Teilstudien als Kontextanalyse für die anderen Teilstudien genutzt werden können. Die Evaluatoren halten die Ausführungen zum Kontext zwar insgesamt recht knapp, wobei wir den Eindruck erhalten, dass die wichtigsten Aspekte abgedeckt worden sind. Die Evaluation erfüllt dieses Kriterium vollständig.

3.4.3. Beschreibung von Zielen und Vorgehen

Grundsätzlich kann man feststellen, dass der Schlussbericht zwar die grundlegenden Ziele und Vorgehensweisen ausreichend genau beschreibt, dass hingegen keine präzise Darstellung des Evaluationsprozesses erfolgt. Vielfach wird aus dem Bericht nicht klar wie die Evaluatoren nun konkret ihre Untersuchungskonzepte umgesetzt haben. Ebenso haben die Evaluatoren fast gänzlich darauf verzichtet, auf die Gründe für die im Evaluationsprozess getroffenen Forscherentscheide einzugehen. Insgesamt gesehen fehlt dem Bericht dadurch die Transparenz, welche dem Leser Einblick in das Vorgehen und die dabei gemachten Überlegungen verschaffen würde. Beispielsweise wären folgende Fragen genauer zu klären:

a) Fallstudie Grenzwertsetzung (SB: 28-56):

- Warum bezieht sich die Fallstudie lediglich auf die Entwicklung der Grenzwertsetzung zu den NO_x-Emissionen in der LRV 92? Warum sind andere Grenzwerte ausgeklammert worden? Warum sind andere Erlasse nicht berücksichtigt worden?
- Wie wurde das Dokumentenstudium gestaltet? Welche Dokumente hat man verwendet?
- Wie sind die Evaluatoren bei der Erarbeitung der Fallstudie vorgegangen?

b) Marktanalyse Typenprüfung (SB: 57-83):

- Welche Statistiken wurden beigezogen? Wie wurden diese qualitativ und quantitativ bearbeitet? Welchen Zeitraum umfasst die Analyse?
- Wie wurde die Trennung zwischen Brenner- und Kesselfabrikanten, wie jene zwischen Produzenten und Installateuren vorgenommen? Werden Unit-Lösungen den Kesseln oder/und den Brennern zugeordnet? (siehe etwa Darstellungen 4.5. und 4.6., SB: 74-76)

c) Quantitative Analyse Feuerungskontrolle (SB: 84-114):

- Wie ist die Auswahl der untersuchten Kantone bzw. Gemeinden zustande gekommen? Warum soll die Auswahl der untersuchten Berner Gemeinden weitgehend repräsentativ sein (SB: 103)?
- Welcher Umfragetyp ist bei der Datenerhebung verwendet worden? Schriftlich oder mündlich (telefonisch oder persönlich)? Standardisiert? Geschlossen? Strukturiert?
- Wie wurden die kontrolleurspezifischen Faktoren auf der Ebene der Gemeinden aggregiert? Wie wurden Fälle mit fehlenden Werten behandelt?
- Welche Art von Varianzanalyse ist benutzt worden (SB: 108)?
- Wie lauten die präzisen operationalen Definitionen der Variablen? Wie lautet die Definition der Variable "Anzahl Anlagen" (Darstellung 5.2., SB: 102), "kontrollierte Anlagen" (Darstellung 5.7., SB: 110) bzw. "die Gemeindegrösse - in Form der Anzahl kontrollierter Anlagen" (SB: 111)? Wenn der Indikator die Anzahl der kontrollierten Anlagen anzeigt, wäre dies ein ungeeigneter Indikator für die Gemeindegrösse, da der Kontrollrhythmus ein oder zwei Jahre beträgt (vgl. SB: 103) und die Grösse der Anlagen nicht berücksichtigt werden (vgl. dazu auch unten Kapitel 3.4.5.).
- Wie lauten die nicht-standardisierten Regressionskoeffizienten (SB: 110)? Warum wurde ein bereinigtes R² Mass benutzt (SB: 110)? Sind die Gesamtschätzungen signifikant? Wie verhalten sich die Regressionsschätzungen bezüglich Multikollinearität? Mit welchem Schätz-

verfahren wurde gearbeitet? Wie gross ist die Anzahl der für die Schätzung zur Verfügung stehenden Fälle?

- Auf welche Leistungsklasse von Heizungen bezieht sich das Datenmaterial zur Feuerungskontrolle? Sind lediglich Ölfeuerungskontrollen berücksichtigt oder auch die Gasfeuerungskontrollen? (vgl. Kapitel 3.4.5. und 3.4.8.)
- Warum wird das 'Basler Modell' "in zu wenigen Gemeinden angewandt, als dass eine komparative Analyse von kommunalen Daten möglich wäre" (SB: 100)? Die Kantone Schwyz, Obwalden, Schaffhausen, und beide Basel, welche gemäss den Angaben im Schlussbericht nach dem Basler Modell vollziehen (siehe Darstellung 5.1., SB: 90), umfassen ja zusammen immerhin 74 Gemeinden.

d) Zeitreihenanalyse Gesamtprogramm (SB: 115-131):

- Warum wurden die beiden Städte Zürich und Luzern für diesen Untersuchungsteil ausgewählt? (siehe SB: 116)
- Im "Luzerner Modell" wird die Feuerungskontrolle gemäss den Angaben der Evaluatoren "ausschliesslich im Winterhalbjahr angesetzt" (SB: 96), was aufgrund der uns vorliegenden Informationen nicht zutrifft (vgl. oben Kapitel 3.4.1.). In diesem Zusammenhang kommt aber die Frage auf, wie die Jahreszahlen beispielsweise in Darstellung 6.2. (SB: 118) zu interpretieren sind. Steht die Zahl "64" für den Winter 1963/64 oder für den Winter 1964/65, oder ist damit wirklich das Kalenderjahr 1964 subsumiert?
- Stammt das zugrundeliegende Datenmaterial für die Stadt Zürich nur aus Routinekontrollen oder sind darin auch Nachkontrollen und Kontrollen, welche auf Klage hin durchgeführt wurden, enthalten? (siehe etwa Darstellung 6.2., SB: 118)
- In den Darstellungen fehlen zumeist Hinweise, wie gross die Anzahl der Fälle für die verschiedenen Kategorien ist. Beispielsweise ist zu vermuten, dass von den 2298 Routinemessungen, die im Jahre 1991 in der Stadt Luzern durchgeführt wurden, nur wenige darunter sind, die bei einer Anlage mit einem Kessel aus dem Jahre 1959 vorgenommen wurden (siehe Darstellungen 6.7. und 6.8.). Werden unter der Bezeichnung Kesseljahrgang 1959 auch Kessel älteren Datums aufgeführt (siehe Darstellung 6.6., SB: 123)?

e) Fragen zu allen Untersuchungsteilen:

- Wie wurden die Expertengespräche durchgeführt? Welche Experten wurden dafür beigezogen? Wieviele Gespräche wurden geführt? Das Forschungsgesuch spricht von 35 Expertengesprächen (FG: 5.3), im Vorwort zum Schlussbericht wird auf Gespräche mit rund 50 Expertinnen und Experten hingewiesen. Wie lange dauerten diese? Hat man einen Leitfaden

verwendet? Wurde das Gespräch vorstrukturiert? Welche Zielsetzungen verfolgte man mit den Expertengesprächen? Wurden die Gespräche protokolliert? Wie?

Positiv hervorzuheben ist im Gegensatz dazu, dass die Evaluatoren ihr Vorgehen methodisch reflektieren und versuchen, die Vorteile und Grenzen der verschiedenen Ansätze darzustellen (SB: 156-164). Damit leistet der Schlussbericht einen wichtigen Beitrag zur methodischen Zielsetzung des Nationalen Forschungsprogramms 'Wirksamkeit staatlicher Massnahmen' (NFP 27), in dessen Rahmen die Evaluation entstanden ist. Der Schlussbericht richtet sich an einen breiteren Adressatenkreis. Aus diesem Grund haben sich die Autoren dafür entschieden, die methodische Berichterstattung nicht zu stark zu betonen. Die im Schlussbericht bereitgestellten Informationen zur Vorgehensweise können deshalb die Bedürfnisse eines spezifisch methodisch interessierten Publikums nicht in allen Teilen befriedigen. Der Schlussbericht erlaubt es dem interessierten Leser nur beschränkt, das gewählte Vorgehen vollständig nachzuvollziehen. Möglicherweise hätte sich diese Problematik mit einem methodisch orientierten Anhang entschärfen lassen. Obwohl die Evaluation hinsichtlich der Beschreibung der Ziele und des Vorgehens durchaus einige positive Aspekte aufweist, vermag die Evaluation das Kriterium nicht zu erfüllen.

3.4.4. Verlässliche Informationsquellen

Wir haben bereits im vorangehenden Kapitel detailliert auf die Lücken hingewiesen, die bei der Beschreibung der Vorgehensweise bestehen. Die Hinweise bezüglich der Informationsquellen sind auch bei diesem Kriterium zu berücksichtigen. Besonders wichtig sind hier:

- das Fehlen von Angaben zu den Gesprächspartnern für die Expertengespräche,
- die Unklarheiten, welche Dokumente für das Dokumentenstudium genutzt wurden und
- die Frage, welche Statistiken beigezogen wurden.

An diesen Stellen liegen eindeutig erhebliche Lücken in der Beschreibung vor, so dass die Adäquatheit der Informationen in den betroffenen Teilen der Untersuchung nicht eingeschätzt werden kann. Aufgrund dieser Überlegungen kommen wir zum Schluss, dass das Kriterium nicht eingehalten wurde.

3.4.5. Valide Messung

Die einzelnen Teilstudien sind hier separat zu betrachten.

a) Fallstudie Grenzwertsetzung (SB: 28-56):

Aufgrund der uns zugänglichen Informationen sind die Instrumente und Verfahren zur Erhebung von Informationen zweckmässig gewählt und umgesetzt worden. Im Gegensatz zum Vorverständnis, wie es von den Forschern im einleitenden Kapitel dargelegt worden ist (SB: 18-21), kommt in der Aktoranalyse eine weitere Akteurgruppe hinzu, die Interessenverbände (vgl. dazu auch oben Kapitel 3.1.1.). Die Aktoranalyse baut auf dem ökonomischen Kalkül auf, und versucht zu erklären, wie es zu einem Zusammengehen der verschiedenen Akteurgruppierungen kommen konnte. Sowohl die Deskription wie auch die Interpretation vermögen Informationen hervorzubringen, welche die verfolgten Ziele erreichen.

b) Marktanalyse Typenprüfung (SB: 57-83):

Die wenigen zur Verfügung stehenden quantitativen Angaben (vgl. Kapitel 3.1.3.) beruhen oft auf Schätzungen oder Hochrechnungen, deren Gültigkeit nicht über allen Verdacht erhaben sind. Die qualitative Analyse weist relativ viele Ungenauigkeiten oder gar Unstimmigkeiten auf (siehe Kapitel 3.4.1.). Aus diesen Überlegungen heraus kommen wir zum Schluss, dass die Marktanalyse dieses Kriterium nicht vollumfänglich erfüllt. Dabei ist jedoch deutlich darauf hinzuweisen, dass den Evaluatoren der fehlende Datenzugang nicht angelastet werden kann. Eventuell hätte die Einbindung der Branchenorganisationen in die Evaluation (etwa in der Form einer Begleitgruppe) dazu beitragen können, den Datenzugang zu verbessern.

c) Quantitative Analyse Feuerungskontrolle (SB: 84-114):

In diesem Analyseteil erfolgt vor der quantitativen Analyse mittels eines Regressionsmodells eine qualitative Aufarbeitung der verschiedenen Vollzugstypen. Dazu werden die kantonalen Vollzugsformen drei Modellen zugeordnet. Innerhalb der drei dargestellten Modelle bestehen teilweise grössere Unterschiede. Darauf weisen die Autoren in pauschaler Form hin (SB: 89), eine detailliertere Aufarbeitung der Differenzen wird lediglich für eines der Modelle geleistet (Kapitel 5.3., SB: 99-114). Weiter ist hier darauf hinzuweisen, dass sich in diesem Studienteil auch einige ungenaue oder falsche Angaben finden. 1982 haben gemäss einer Umfrage der Schweizerischen Vereinigung für Gesundheitstechnik nicht "nur 20 Prozent der Gemeinden, welche Ölfeuerungskontrollen durchführten, auch die Energieverluste" (SB: 86) kontrolliert, sondern es waren 32 von 142 Gemeinden, also etwa 22,5 Prozent (SVG 1983b: 27). Die vom EDI 1978 erlassenen und den Kantonen und Gemeinden zur Nachahmung empfohlenen und hier

relevanten Richtlinien sind nicht jene mit der Bezeichnung "Richtlinien über die Auswurfbegrenzung bei Haus- und Industriefeuerungen" (Anm. 71, SB: 176, vgl. auch SB: 87; BBl 1972/I: 1089-1095), sondern heissen "Richtlinien zur Prüfung der Abgase von Ölf Feuerungen" (BBl 1978 I: 1396-1402). (Vgl. auch Kapitel 3.1.3., 3.1.5. und 3.4.1.)

Unsere Analyse der quantitativen Auswertungen erfolgt hier entlang der Variablen, welche in die quantitative Analyse einbezogen worden sind:

ca) Abhängige Variable:

- *Anteil 'moderner Heizungen':* Zu den modernen Heizungen werden jene Anlagen gezählt, die über einen Kessel mit Jahrgang 1986 oder jünger verfügen. Diese sollen, folgt man der Argumentation der Autoren, einer Typenprüfung unterstellt sein. Wie bereits dargelegt worden ist, wurden einerseits bereits vor 1986 Typenprüfungen vorgenommen und andererseits sind nicht alle Heizungen, die ab 1986 installiert wurden, typengeprüft. Weiter ist der Indikator nicht in der Lage, Informationen zum Brenner zu liefern. Es ist sehr wohl denkbar, dass ein älterer Kessel mit einem neueren (typengeprüften) Brenner ausgestattet ist. Das heisst mit anderen Worten, dass dieser Indikator nicht in der Lage ist, auszuweisen, wie gross der Anteil typengeprüfter Anlagen ist. Der Indikator soll dazu dienen, die Verbreitung moderner Technologien zu erfassen. Der Jahrgang alleine kann nicht als Indikator für den Einsatz moderner Technologien verwendet werden. Er kann auch nicht als Indikator für die Vollzugsqualität gewertet werden. Neuere Heizungen sind nicht per se bessere Heizungen. Probleme des Übervollzugs oder des Fehlvollzugs finden keine Beachtung. Aufgrund der angebrachten Bemerkungen ist die Validität dieses Indikators als problematisch einzustufen. Andererseits ist aber auch darauf hinzuweisen, dass es in diesem Fall äusserst schwer fällt, eine valide Operationalisierung dieser Variable zu erreichen.

cb) Kontrolleurspezifische unabhängige Variablen:

Alle kontrolleurspezifischen Variablen mussten nicht auf der Ebene der Gemeinde, sondern auf der Ebene des einzelnen Kontrolleurs erfasst worden sein. Agiert pro Gemeinde nur ein Kontrolleur, stellt dies kein Problem dar. Die Untersuchungseinheiten fallen zusammen. Ist jedoch mehr als ein Kontrolleur in einer Gemeinde tätig, ergeben sich Probleme. Wie die Autoren damit umgegangen sind, geht aus dem Schlussbericht nicht hervor. Es ist zu vermuten, dass in diesen Fällen die Daten in irgendeiner Form aggregiert wurden. Die Validität der kontrolleurspezifischen Indikatoren ist jedoch auf alle Fälle in Frage gestellt. Zur weiteren Beurteilung gehen wir zunächst davon aus, dass der günstigste der denkbaren Fälle vorliegt, dass nämlich pro Gemeinde nur eine Person als Kontrolleur tätig ist. Damit nun also zu den einzelnen Indikatoren.

- *Anstellungsart des Kontrolleurs: Voll- oder Nebenamt:* Die Autoren gehen davon aus, dass Feuerungskontrolleure im Vollamt professioneller und besser arbeiten als Personen, welche diese Aufgabe im Nebenamt wahrnehmen. Unter Beachtung einer *ceteris paribus* Klausel ist die Annahme, dass die Anstellungsart einen Hinweis auf die Professionalität des Kontrolleurs gestattet, durchaus plausibel.
- *Beruf des Kontrolleurs: Gemeindebeamte, Kaminfeger, Ölfeuerungsmonteure und andere Berufe:* Für die vier 'Berufsgruppen', die hier genannt werden, ist die Ausschliesslichkeit nicht gegeben. Das heisst, dass beispielsweise ein gelernter Kaminfeger auch als Gemeindebeamter tätig sein könnte. Beruf und Stellung werden vermischt. Ein weiteres Problem besteht darin, dass die Kategorien zu wenig präzise definiert sind. Ist ein *Gemeindeangestellter* der Gruppe der 'Gemeindebeamten' zuzuordnen oder gehört er in die Kategorie 'andere Berufe'? Im Kommentar zum empirischen Datenmaterial wird die Verwirrung noch verstärkt durch folgende Feststellung: "Gemeindeangestellte sind im Kanton Bern besonders häufig Kontrolleure." (SB: 104, Hervorhebung TW) Aufgrund der Zahlenangaben in einer Tabelle erreicht der Kanton Basel-Land einen höheren Anteil an *Beamten* als der Kanton Bern (Darstellung 5.4., SB: 105). Der Schlussbericht gibt keine Auskünfte zu diesen Problemen. Die Validität dieses Indikators ist nicht gewährleistet. Der Qualifikationsgrad der Berufsausbildung kann so nicht erfasst werden. Die Berufsausbildung und die Stellung im Beruf sollten getrennt beobachtet werden.
- *Qualität der Kontrolleure:* Die Angaben zur Qualität der Kontrolleure sind durch eine Umfrage bei der für die Feuerungskontrolle verantwortlichen Beamten des jeweiligen Kantons erhoben worden. Da es sich dabei um die Einschätzung durch drei verschiedene Einzelpersonen handelt, welche Qualitäten ein guter Kontrolleur aufzuweisen hat, wie diese zu gewichten sind und wie hoch die Ansprüche gesetzt werden sollen, sind diese Angaben zumindest für den interkantonalen Vergleich wenig geeignet. Weiter verfügt dieser Indikator nicht über ein intervallskaliertes Messniveau, wie es für die Bildung arithmetischer Mittelwerte (vgl. Darstellung 5.4., SB: 105) zwingend erforderlich ist. Selbstverständlich reicht das Messniveau auch für die Regressionsanalyse nicht aus (SB: 108-114). Aufgrund der empirischen Werte in Darstellung 5.4. (SB: 105) zeigt sich die Problematik sehr deutlich. Geht man von der Annahme aus, dass die Qualität der Kontrolleure in allen drei Kantonen gleich hoch ist, müssten die Daten bezüglich Zentralität und Dispersion standardisiert werden, was bei ordinal-skaliertem Datenmaterial technisch nicht ganz einfach ist. Die Autoren gehen aber von der Annahme aus, dass sich die Mittelwerte pro Kanton auf effektive Qualitätsunterschiede der Kontrolleure in den drei Kantonen zurückführen lassen. Wir haben oben darauf hingewiesen, dass dieser Schluss nicht möglich ist. Aus diesen Gründen ist dieser Indikator nicht ausreichend valide. Andererseits ist an dieser Stelle auch darauf zu verweisen, dass alternative Erfassungsmöglichkeiten mit höherer Validität kaum auszumachen sind. Aus dieser Perspektive

heraus erscheint die gewählte Operationalisierung als vertretbar. Eine kritische Diskussion der Validitätsproblematik ist unseres Erachtens dennoch erforderlich.

- *Amtsdauer des Kontrolleurs*: Dieser Indikator wird nicht, wie das zu vermuten wäre, in Zeiteinheiten erfasst. Stattdessen werden drei Kategorien gebildet, nämlich 'bis 5 Jahre' (kurze Amtsdauer), '5 - 10 Jahre' (mittlere Amtsdauer) und 'mehr als 10 Jahre' (lange Amtsdauer) (siehe Darstellung 5.4., SB: 105 und Darstellung 5.7., SB: 110). Diese Operationalisierung ist grundsätzlich als valide zu bewerten. Es sei hier jedoch trotzdem noch eine Zusatzbemerkung gestattet. Uns erscheint diese Kategorisierung nicht zweckmässig. Und zwar aus folgendem Grund: Die Analyse erfolgt ja mittels einer multiplen Regressionsanalyse. Die offensichtlich ordinale Charakteristik der gewählten Erfassungsart zwingt die Autoren, diese Variable für die Auswertung in zwei dichotome Merkmale ('dummies') aufzuteilen. Dies ist wenig sinnvoll, da dadurch für die Regressionsschätzung Freiheitsgrade verlustig gehen. Wäre die Amtsdauer von Beginn weg in Zeiteinheiten erfasst worden, hätte sich dieser Kunstgriff erübrigt. Der Informationsgehalt - und damit die Aussagekraft - des Indikators hätte so gleichzeitig noch erhöht werden können.
- *Beanstandungsquote*: Hier sind sich die Autoren nicht ganz im Klaren darüber, in welche Variablen­gruppe (kontrolle­urspezifisch oder nicht) diese Variable einzuordnen ist. Sie schreiben: "Schliesslich hängt auch die kommunale Beanstandungsquote von den Kontrolleuren ab." (SB: 101) Infolgedessen ordnen sie diese Variable zunächst der Gruppe kontrolle­urspezifischer Faktoren zu. Bei der Darstellung der empirischen Daten hingegen finden sich diese Variablen nicht in Darstellung 5.4. ("Art und Qualität der Kontrolleure", SB: 105), sondern bei den übrigen Faktoren (Darstellung 5.5., SB: 106). Diese Inkonsistenz ist unseres Erachtens darauf zurückzuführen, dass dieser Indikator auf der inhaltlichen Ebene zwar ein kontrolle­urspezifisches Merkmal darstellen mag, dass die empirischen Daten aber auf der kommunalen Ebene erfasst wurden.

cc) *Weitere unabhängige Variablen*:

- *Kantonszugehörigkeit der Gemeinde*: Dieser Indikator ist als valide zu betrachten.
- *Anzahl kontrollierter Anlagen*: Bei dieser Variablen liegen die Probleme vor allem darin begründet, dass nicht ganz klar ist, welches Konstrukt damit erfasst werden soll. Soll der Indikator die Gemeindegrösse erfassen, wie dies in der Interpretation suggeriert wird (SB: 111 und 139), ist die Validität nicht gegeben, da die Heizungsanlagen in einem Teil der Gemeinden im jährlichen Rhythmus kontrolliert werden, während andere Gemeinden einem Zweijahresintervall folgen. Weiter kommt hinzu, dass dabei auch die Anlagengrösse mitzuberechnen wäre. Bei der Variablenbeschreibung äussern sich die Autoren folgendermassen: "Weiter haben wir die Anzahl der kontrollierten Anlagen in die Untersuchung

- einbezogen. Mehr Kontrollen bedeutet mehr Professionalität und mehr Sanierungsverfügungen, was zur Folge hat, dass tendenziell ein höherer Anteil moderner Anlagen in der Gemeinde stehen." (SB: 101) Die Anzahl der kontrollierten Anlagen bestimmt sich - wenn wir hier von möglichen Vollzugsdefiziten und weiteren Faktoren absehen - einerseits durch die Zahl der Anlagen in einer Gemeinde und andererseits durch den Kontrollrhythmus, der in der Gemeinde zur Anwendung gelangt. Dieser Indikator ist in absoluten Zahlen erfasst und nicht - wie etwa die abhängige Variable - relativ erhoben worden. Der Indikator setzt sich aus zwei verschiedenen Merkmalen zusammen, die einzeln hätten untersucht werden müssen. Da hier nicht klar ist, was der Indikator erfassen soll, kann seine Validität nicht eingeschätzt werden.
- *Kommunale Luftqualität*: Zur operationalen Definition dieser Variablen schreiben die Autoren: "Als Hinweis auf eine schlechte kommunale Luftqualität haben wir die Tatsache interpretiert, dass eine Gemeinde einem speziellen kantonalen Massnahmenplan zur Reinhaltung der Luft unterstellt ist." (SB: 102) In einer kürzlich erschienen Studie über die Massnahmenpläne zur Luftreinhaltung findet sich dazu folgende Aussage: "Ursprünglich hatten wir die Absicht, die geographische Verteilung der Massnahmenggebiete auf einer Karte wiederzugeben, um einen Überblick über die schweizerischen Belastungsgebiete zu erhalten. Wir mussten jedoch feststellen, dass das Einzugsgebiet des Massnahmenplanes und das Gebiet mit übermässigen Immissionen nicht in jedem Fall deckungsgleich sind." (Knoepfel/Imhof/Zimmermann 1994: 27-28) Dies schränkt die Validität des Indikators ein. Da aber validere Indikatoren (beispielsweise flächendeckende Immissionsmesswerte) nicht verfügbar sind, ist die gewählte Vorgehensweise vertretbar.
 - *Gebäudezunahme*: Die Gebäudezunahme als "Veränderung des Gebäudebestandes zwischen 1980 und 1990" (SB: 103) wird in Prozenten gemessen (Darstellung 5.5., SB: 106). Dieser Indikator ist als valide zu bewerten.
 - *Zunahme Anteil Ölheizungen*: Die "Veränderung des Anteils der Ölheizungen am Anlagebestand" (SB: 103) einer Gemeinde wird ebenso in Prozenten erfasst (Darstellung 5.5., SB: 106). Auch hier ist die Validität gegeben.
 - *Anteil Anlagen mit Serviceabonnements*: Dieser Indikator soll Auskunft geben zur Qualität der Heizungswartung. Da nicht ganz klar wird, wie der Indikator erhoben worden ist, könnte möglicherweise die Gültigkeit der operationalen Definition in Frage gestellt sein. Beispielsweise muss eine Heizung nicht im Rahmen eines Serviceabonnements unterhalten werden. Es sind auch andere Formen denkbar. Da dies nur auf eine relativ kleine Anzahl von Anlagen zutreffen dürfte, reicht die Validität des Indikators aus.

Die Auswahl dieser elf Variablen erfolgte "aufgrund von Vorabklärungen" (SB: 100). Diese Variablen fliessen somit in ein induktiv hergeleitetes Regressionsmodell ein. Die zugrundeliegenden Hypothesen sind ad-hoc-Erklärungen. Bei dieser Vorgehensweise kann natürlich immer auf

Merkmale hingewiesen werden, die auch noch in das Modell hätten integriert werden müssen oder die trotz mangelnder Relevanz Aufnahme ins Modell gefunden haben. Wir werden auf die Regressionsanalyse zu einem späteren Zeitpunkt nochmals zu sprechen kommen (vgl. unten Kapitel 3.4.8)

Aufgrund der angeführten Überlegungen stellen wir fest, dass bei der Validität in diesem Untersuchungsteil einige Mängel vorliegen. Dem entgegenzuhalten ist aber, dass sich die Evaluatoren darum bemühten, möglichst valide Operationalisierungen zu den relevanten Konstrukten zu finden und dass sich in einigen Fällen kaum bessere Alternativen hätten einsetzen lassen. Es zeigt sich, dass der gewählte Ansatz äusserst hohe Anforderungen an die operationale Qualität stellt, die äusserst schwierig zu implementieren ist. So betrachtet sind vor allem die hohen Ansprüche - und nicht etwa eine ungenügende Umsetzung - für die aufgeführten Schwächen verantwortlich. Kritischer ist dagegen der Aspekt zu bewerten, dass im Schlussbericht kaum auf die Validitätsproblematik eingegangen wird. Unter diesen Umständen gelangen wir zum Schluss, dass die Evaluation das Kriterium nur teilweise zu erfüllen vermag.

d) Zeitreihenanalyse Gesamtprogramm (SB: 115-131):

Die Autoren gehen von folgender Prämisse aus: "Technische Entwicklungen im Bereich der Heizungen haben zur Folge, dass neuere Anlagen geringere Schadstoffemissionen aufweisen als ältere und dass sie weniger Energie verbrauchen." (SB: 115) Diese Annahme wird aufgrund des Datenmaterials, wie es von den Autoren präsentiert wird (vgl. Darstellung 6.8., SB: 125 und Darstellung 6.9., SB: 126), nicht gestützt. Dieser Umstand wird von den Autoren in keiner Weise diskutiert. Es gäbe dafür aber eine relativ einfache Erklärung. Das zitierte Datenmaterial beruht auf einer Erhebung der Feuerungskontrollergebnisse in der Stadt Luzern im Jahre 1991. Die Jahrgänge der Kessel der Heizungen werden (stellvertretend für das Alter der Heizungen) in Beziehung gesetzt mit Indikatoren wie NO₂-Emissionen (als Hinweis für NO_x-Emissionen) und den Abgasverlusten. Diese Vorgehensweise führt dazu, dass die Auswahl der zur Verfügung stehenden Daten über die Jahrgänge hinweg, auf unterschiedliche Weise zustandekommt. Das Datenmaterial ist 'zensoriert', das bedeutet, dass die Werte für frühe Jahrgänge (etwa 1960) nur auf Heizungsanlagen beruhen, die im Jahre 1991 noch in Betrieb sind und also nur einen kleinen Anteil jener Anlagen darstellen, welche im betreffenden Jahr (etwa 1960) installiert worden waren. Da es nicht von der Hand zu weisen ist, dass diese Auswahl auch aufgrund von hier interessierenden Faktoren (wie Emissionsverhalten und sparsamer Energieverwendung) zustandekommt, fallen die Resultate der Erhebung dementsprechend verzerrt aus. An einer Stelle kommen die Autoren auf diese Problematik zu sprechen, und zwar im Zusammenhang mit der Beanstandungsquote für Heizungen mit unterschiedlichen Jahrgängen. Sie schreiben: "Es ist aber bemerkenswert, dass der Zusammenhang zwischen Kesseljahrgang und Beanstandungsquote in

unseren Daten nur sehr schwach ausgeprägt ist. Dies kann darauf zurückgeführt werden, dass ältere Heizungen, die zu Problemen Anlass gaben, bereits ersetzt wurden. Wichtiger scheint aber zu sein, dass die heute gültigen Grenzwerte die Schwachstellen alter, überdimensionierter Anlagen kaum einzufangen vermögen." (SB: 124) Die Evaluatoren klären den Leser nicht auf, weshalb die heutigen Grenzwerte bei älteren Anlagen versagen. Erstaunlicherweise gehen die Autoren auch später auf diese Thematik nicht mehr ein, obwohl dieser Aspekt doch von zentraler Bedeutung für die Evaluation zu sein scheint. Warum dieser Aspekt 'wichtiger' sein soll als der Zensorierungseffekt, wird nicht begründet. Andererseits lässt sich der Zensorierungseffekt auch kaum genauer abschätzen, erst recht nicht, wenn die Ausfallraten nicht bekannt sind.

Der andere Teil der Zeitreihenanalyse des Gesamtprogramms stützt sich auf Datenmaterial von Feuerungskontrollen in der Stadt Zürich aus den Jahren 1964 bis 1991 oder 1992. Es handelt sich dabei somit um echte Zeitreihendaten. Wie dies bei dieser Untersuchungsanlage häufig der Fall ist, stellen sich bezüglich der Validität hier zwei Probleme. Erstens bestehen teilweise erhebliche Datenlücken (vgl. Kapitel 3.4.4.). Zweitens besteht das Problem, dass sich die Qualität des Datenmaterials über die Zeit hinweg verändern kann, zum Beispiel aufgrund wechselnder Vorgehensweisen bei der Datenerhebung (in unserem Fall etwa der angewandten Messverfahren). Da die Autoren keine Informationen dazu liefern, ist es uns nicht möglich, eine fundierte Evaluierung der Validität dieses Datenmaterials vorzunehmen. Es ist hier noch eine weitere Anmerkung anzufügen. Soweit uns Informationen vorliegen, bezieht sich das verwendete Datenmaterial nicht ausschliesslich auf Feuerungskontrollen von Anlagen, die für diese Studie von Interesse sind (Leistungsklasse 350kW bis 1 MW, evtl. Gasheizungen).

Zur Bewertung der Validität der Gesamtuntersuchung ist vorzuschicken, dass zwischen den einzelnen Untersuchungsteilen sehr grosse Unterschiede bestehen, die unseres Erachtens mit dem Schwierigkeitsgrad der eingesetzten Verfahren in Zusammenhang stehen. Die Evaluatoren haben sich darum bemüht, eine hohe Validität der Ergebnisse zu erreichen, was in vielen Fällen auch gelungen ist. Andererseits zeigt sich in jenen Untersuchungsteilen, die sehr hohe Anforderungen stellen, dass die Etablierung einer hohen Validität teilweise nicht zu realisieren war. Wir kommen zum Schluss, dass die Evaluation das Kriterium der validen Messung teilweise erfüllt.

3.4.6. Reliable Messung

Da ein direkter Zusammenhang zwischen Validität und Reliabilität besteht, werden wir uns hier auf Aspekte beschränken, die nicht schon im Zusammenhang mit der Validität diskutiert wurden. Einige der bei der Validitätsdiskussion (vgl. oben Kapitel 3.4.5.) angeschnittenen Problempunkte sind aber auch bezüglich der Reliabilität der Messungen relevant.

a) *Fallstudie Grenzwertsetzung (SB: 28-56):*

Die Zuverlässigkeit der Informationen, wie sie durch die Fallstudie hervorgebracht werden, ist als ausreichend zu bewerten. Wie bei Fallstudien üblich, stellen sich natürlich auch hier Fragen bezüglich der Verlässlichkeit der Informationsauswahl, der Gewichtung der Informationen etc. Hinsichtlich der Auswahl der Informationen also etwa:

- Warum wurde in der Fallstudie die bereits im März 1988 publizierte Studie, welche von der City-Vereinigung in Auftrag gegeben wurde, nicht erwähnt (City-Vereinigung 1988)? Die Broschüre vom Juni 1989 wurde dagegen in die Fallstudie aufgenommen (vgl. 19. Sequenz, SB: 40-41).
- Warum fehlen in der Fallstudie die parlamentarischen Vorstösse zum Thema, wie etwa die Motion Hess vom 19. Dezember 1986 (Nationalrat Peter Hess (CVP ZG), Amtl. Bulletin, Nationalrat, 1987 I: 269 ff.) oder die Interpellation Lüchinger vom 18. Dezember 1986 (Nationalrat Hans-Georg Lüchinger (FDP ZH und damaliger Präsident der City-Vereinigung), Amtl. Bulletin, Nationalrat, 1987 I: 271 ff.)

Es gäbe noch eine Vielzahl weiterer derartiger Fragen. Deshalb jedoch die Zuverlässigkeit der Fallstudie in Frage stellen zu wollen, wäre falsch, bestehen doch solche Fragen bei jeder Fallstudie, besonders dann, wenn man versucht, die Darstellung auf die wichtigsten Aspekte zu beschränken.

Etwas Erstaunen könnte der Umstand erregen, dass die Kreise, welche die Massnahmen zu finanzieren haben, also die Hauseigentümer beziehungsweise (durch die Überwälzung auf die Mietzinse) die Mieterschaft, im Prozess, wie er geschildert wird, gar nicht in Erscheinung treten. Dies hat damit zu tun, dass auch von dieser Seite ein Konsens über den Handlungsbedarf bestanden hat und dass die Kosten (im Regelfall) vergleichsweise niedrig waren, besonders auch weil sich die Modernisierung der Heizungstechnologie positiv auf die Betriebskosten auswirkten (Einsparung von Energie).

Die Fallstudie kann als reliabel bezeichnet werden.

b) *Marktanalyse Typenprüfung (SB: 57-83):*

Die Reliabilität des Datenmaterials, welches in der Marktanalyse beigezogen wird, ist mehrheitlich nicht gewährleistet. Die Angaben beruhen häufig auf Schätzungen und Hochrechnungen, die von ihrer Zuverlässigkeit her als eher schwach zu beurteilen sind. Die Evaluatoren haben sich darum bemüht, innerhalb des durch die schmale Datenbasis (siehe Kapitel 3.1.3.) eingeschränkten Rahmens, möglichst zuverlässiges Datenmaterial zu beschaffen und zu verwenden.

c) Quantitative Analyse Feuerungskontrolle (SB: 84-114):

Die Aussagen zu den drei Modellen des Vollzugs der Feuerungskontrolle sind bezüglich ihrer Reliabilität schwierig einzustufen, da keine genaueren Angaben vorliegen, wie die Informationen erarbeitet wurden (siehe Kapitel 3.4.3.). Die Untersuchung der Reliabilität des quantitativen Untersuchungsteils ist deshalb nicht möglich.

d) Zeitreihenanalyse Gesamtprogramm (SB: 115-131):

Die Zuverlässigkeit des Datenmaterials ist in unterschiedlichem Ausmass gewährleistet. Bei jenen Daten, welche auf einer Erhebung der Feuerungskontrolldaten aus dem Jahre 1991 beruhen, ist sie durch den Zensorierungseffekt tangiert (vgl. dazu Kapitel 3.4.5.). Die Daten zu den NO₂-Emissionen (Darstellung 6.8., SB: 125) sind in Frage zu stellen, da zur Qualität des Heizöls keine Angaben vorliegen. Der schwankende Stickstoffgehalt des Heizöls kann unter Umständen zu stark verzerrten Messwerten führen. Inwiefern diese Verzerrungen systematischer Art sind, kann nicht beurteilt werden. Die Problematik der sich über die Zeit ändernden Messverfahren bei echten Zeitreihendaten stellt sich auch bei der Reliabilitätsdiskussion. Insgesamt gesehen liegt in diesem Untersuchungsteil nur eine eingeschränkte Reliabilität vor.

Über alle Untersuchungsteile hinweg, kommen wir zum Schluss, dass die Evaluation das Kriterium der Reliabilität teilweise erfüllt.

3.4.7. Systematische Datenüberprüfung

Da die Angaben zur Vorgehensweise nur lückenhaft vorliegen (siehe Kapitel 3.4.3.), kann nur indirekt auf die Qualität der Datenüberprüfung geschlossen werden. Aufgrund der aufgezeigten punktuellen Ungenauigkeiten und Fehler hätte hier sicherlich noch Verbesserungspotential bestanden. Da uns jedoch zu diesem Kriterium keine ausreichenden Informationen vorliegen, verzichten wir auf eine Bewertung.

3.4.8. Analyse quantitativer Informationen

Wegen den unterschiedlichen Vorgehensweisen in den Teilstudien beschränken wir uns hier auf die Beurteilung der Analyse quantitativer Informationen zur Feuerungskontrolle und zum Gesamtprogramm. Die übrigen Untersuchungsteile konzentrieren sich eher auf qualitative Analysen (siehe dazu Kapitel 3.4.9.).

a) *Quantitative Analyse Feuerungskontrolle (SB: 84-114):*

Aufgrund von "Vorabklärungen" (SB: 100) entwickeln die Evaluatoren ein Regressionsmodell mit elf exogenen Faktoren. Wir sind bereits anlässlich der Validitätsproblematik auf diese und auf die abhängige Variable, der Anteil moderner Anlagen in einer Gemeinde, zu sprechen gekommen.

Die Evaluatoren gehen so vor, dass sie

1. die Hypothesen formulieren (SB: 101-103, zusammenfassend Darstellung 5.2., SB: 102),
2. die Stichprobe, auf der das Datenmaterial beruht, beschreiben (SB: 103-104),
3. eine univariate bzw. (unter Bezug der Kantonszugehörigkeit der Gemeinden) eine bivariate Analyse durchführen (SB: 104-108), um dann
4. eine multivariate Regressionsanalyse zu erarbeiten und zu interpretieren (SB: 108-114).

aa) *Zum ersten Schritt:*

Die Hypothesen erscheinen soweit plausibel. Es stellt sich die Frage, weshalb andere auf der Hand liegende Faktoren (wie Kontrollrhythmus der Feuerungskontrolle, Tradition der Feuerungskontrolle in der Gemeinde, 'Leistungsklassen-Split' in den Gemeinden usw.) nicht in die Analyse einbezogen worden sind. Problematisch erscheint dabei nicht primär, dass diese Variablen in der Regressionsanalyse fehlen, sondern vielmehr, dass die dahinterstehenden Überlegungen nicht dargestellt werden. Andererseits hätten derartige Ausführungen dem Entscheid der Evaluatoren widersprochen, die Berichterstattung möglichst von methodisch ausgerichteten Ausführungen zu entlasten.

Bei drei Fällen erweist es sich als schwierig, aufgrund der Vorüberlegungen festzustellen, ob eher ein positiver oder ein negativer Wirkungszusammenhang bestehen könnte. Für beide Spezifikationen bestehen nämlich stringente Argumentationslinien. Dies betrifft die unabhängigen Variablen 'Beanstandungsquote', 'Kantonszugehörigkeit' und 'Anteil Anlagen mit Serviceabonnements'. Diese Konstellation ist für die (kausale) Interpretation der empirischen Resultate der Regression problematisch.

ab) *Zum zweiten Schritt:*

Wie bereits erwähnt, bezieht sich die quantitative Querschnittsanalyse ausschliesslich auf das sogenannte 'Luzerner Modell', welches in vierzehn Kantonen angewandt wird (siehe Darstellung 5.1., SB: 90). Aus diesen Kantonen wählen die Autoren die drei Kantone Basel-Land, Bern und Luzern aus. Für diese Auswahl bleiben die Evaluatoren eine Begründung schuldig. Aufgrund der

Datenqualität konnten von den insgesamt 592 Gemeinden der drei Kantone nur 419 Gemeinden berücksichtigt werden. In den Kantonen Basel-Land (72 aus 73 Gemeinden) und Luzern (100 aus 107) ist dies unproblematisch. Im Falle des Kantons Bern hingegen ist die Situation schwieriger. Nur von 247 der insgesamt 412 Gemeinden liegt ausreichendes Datenmaterial vor, das entspricht immerhin einer Ausfallquote von rund 40 Prozent. Die Evaluatoren bemerken dazu: "Es handelt sich um eine weitgehend repräsentative Auswahl." (SB: 103) Weshalb hier Repräsentativität vorliegen soll, lassen sie offen. Die Evaluatoren nennen zwei Anforderungen an das Datenmaterial zur Zulassung einer Gemeinde in die Untersuchung:

1. Mindestens zehn Prozent des Anlagenbestandes müssen in der Kontrollperiode 1990/91 kontrolliert worden sein.
2. Es müssen "Angaben über den Anteil der Kessel mit Jahrgang 1986 und jünger verfügbar" (SB: 103) sein.

Aufgrund dieses Anforderungsprofils an das Datenmaterial sind also auch Gemeinden in der Stichprobe enthalten, welche bei exogenen Faktoren fehlende Werte aufweisen. Leider fehlen dazu jegliche Angaben. Es wird nicht klar, wie gross die Zahl der analysierten Fälle in den uni-, bi- und multivariaten Analysen ist. Die Darstellung der Analyseresultate erfolgt jeweils ohne Nennung der relevanten Fallzahl. Möglicherweise wurden die fehlenden Werte auch durch Mittelwerte ersetzt oder andere Techniken der Substitution verwendet, was zu einer konstanten Fallzahl führen würde. Hier besteht Klärungsbedarf.

ac) Zum dritten Schritt:

Aufgrund der Mittelwerte der Angaben für alle Gemeinden zusammen und für die Kantone separat werden die zehn verbleibenden exogenen Variablen und die abhängige Variable einer univariaten Analyse unterzogen. Als Analyseinstrumente werden arithmetische Mittel und teilweise Standardabweichungen verwendet. Dabei sind drei Punkte zu kritisieren:

1. Die Berechnung des arithmetischen Mittels bei ordinal-skalierten Variablen ist, obwohl recht häufig anzutreffen, nicht zulässig (so im Falle der Angaben zur Qualität der Feuerungskontrolleure, siehe Darstellung 5.4., SB: 105).
2. Die Verwendung der Standardabweichung zum Vergleich der Streuung verschiedener empirischer Verteilungen ist nicht zweckmässig, da dieser Vergleich durch Differenzen in den Mittelwerten verzerrt ist. Dafür geeignet ist der Variationskoeffizient.
3. Die Nichtbeachtung unterschiedlicher Untersuchungsniveaus, wie es bereits oben dargelegt worden ist, führt zu Problemen. Die Angaben in der Tabelle sind zweideutig. So ist nicht klar,

ob sich die prozentualen Angaben in der Tabelle "Art und Qualität der Kontrolleure" auf die Gemeinden oder auf die einzelnen Kontrolleure beziehen (Darstellung 5.4., SB: 105). Haben nun 67 Prozent der Gemeinden einen Feuerungskontrolleur im Vollamt oder sind 67 Prozent der Kontrolleure im Vollamt tätig? Gilt ersteres, stellt sich sofort die Frage, was denn nun mit jenen Gemeinden geschehen ist, in denen mehrere Kontrolleure im Voll- und Nebenamt wirken.

Im Anschluss daran wird von einer Varianzanalyse gesprochen, die auf "systematische Unterschiede der Grössen in den drei Kantonen" (SB: 108) hinweisen soll. Aus den Angaben der Autoren zu dieser Auswertung wird nicht klar, welche Art von Varianzanalyse benutzt wurde.

ad) Zum vierten Schritt:

Die Evaluatoren haben insgesamt fünf verschiedene Regressionsschätzungen vorgenommen. Das erste Modell ist für die Gemeinden aller Kantone geschätzt worden, die Schätzungen drei bis fünf jeweils für einen Kanton. Für die zweite Schätzung ist ein revidiertes Modell mit einer zusätzlichen Variable für alle Gemeinden beigezogen worden. Die zusätzliche Variable ist aus zwei bestehenden Variablen gebildet worden, nämlich einerseits aus der Zugehörigkeit zum 'Kanton Luzern' und andererseits aus der Variable 'Qualität der Kontrolleure'. Dieses Modell ist sicherlich fehlspezifiziert, da die zusätzliche Variable eine lineare Kombination der beiden Ursprungsvariablen darstellt. Dieser Umstand führt per definitionem zu hoher Multikollinearität dieser drei exogenen Faktoren. Die Multikollinearität exogener Faktoren wird im Schlussbericht mit keinem Wort bedacht, obwohl dies aufgrund inhaltlicher Überlegungen auf der Hand liegt. Die beiden Variablen 'Zunahme Anteil Ölheizungen' und 'Gebäudezunahme' beispielsweise korrelieren wahrscheinlich relativ hoch miteinander.

Es fehlen jegliche Angaben zum verwendeten Schätzverfahren. Es ist zu vermuten, dass mit der Methode der kleinsten Quadrate (ordinary-least-squares, OLS) gearbeitet wurde. Angaben über die Fallzahlen fehlen. Durch 'listwise deletion' könnte eine grössere Anzahl von Gemeinden für die Schätzungen ausser Betracht gefallen sein.

Die Evaluatoren machen nur sehr knappe Angaben zu den Schätzergebnissen. Sie führen die Schätzungen der standardisierten partiellen Regressionskoeffizienten, die jeweils erreichten Signifikanzniveaus (0,1%, 1% und 5%) und die bereinigten R^2 -Schätzungen an. Die nicht-standardisierten Schätzungen fehlen. Das ist bedauerlich, da nur sie gestatten würden, die Schätzungen Eins und Zwei miteinander zu vergleichen. Die standardisierten Parameterschätzwerte lassen den Vergleich unterschiedlich spezifizierter Modelle nicht zu. Für einen derartigen Vergleich sind die nicht-standardisierten Regressionskoeffizienten zu verwenden.

Die Resultate der Schätzungen sind mehrheitlich nicht signifikant. Von den geschätzten Parametern sind in der Schätzung Eins 5 von 15 signifikant, in der Schätzung Zwei 5 von 16, in der Schätzung Drei 1 von 13 und in den Schätzungen Vier und Fünf 1 von 12. Zudem sind die Schätzungen bezüglich ihrer Signifikanz sehr volatil. Die Anzahl zu schätzender Parameter weicht von der Anzahl der Variablen ab, weil einzelne Variablen dichotomisiert werden mussten, da sie in ordinaler oder nominaler Form vorlagen. Dies betrifft die Variablen Beruf (drei 'dummies'), Amtsdauer (zwei 'dummies') und Kantonszugehörigkeit (zwei 'dummies'). Die Variable Qualität der Kontrolleure wird als intervallskalierte Variable verwendet, obwohl dies im Prinzip nicht zulässig wäre. Sie beruht ja auf der Einschätzung durch die Verantwortlichen bei den Kantonen, welche die Qualität 'ihrer' Kontrolleure mit 'sehr gut', 'gut', 'durchschnittlich', 'unterdurchschnittlich' und 'schlecht' qualifizierten.

Die bereinigten R^2 -Schätzungen erreichen maximal den Wert von .19. Unter Berücksichtigung der Anzahl Regressoren - in diesem Fall 16 - ist (überspitzt formuliert) festzustellen, dass ein solcher Erklärungsgehalt auch mit zufallsgenerierten exogenen Faktoren erreichbar ist. Die Autoren schreiben dazu: "Das ist zwar wenig, der Wert liegt aber in einer für sozialwissenschaftliche Untersuchung üblichen Grössenordnung." (SB: 109) Das mag vielleicht für eine bivariate Regressionsanalyse zutreffen, bei multiplen Regressionen (mit einer relativ hohen Regressorenzahl) liegen die entsprechenden Werte - auch in den Sozialwissenschaften - doch etwas höher (zumindest bei jenen Regressionsmodellen, die nicht verworfen werden).

Aufgrund der vielfältigen Probleme bei dieser Regressionsschätzung, die noch verstärkt werden durch die oben diskutierten Validitäts- und Reliabilitätsfragen (siehe Kapitel 3.4.5. und 3.4.6.), kommen wir zum Schluss, dass diese Analyse nicht in der Lage ist, haltbare Interpretationen sicherzustellen.

In den methodischen Reflexionen äussern sich die Autoren zum Einsatz quantitativer Verfahren in der Evaluationsforschung folgendermassen: "Quantitative Verfahren tendieren dazu, die hochgradigen Interdependenzen von Variablenzusammenhängen und die Mannigfaltigkeit der Bedingungsfaktoren von Implementationsprozessen zu vernachlässigen." (SB: 162) Unseres Erachtens lässt sich dieser Schluss aufgrund der durchgeführten Regressionsanalyse, welche gerade im Bereich der Multikollinearität (also der Interaktion verschiedener exogener Faktoren) Mängel aufweist, nicht ziehen. Die Qualität eines Verfahrens hängt nicht zuletzt von der Qualität der Anwendung ab, wobei eine ausreichende Datenqualität eine Voraussetzung bildet. Zweitens ist darauf hinzuweisen, dass es nicht zulässig erscheint, aufgrund der Erfahrungen mit *einem* quantitativen Verfahren, Schlüsse auf die gesamte quantitative Methodik zu ziehen.

b) *Zeitreihenanalyse Gesamtprogramm (SB: 115-131):*

Hier ist zuerst eine Vorbemerkung anzubringen. Dieses Kapitel enthält zwei verschiedene Analysesteile, welche sich auf quantitatives Datenmaterial abstützen. Die erste Analyse (SB: 115-122) basiert auf Zeitreihendaten im üblichen Sinne. Das heisst, es wird der Verlauf verschiedener Merkmale über die Zeit hinweg betrachtet. Der zweite Teil (SB: 122-128) dagegen ist *keine* Zeitreihenanalyse, sondern eine Querschnittsanalyse, da die Datenerhebung zu einem Zeitpunkt (bzw. im Zeitraum einer Kontrollperiode) erfolgte. Dies gilt auch dann, wenn eine der erfassten Variablen - hier das Alter der Heizungen - einen zeitlichen Aspekt beinhaltet.

Die Autoren beschränken sich bei der Analyse des Gesamtprogramms auf "intelligente Interpretationen des Datenmaterials" (SB: 163). Um dies zu erleichtern, werden die Daten optisch aufbereitet. Im Gegensatz zur Darstellung der Autoren sind darüber hinaus keine analytischen Schritte im Schlussbericht enthalten. Die Autoren dazu: "Für die Analyse der Beobachtungen zogen wir einfache statistische Masszahlen, wie Mittelwerte und Standardabweichungen bei. Ferner haben wir gleitende Mittelwerte berechnet und Interpretationen für markante Abweichungen gesucht." (SB: 164) In der Darstellung der Analyse sucht man vergebens nach diesen Analysen. Einzig die 'Interpretation markanter Abweichungen' ist im Schlussbericht enthalten.

Die echte Zeitreihenanalyse basiert auf Daten aus der Stadt Zürich zu folgenden Merkmalen:

- Beanstandungsquote von 1964 bis 1992 (ohne 1970) in % (Darstellung 6.2., SB: 118).
- Mittlere Abgasverluste von 1975 bis 1991 (ohne 1976, 1977 und 1986) (Darstellung 6.3., SB: 119).
- Anteil maximal fünfjähriger Kessel von 1964 bis 1991 (ohne 1965 bis 1969, 1971 bis 1974, 1976, 1977, 1979 und 1986, ergänzt mit Angaben zum Anteil Neu- und Umbauten am Bestand für die Jahre 1981 bis 1990) (Darstellung 6.4., SB: 120).
- Durchschnittsalter der Heizkessel von 1980 bis 1991 (ergänzt mit Angaben aus dem Kanton Basel-Land für die Jahre von 1987 bis 1991) (Darstellung 6.5., SB: 122).

Die Interpretationen der Zeitreihen erscheinen uns - unter Vorbehalt der bereits in den Kapiteln 3.4.5. und 3.4.6. geäusserten Bedenken - grösstenteils angemessen erfolgt zu sein.

Die Ungenauigkeiten bezüglich der relevanten Verordnungen zeigen hier ihre Auswirkungen, so etwa bei der Analyse der energetischen Vorschriften. Darstellung 6.2. zeigt, dass die energetischen Beanstandungen im Jahre 1986 sprunghaft zugenommen haben (SB: 118). Die Angaben, welche von den Autoren ungenauerweise mit 1986 beschriftet wurden, beziehen sich auf die Heizperiode 1985/1986. Die Autoren interpretieren den Anstieg der Quote folgendermassen: "Weiter sieht man, wie mit der Einführung der LRV 1986 die Beanstandungen aus energetischen

Gründen stark zugenommen haben." (SB: 117) Die LRV 1986 ist per 1. März 1986 in Kraft getreten. Das bedeutet, dass die Beanstandungen im Winter 1985/1986 kaum auf diesen Rechtserlass zurückzuführen sind. Vielmehr könnte der Anstieg der energetischen Beanstandungsquote auf die per 1. Januar 1985 in Kraft getretene LMFV hinweisen. Vergleicht man die energetischen Vorschriften der LMFV und der LRV 1986, zeigt sich, dass die Vorschriften in der LRV 1986 weniger restriktiv gestaltet sind, als dies in der LMFV der Fall war. Die zulässigen Prozentwerte schwanken in der LMFV zwischen 11 und 13 Prozent. Die LRV 1986 fordert dagegen lediglich, dass sich die Abgasverluste je nach Feuerungswärmeleistung der Anlage unter 14 bzw. 16 Prozent bewegen. Ähnliche Ungereimtheiten zeigen sich auch beim Vergleich, der effektiv erreichten mittleren Abgasverlusten und den Vorschriften in den LRV 1986 und 1992 - auch hier wird die Existenz der LMFV nicht erwähnt (SB: 118-119, Darstellung 6.3., SB: 119). Die Autoren geben an, dass auf "den 1. Januar 1986" ... "für Altanlagen 14 bis 16 Prozent Abgasverluste als Grenzwert festgeschrieben" (SB: 119) wurden. Das trifft nicht zu. Zu diesem Zeitpunkt war noch die LMFV in Kraft, die Grenzwerte der LRV 1986 galten erst ab dem 1. März 1986.

Beim zweiten Teil, der Querschnittsanalyse, verwenden die Evaluatoren Datenmaterial, das aufgrund der Feuerungskontrollen in der Stadt Luzern in der Kontrollperiode 1990/1991 entstanden ist. In einem ersten Schritt wird eine Alterspyramide der Kessel und Brenner interpretiert. Dann werden aufgrund der Jahrgänge der Heizkessel folgende Indikatoren analysiert:

- Beanstandungsquote (Darstellung 6.7., SB: 124)
- Mittlere NO_x -Emissionen (anhand der NO_2 -Emissionen) (Darstellung 6.8., SB: 125)
- Mittlere Abgasverluste (Darstellung 6.9., SB: 126)

Im Gegensatz zu den echten Zeitreihendaten fällt bei diesen Datenreihen deutlich auf, dass die Reduktion deutlich geringer ausfällt als erwartet, relativ stationäre Wertereihen resultieren oder im Extremfall zumindest tendenziell steigende Reihen festzustellen sind. Wir haben den Zensorierungseffekt bereits weiter oben ausführlich dargelegt. Dieses Phänomen lässt eine stichhaltige Interpretation des Datenmaterials praktisch nicht zu. Die einzig möglichen, sinnvollen Interpretationen sind jene, welche auf starke lokale Veränderungen hinweisen. Auf diese gehen jedoch die Autoren kaum ein. Interessant wäre etwa die Frage, weshalb im Jahre 1991 noch sehr viele Anlagen aus dem Jahre 1970 in Betrieb sind (siehe Darstellung 6.6., SB: 123). Oder warum Heizungen mit dem Kesselbaujahr 1986 verhältnismässig häufig Anlass für Beanstandungen boten (siehe Darstellung 6.7., SB: 124). Wegen der mangelnden Aussagekraft des Datenmaterials aufgrund der Zensorierung sind die gemachten Interpretationen nicht ausreichend gestützt. Das Kriterium ist nicht erfüllt.

3.4.9. Analyse qualitativer Informationen

Wir beschränken uns an dieser Stelle auf jene Untersuchungsteile, in denen schwergewichtig qualitative Analyseverfahren zur Anwendung kommen, also erstens die Fallstudie zur Grenzwertsetzung und zweitens die Marktanalyse zur Typenprüfung.

a) Fallstudie Grenzwertsetzung (SB: 28-56):

Die Fallstudie zur Grenzwertsetzung besteht einerseits aus einer Deskription und andererseits aus einer Aktoranalyse. Der analytische Gehalt der chronologischen Beschreibung ist per definitionem klein. Wir konzentrieren uns aus diesen Gründen auf die Aktoranalyse. Diese zeigt - aufgrund rationaler Kosten-Nutzen-Überlegungen - wie sich die Interessenlagen der einzelnen Akteurgruppen verhalten hatten. Etwas überraschend wirkt der Einbezug der vierten Gruppe von Akteuren, der Interessenverbände. Interessant sind die Ausführungen zum Verhalten der Vertreter der automobilen Interessen, welche sich für lufthygienische Massnahmen im stationären Bereich stark machten, um derartige Eingriffe für den Individualverkehr zu verhindern. Dies gilt ebenso für die Feststellung, dass die Umweltverbände am konkreten Aushandlungsprozess einer primär umweltpolitischen Massnahme kaum beteiligt wurden. Ebenso erging es aber auch den Organisationen der Immobilienbesitzer und Mieter. Dieser Hinweis fehlt jedoch in der Studie.

Die Analyse ergibt plausible Resultate. Inwiefern diese durch angemessene und systematische Analysen zustande kamen, kann nicht beurteilt werden, da die entsprechenden Informationen aufgrund des Schlussberichts nicht erschliessbar sind. Das Kriterium kann aber - soweit uns Informationen vorliegen - als erfüllt bezeichnet werden.

b) Marktanalyse Typenprüfung (SB: 57-83):

Zuerst ist darauf hinzuweisen, dass die Evaluatoren ursprünglich für diesen Analyseteil eine quantitative Vorgehensweise vorgesehen hatten, wegen der dürftigen quantitativen Datenbasis dann jedoch fast gänzlich darauf verzichten mussten.

Die Interpretationen der Autoren vermögen gesamthaft gesehen nicht voll zu überzeugen. Den Evaluatoren gelingt es nur punktuell, gegenüber anderen Arbeiten (Wüest & Partner 1992, Balthasar/Müller 1991) Informationen mit Neuigkeitswert zu erarbeiten. Die Evaluatoren beklagen zwar den dürftigen Informationsstand, verzichten jedoch darauf, zur Verfügung stehende Informationen in die Analyse miteinzubeziehen (so etwa Estermann/Feldmann 1990). Die Aussagen haben einen statischen Charakter. So verzichten die Autoren auf den Vergleich der Zahlen zu den Marktanteilen, wie sie im Rahmen dieser Studie erhoben wurden, mit früheren Erhebungen. Zur Illustration dieser Kritik möchten wir ein Beispiel anführen. Andreas Balthasar und

Ueli Müller haben in der 1991 erschienen Vorläuferstudie festgehalten, dass der Importanteil am schweizerischen Heizkessel-Markt rund 45 Prozent beträgt (Balthasar/Müller 1991: 11). In der hier betrachteten Evaluationsstudie wird dieser Anteil mit rund 60 Prozent angegeben (SB: 73). Diese Veränderung kann in Anbetracht der kurzen zeitlichen Differenz als massiv bezeichnet werden. Ein weiteres Beispiel stellt die Darstellung der Rahmenbedingungen dar. Hier wird beispielsweise kein Wort darüber verloren, dass sich diese Rahmenbedingungen über die Zeit hinweg verändert haben. So etwa der Geltungsbereich der Typenprüfung, die lufthygienischen und energetischen Vorschriften, die Messempfehlungen usw. Die Hinweise zum internationalen Vergleich sind sehr bruchstückhaft. So werden lediglich die Vorschriften oder Empfehlungen, welche in Deutschland zur Anwendung gelangen, diskutiert. Es "kann pauschal festgehalten werden, dass die neuen Anforderungen, wie sie die Typenprüfung heute an die Hersteller stellt, im internationalen Vergleich mit Ausnahme einiger lokaler Spezialregelungen (wie in der Stadt Hamburg) am strengsten sind." (SB: 66) Für diese Aussage werden keinerlei Quellenangaben gemacht - wie dies übrigens in diesem Untersuchungsteil mehrfach zu bemängeln ist (siehe etwa Darstellung 4.2., SB: 66 und Darstellung 4.4., SB: 72). Zumindest im österreichischen Bundesland Steiermark sind ähnlich strenge Normen in Kraft (vgl. Landesgesetzblatt (LGBL) der Steiermark, Stück 6, Nr. 33, ausgegeben am 30. Juni 1992: 44).

Bei der Analyse der Marktstruktur (SB: 69-76) sind Inkonsistenzen festzustellen. So etwa bei den Aussagen zur Lebensdauer der Heizkessel. Nach der Studie kann von einer Lebensdauer von 20 bis 25 Jahren ausgegangen werden (SB: 70). Eine Seite später findet sich ein Zitat, das übrigens nicht als solches ausgewiesen wird, aus einer anderen Untersuchung, das folgendermassen lautet: "Die Altersgrenze von +/- 15 Jahren liegt ausserdem bereits im Bereich des 'natürlichen' Ersatzes von Feuerungsanlagen." (Wüest & Partner 1992: 15, vgl. SB: 71). Da die Autoren zuvor feststellten, dass "die Jahrgänge der Kessel stellvertretend für die ganze Heizung" (SB: 71) betrachtet werden, entsteht so ein Widerspruch. Wiederum finden sich fehlerhafte Angaben zum Zeitpunkt der Einführung einer Typenprüfung (SB: 71), die sich auch in der (verschwiegenen) Quelle (Wüest & Partner 1992: 15) finden.

Die Darlegungen aufgrund der Rapporte aus der Feuerungskontrolle sind aus zwei Gründen wenig aussagekräftig:

- *Terminierung*: Die Daten zum Heizungsbestand in der Stadt Luzern und im Kanton Bern werden in drei Altersgruppen (Jahrgänge 1971-1979, 1980-1985 und 1986-1991) aufgeteilt. Neben dem Zensorierungseffekt kommt erschwerend hinzu, dass die Bildung der Altersgruppen nicht gemäss der Einführung der Typenprüfung erfolgte.
- *Leistungsklassen*: Das Datenmaterial erfasst zu einem wesentlichen Teil auch Anlagen, die erst durch die LRV 92 einer Typenprüfung unterstellt wurden oder auch heute noch nicht

einer solchen unterstellt sind. Die Angaben aus der Stadt Luzern beispielsweise beziehen sich auf Heizungsanlagen mit Leistungen bis zu 1 MW. Gemäss LMFV und LRV 86 sind nur Anlagen bis 70 kW einer Typenprüfung unterstellt. Die LRV 92 erweitert den Geltungsbereich auf Anlagen bis 350 kW.

Die Marktanalyse weist eine weitere bedeutsame Lücke auf. Auf die *Preisentwicklung* kommt die Analyse überhaupt nicht zu sprechen. Es findet sich lediglich ein Hinweis zu den Preisen der Heizungsanlagen, und zwar jener, dass das Preisniveau in der Schweiz höher liegt als im Ausland (SB: 79-80). Wie sich die Preise im zeitlichen Verlauf verhalten, behandelt die Studie nicht.

Aufgrund der verschiedenen Ungenauigkeiten und Fehlerhaftigkeiten sind die hier vorgenommenen Analysen nicht in der Lage, in allen Teilen plausible Interpretationen sicherzustellen. Das Kriterium wird nur teilweise erfüllt.

3.4.10. Gerechtfertigte Folgerungen

Aufgrund der verschiedenen Analyseschritte führen die Evaluatoren sowohl inhaltliche wie auch methodische Folgerungen an. Wir werden diese getrennt untersuchen.

a) Folgerungen inhaltlicher Art (SB: 136-154):

Generell betrachtet sind die 'wirtschaftspolitischen Schlussfolgerungen' klar ausgeführt, gut begründet und sinnvoll durch das empirische Material abgestützt. Geglückt ist der Versuch der Autoren, die Erfolgsbedingungen eines technisch orientierten Umweltschutzes aufzuzeigen (SB: 153-154). Daraus ergeben sich für die Praxis (Übertragbarkeit auf andere Bereiche) und für die Wissenschaft (induktive Theorieentwicklung) interessante Einsichten. Zu kritisieren ist dabei lediglich, dass es angemessener gewesen wäre, diese Resultate in der Form von Thesen darzulegen, als sie in Form von gesicherten Erkenntnissen zu präsentieren - was sie ja nicht sind.

Die Evaluatoren machen drei konkretere Anregungen zur zukünftigen Politikgestaltung, nämlich:

- die Dynamisierung der Grenzwerte,
- die Rhythmisierung der Revisionen der Typenprüfungen und
- die Qualifizierung der Feuerungskontrolle. (SB: 146)

Die ersten zwei Massnahmen sollen dazu beitragen, "einen permanenten Druck zur technischen Verbesserung der Heizanlagen auszuüben" (SB: 146) und gleichzeitig den Entwicklungsbemühungen der Unternehmen eine klare Orientierung zu geben. Die Autoren gehen leider nicht auf die praktischen Probleme ein, die eine derartige Gestaltung mit sich bringt. Gemeint

sind damit etwa die mangelnden Informationen über den Stand der Technik in zehn oder gar zwanzig Jahren, das Problem der sinkenden Grenzeffekte technischer Umweltschutzmassnahmen usw. Neben diesen inhaltlichen Fragen, die sicherlich einer breiteren Diskussion bedürfen, fällt bei der Lektüre noch etwas weiteres auf. Diese Anregungen werden an die Verwaltung gerichtet. Dies leuchtet nicht ein, da die vorgeschlagene Ausgestaltung der hier diskutierten umweltpolitischen Massnahmen eine Gesetzesänderung voraussetzen würde. Der im Umweltschutzgesetz festgelegte Grundsatz, dass vorsorgliche Emissionsbegrenzungen so weit zulässig sind, "als dies technisch und betrieblich möglich und wirtschaftlich tragbar ist" (USG, Art. 11, Ziff. 2, SR 814.01) steht einer Realisierung entgegen. Diese beiden Vorschläge sind deshalb nicht nur an die Verwaltung, sondern auch an die Politik zu richten. Erst recht, wenn man daran denkt, dass sich das Umweltschutzgesetz zurzeit in der parlamentarischen Debatte befindet. Der dritte Vorschlag der Evaluatoren ist hingegen sehr wohl auf Verwaltungsebene zu realisieren.

Die Evaluatoren gehen in zwei Punkten deutlich über die Grenzen des von ihnen untersuchten Gegenstandsbereichs hinaus. Es handelt sich dabei um zwei Aspekte, die sie zu Beginn der Studie ganz explizit als nicht untersuchte Gegenstände festlegen (SB: 7-8). Angesprochen sind damit einerseits die Hinweise zu den lufthygienischen und energetischen Folgen des untersuchten Programms (SB: 146-147 und 149) und andererseits die Abwägungen zwischen einer technischen und einer ökonomischen Ausgestaltung der Umweltpolitik (SB: 149-154). Diese Aspekte haben die Evaluatoren nicht untersucht. Trotzdem erweitern die Autoren die Aussagen ihrer Studie im wirtschaftspolitischen Schlusskapitel auch auf die ökologischen Aspekte (siehe SB: 146-147). Dies ist aufgrund der durchgeführten empirischen Analysen nicht zulässig. Spricht man in diesem Zusammenhang von einer Verminderung der Abgase und einer effektiveren Nutzung knapper Energie, reicht es nicht aus, lediglich die beim Betrieb der Anlagen verbrauchten Ressourcen und emittierten Schadstoffmengen - die Evaluatoren berücksichtigen lediglich die NO_x-Emissionen - zu betrachten. Hier drängt sich eine holistische Sichtweise des ganzen Produktzyklus auf (Stichworte Öko-/Energiebilanz). Darüber hinaus haben sich die Zielsetzungen 'effektiver Energieeinsatz' und 'Abgasverminderung' als konfligierende Ziele herausgestellt (SB: 125, vgl. Darstellungen 6.8. und 6.9., SB: 125 und 126). Von der Presse wird diese Überinterpretation aufgenommen. So lautet der Titel eines Artikels in der Wochenzeitung 'Die Weltwoche': "Wenn Behörden und Techniker einander stossen, gewinnt die Umwelt: Wettstreit um bessere Luft" (WeWo, 16. 6. 1994: 35). Gerade dies wurde von den Evaluatoren nicht überprüft! (siehe auch Kapitel 3.4.1.) Im Gegensatz zu den energetischen und ökologischen Auswirkungen, wird bei den Überlegungen zur Wahl des 'richtigen' Instrumentariums anhand der zugehörigen Titelüberschrift ("Ausblick", SB: 149) deutlicher, dass hier Folgerungen angeführt werden, welche über den eigentlichen Gegenstandsbereich der Untersuchung hinausgehen.

b) *Folgerungen methodischer Art (SB: 155-171)*

Die Folgerungen methodischer Art reflektieren vorerst den Methodeneinsatz in den verschiedenen Teilstudien, um dann in einem zweiten Schritt diese Überlegungen miteinander zu verbinden.

Die Überlegungen zum Einsatz des *Fallstudienkonzepts* (SB: 156-158) überzeugen durch klare Argumentationen und die Darlegung der Möglichkeiten und Grenzen dieses Ansatzes. Einzig die Zuordnung Harry Eckstein's zur quantitativen Schule wirkt seltsam, hat dieser Wissenschaftler doch gerade seine bekanntesten Arbeiten mit einer schwergewichtig qualitativen Vorgehensweise geschaffen (Eckstein 1966 und 1992, Eckstein/Gurr 1975).

Die Ausführungen zur *Marktanalyse* (SB: 158-160) konzentrieren sich auf zwei Schwerpunkte. Zum einen werden inhaltliche Aspekte wieder aufgegriffen und zum anderen wird auf die Probleme der Umsetzung des Analysekonzepts eingegangen. Sicherlich sind inhaltliche und methodische Fragen nicht immer klar voneinander zu trennen. Trotzdem sei hier bemerkt, dass die angefügten inhaltlichen Hinweise keinen Beitrag zur methodischen Reflexion zu leisten vermögen. Bei der Umsetzung des Ansatzes der Marktanalyse diagnostizieren die Forscher drei Problemkreise, die sich allesamt auf die Schwierigkeiten bei der Datenerhebung beziehen. Im Gegensatz zur Vorläuferstudie ("eine Produktgruppe, welche gegenüber anderen Erzeugnissen leicht abgrenzbar ist", Balthasar/Müller 1991: 5) machen die Evaluatoren zum einen die "Schwierigkeiten der Marktabgrenzung" (SB: 159) fest, weshalb viele verschiedene Statistiken als Datenquellen nicht genutzt werden konnten. Zweitens habe den Evaluatoren bei der Anwendung der Marktanalyse der Umstand Schwierigkeiten bereitet, dass "viele Schweizer Hersteller zumindest Komponenten, wenn nicht gleich die ganzen Anlagen im Ausland fertigen" (SB: 159). Dies führt dazu, dass die Aussenhandelsstatistik keine verlässliche Datenquelle darstellte. "Drittens waren von der Branche nur spärliche Informationen zu haben." (SB: 159) Das Zusammenwirken dieser drei Aspekte führte dazu, dass eine quantitative Marktanalyse im herkömmlichen Sinn nicht zu realisieren war. Damit wurde die Marktanalyse - wie die Autoren aufrichtigerweise feststellen (SB: 160) - im Endeffekt eher zu einer Fallstudie. Die Evaluatoren folgern: "Marktanalysen, das zeigte sich deutlich, sind vor allem für die Untersuchung von Marktsegmenten geeignet, welche über eine längere Zeit statistisch differenziert dokumentiert sind." (SB: 160) Diese Folgerung ist sicherlich aufgrund der gemachten Erfahrungen gerechtfertigt.

Die Bewertung des Einsatzes *statistischer Verfahren* (SB: 160-162) durch die Evaluatoren lassen sich mit den folgenden zwei Kernaussagen darstellen:

"Es hat sich gezeigt, dass ein statistisches Verfahren bei einer relativ spezifischen Fragestellung auch in der Evaluationsforschung zu aufschlussreichen Ergebnissen führen kann." (SB: 161)

"Quantitative Verfahren tendieren dazu, die hochgradigen Interdependenzen von Variablenzusammenhängen und die Mannigfaltigkeit der Bedingungsfaktoren von Implementationsprozessen zu vernachlässigen." (SB: 162)

Diese Aussagen haben zwei Schwächen. Einerseits besteht ein Abgrenzungsproblem zur Marktanalyse (siehe oben) und zur Zeitreihenuntersuchung (siehe unten), die ja üblicherweise auch mit 'quantitativen Verfahren' operieren. Zweitens erscheint es etwas gewagt, aufgrund des einmaligen Einsatzes eines spezifischen Verfahrens (multiple Regression) für die Gesamtheit aller quantitativen Verfahren Schlüsse zu ziehen. So existieren durchaus quantitative Verfahren, die sowohl die 'Interdependenzen von Variablenzusammenhängen' oder auch die Prozesshaftigkeit der Politik besser abzubilden vermögen, als dies im Rahmen der klassischen Regressionsanalyse möglich ist. Weiter ist zu bemerken, dass die Evaluatoren auch die Möglichkeiten der Regressionsanalyse, auf die problembeladenen Randbedingungen der Evaluationsforschung einzugehen, nicht voll ausgeschöpft haben (siehe die Hinweise in Kapitel 3.4.8.).

Die Bemerkungen der Evaluatoren zur *Zeitreihenuntersuchung* (SB: 162-164) beziehen sich einerseits auf den Einsatz statistischer Zeitreihenverfahren und andererseits auf die in der Evaluation gewählte, pragmatisch angeleitete Vorgehensweise mittels Interpretationen graphischer Darstellungen. Die Evaluatoren argumentieren, dass die Einsatzmöglichkeiten quantitativer Zeitreihenanalysetechniken durch zwei Faktoren stark reduziert, wenn nicht gar verunmöglicht werden. Den einen Faktor bilden die hohen Ansprüche an das zur Verfügung stehende Datenmaterial, den zweiten lokalisieren die Evaluatoren in der Schwierigkeit dieser Verfahren, mit inkrementell angelegten Programmentwicklungen umzugehen (SB: 163). Da die Bandbreite statistischer Zeitreihenverfahren sehr weit ist, treffen beide dieser Bemerkungen sicherlich nicht gleichzeitig für alle Verfahren zu. Einerseits bestehen Verfahren, welche auch mit einer relativ schmalen Datenbasis verwendet werden können (wie beispielsweise einfache 'pretest-/posttest-designs' anhand von Mittelwert- und Varianzvergleichen). Andererseits existieren auch Verfahren, die hochkomplexe Interventionsverläufe nachzuzeichnen vermögen (multivariate ARIMA-Techniken, Zeitreihenregression). Stellt sich aber das Problem, dass beide Einschränkungen gleichzeitig wirksam sind, sind quantitative Zeitreihenanalysen, die über eine graphische Analyse hinausgehen, kaum implementierbar. Die Evaluatoren stellen fest, "dass sich traditionelle statistische Zeitreihenanalysen in der Evaluationsforschung nur in wenigen Fällen nutzbringend einsetzen lassen" (SB: 164). Diese Folgerung ist aufgrund der Nicht-Anwendbarkeit in einem konkreten Fall sicherlich nicht gerechtfertigt. Wie schon in anderem Zusammenhang bemerkt worden ist, führen die Autoren an, statistische Verfahren (sic!) - wie "gleitende Mittelwerte" (SB: 164) - für die Analyse verwendet zu haben, die wir im Schlussbericht jedoch vergeblich suchten. Aufgrund der schmalen empirischen Basis und wegen den Mängeln in der Verwendung

quantitativer Analysetechniken halten wir die hier vorgebrachten Folgerungen der Autoren für nicht gerechtfertigt (vgl. auch Kapitel 3.4.11.).

Die *weitergehenden Überlegungen* der Autoren beschäftigen sich schwerpunktmässig mit Fragen im Zusammenhang mit der Triangulation (SB: 164-171). Sie entwickeln dabei einen Ansatz, welchen sie als '*konzeptionelle Triangulation*' bezeichnen. Bei dieser Vorgehensweise wird ein Gegenstand mittels verschiedener Analysekonzepte oder Zugangsweisen untersucht. Unter dem Begriff der Analysekonzepte verstehen die Forscher zum Beispiel die von ihnen verwendeten Zugänge, wie Fallstudie, Marktanalyse, statistische Verfahren oder Zeitreihenuntersuchung. Diese Vorgehensweise ist indessen nicht dermassen neu, wie sie von den Evaluatoren postuliert wird (siehe SB: 169). Analoge Konzepte sind in der Literatur unter anderer Namensgebung - Denzin etwa verwendet die Bezeichnung '*between-method triangulation*' (Denzin 1989: 244) - schon seit längerer Zeit weit verbreitet (siehe etwa Trend 1978, Jick 1979, Cook 1985, Fielding/Fielding 1986, Smith/Kleine 1986, Denzin 1989: 234-247, Patry 1989, Bryman 1990: 131-134, Patton 1990: 464-472, Hakim 1992: 144-145). Die Evaluatoren folgen in der Evaluation jedoch gar nicht dieser Konzeption. Diese würde ja bedingen, dass die verschiedenen Analysekonzepte auf *denselben* Gegenstand angewendet werden. Die Autoren setzen jedoch die verschiedenen Instrumente zur Analyse *verschiedener* Gegenstände ein, also der Grenzwertsetzung, der Typenprüfung und der Feuerungskontrolle. Darauf weisen die Autoren auch deutlich hin (SB: 167, vgl. Darstellung 8.1., SB: 165). Trotzdem hatten die Autoren - da die drei Massnahmen doch relativ eng zusammenhängen - die Möglichkeit, teilweise von der 'Triangulation' zu profitieren.

c) *Generelle Anmerkungen*

Die Autoren tragen die Folgerungen äusserst klar vor und begründen diese aufgrund ihrer empirischen Arbeit. Punktuell gehen sie jedoch mit ihren Folgerungen über den empirisch untersuchten Gegenstand hinaus. Diese weitergehenden Interpretationen sind sicherlich nicht per se abzulehnen, jedoch wäre es redlicher, dies in Form von Thesen oder Vermutungen - und nicht als gesicherte Erkenntnisse - zu präsentieren. Unter diesem Vorbehalt kann das Kriterium als erfüllt beurteilt werden.

3.4.11. Objektive Berichterstattung

Die Objektivität der Berichterstattung zu bewerten ist kein einfaches Unterfangen. Um dies zu leisten, wäre im Prinzip die Durchführung von längeren Interviews mit einer Vielzahl Beteiligter und Betroffener unabdingbar. Aufgrund kurzer telefonischer Kontakte mit Personen aus dem Forschungsfeld (vgl. Kapitel 5.2.) haben wir den Eindruck erhalten, dass die Berichterstattung im Grossen und Ganzen sicherlich nicht verzerrt ist. An zwei Stellen im Bericht hatten wir den

Eindruck, dass die Evaluatoren etwas befangen waren. Einmal betrifft dies das Argument der Evaluatoren, dass die Qualität der Feuerungskontrolle nicht unwesentlich von den Kompetenzen der Kontrolleure abhängt. Zum Beleg folgendes Zitat: "Ein illustratives Beispiel dafür ist der Umgang mit der Variablen 'Qualität der Kontrolleure'. Aufgrund der Voruntersuchung war uns die Bedeutung dieses Faktors klar. Die ersten Auswertungen haben diese Vermutung bestätigt. Allerdings hätten wir uns deutlichere Ergebnisse gewünscht." (SB: 162)

Der zweite Aspekt betrifft die Methodenkritik im Bereich der quantitativen Verfahren. Die dargelegte Argumentation ist unseres Erachtens in diesem Bereich nicht schlüssig. Im Kapitel zur Zeitreihenuntersuchung beispielsweise heben die Autoren die hohen Datenerfordernisse hervor, die für 'statistische Zeitreihenanalysen' zu erfüllen sind. Eine Definition, was unter 'statistischen Zeitreihenanalysen' zu verstehen ist, fehlt. Es trifft sicherlich zu, dass Verfahren existieren, die hohe Anforderungen an das Datenmaterial stellen. Dies gilt jedoch nur für einen Teil der Verfahren - so lassen sich graphische Analyseverfahren auch dann einsetzen, wenn nur wenige Datenpunkte zur Verfügung stehen. Gegenüber der 'statistischen Zeitreihenanalysen' wird ein weiterer Vorbehalt angefügt: "Dazu kommen aber grundsätzliche Bedenken gegenüber der Anwendung solcher Verfahren in der Evaluationsforschung. Zeitreihenanalysen gehen nämlich von Vorstellungen über Implementationsprozesse aus, wie sie kaum der Praxis entsprechen. Die Einführung politischer Massnahmen geschieht nämlich in der Regel nicht als Intervention, sondern als Prozess." (SB: 163) Diesen Ausführungen können wir nicht folgen, da es nicht zutrifft, dass Zeitreihenanalysen nicht in der Lage sein sollen, dynamische Prozesse zu erfassen. Unseres Erachtens sind gerade Analysen in der Zeitdimension dafür geeignet, Prozesse im Zeitablauf zu untersuchen.

Ebensowenig erscheint die nachfolgende Argumentation stichhaltig: "Für die Analyse inkrementell angelegter Programmentwicklungen stehen auch bei genügender Datenbasis - wenn überhaupt - nur sehr aufwendige statistische Verfahren zur Verfügung. Dadurch treten Interpretationstechniken in den Vordergrund, welche nicht auf statistischen Berechnungen, sondern auf intelligenten Interpretationen des Datenmaterials beruhen." (SB: 163) Und weiter heisst es im Schlussbericht: "Für die Analyse der Beobachtungen zogen wir einfache statistische Masszahlen, wie Mittelwerte und Standardabweichungen bei." (SB: 164) Wir sind der Ansicht, dass auch 'statistische Berechnungen' 'intelligenter Interpretationen' bedürfen und sehen zudem in keiner Weise ein, weshalb 'einfache statistische Masszahlen' keine 'statistischen Berechnungen' darstellen sollen. Aufgrund der Vielfalt der bisher entwickelten Verfahren zur Analyse von Zeitreihen erscheint uns die Behauptung gewagt, dass "- wenn überhaupt - nur sehr aufwendige statistische Verfahren" (SB: 163) bestehen sollen. Unklar bleibt, auf welche Weise die Evaluatoren zu diesen Schlüssen kommen. Diese sind weder durch die durchgeführte Analyse noch auf angeführte Literatur abgestützt. Das Kapitel enthält gerade einen Literaturnachweis (Widmer 1991;

für einen ersten Überblick, vgl. die darin angegebene Literatur und Menard 1991, Chatfield 1989). In diesem Bereich scheint eine Verzerrung durch persönliche Gefühle oder Vorlieben gegeben zu sein.

Abgesehen von diesen punktuellen Einseitigkeiten ist der Bericht aber - soweit wir dies beurteilen können - objektiv abgefasst worden. Die Vorgehensweise, also der Methodenmix, hat sicherlich im positiven Sinne dazu beigetragen.

4. Gesamteindruck

Die Evaluationsstudie 'Umweltpolitik und technische Entwicklung' überzeugt durch ihre Konzeption. Der Einsatz verschiedener Ansätze zur Evaluation verschiedener Teilprogramme und des Gesamtprogramms ist gelungen. Durch eine relativ eng abgegrenzte Fragestellung und ein vergleichsweise grosses Evaluationsvolumen wird die Evaluation realisierbar, obwohl verschiedene Massnahmen auf unterschiedlichen politischen Ebenen mit einer breiten Methodenkombination untersucht werden. Der Evaluationsbericht verfügt über eine klare Grob- und Feinstrukturierung, ist sprachlich sorgfältig aufgearbeitet und präsentiert gut. Er ist auf ein breiteres Publikum ausgerichtet und verzichtet aus diesem Grund auf eine detaillierte methodische Berichterstattung. Für ein spezifisch methodisch orientiertes Publikum verliert der Bericht dadurch etwas an Wert, da die Nachvollziehbarkeit nicht immer gewährleistet ist. Die grössten Schwächen der Evaluation liegen bei der Präzision der Methodenumsetzung. Diesen Befund relativierend muss aber hinzugefügt werden, dass diese Evaluation einerseits durch ihre methodische Vielfalt und andererseits durch die weitergehenden Verfahren in der Datenauswertung das übliche Mass weit übertrifft. Dies ist natürlich auch mit gewissen Risiken verbunden.

Insgesamt betrachtet kommen wir zur Einschätzung, dass diese innovative Studie konzeptionell sowie in der allgemeinen Berichterstattung sehr positiv zu bewerten ist, auch wenn in methodischen Details und in der methodischen Berichterstattung einige kleinere Mängel festgestellt wurden.

5. Anhang

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 5.2.). Zur Erarbeitung dieser Fallstudie sind folgende schriftlichen Unterlagen verwendet worden:

- Forschungsskizze (1/1990)
- Protokoll einer Besprechung vom 17/4/1990
- Schreiben der Programmleitung (Aufforderung Gesuchsauserbeitung) vom 8/5/1990
- Forschungsgesuch vom 12/6/1990 (FG)
- Bewilligungsschreiben Abt. IV des SNF vom 23/8/1990
- Schreiben Revisorat NFP (Verschiebung Projektbeginn), vom 5/11/1990
- Schreiben Programmleitung vom 29/4/1992
- Aktennotiz zu einer Besprechung vom 2/6/1992
- Konzept Teilstudie Hausfeuerungsanlagen vom 2/7/92
- Memorandum "Technologische Fallstudien" vom 21/9/1992
- Entwurf Zwischenbericht vom 22/9/1992
- Konzeption Buch, Entwurf vom 23/9/1992
- Konzeption Buch, Entwurf vom 28/9/1992
- Wissenschaftlicher Zwischenbericht (ohne Datum)
- Finanzieller Zwischenbericht vom 31/10/1992
- Schreiben Programmleitung vom 4/11/1992
- Konzeption Buch, Entwurf vom 11/11/1992
- Referat Wirkungserfassung durch/unabhängig von den Akteuren, Tagung NFP vom 14/1/1993
- Konzeption Buch, Entwurf vom 21/1/1993
- Vortrag "Triangulation" am Forschungszentrum für Schweizerische Politik vom 15/2/1993
- Entwurf des Schlussberichts vom 5/7/1993 (ESB) mit Begleitschreiben
- Aktennotiz Programmleitung zuhanden SNF, Abt. 4 vom 17/8/1993
- Schreiben an SNF Abt. NFP mit Liste Umsetzung und Publikationen vom 3/11/1993
- Finanzieller Schlussbericht vom 18/2/1994
- Schlussbericht in Buchform (Verlag Helbing & Lichtenhahn) (SB)
- Publikationsprospekt Verlag Helbing & Lichtenhahn
- Kurzbericht (NFP27-Broschüre) "Umweltvorschriften und technische Entwicklung" (KB)

5.2. Liste der kontaktierten Personen

- Andreas Balthasar, Interface, Luzern
- Werner Bosshardt, Leiter Prüfstand, Oertli Wärmetechnik, Dübendorf
- Lucius Dürr, Direktor, Verband Schweizerischer Heizungs- und Lüftungsfirmen (VSHL)
- Rudolf Eggi, Umweltschutzbeauftragter, Stadt Zürich
- Ulrich Jansen, Sektion Feuerungen und Energie, Abteilung Luftreinhaltung, Hauptabteilung Immissionsschutz, BUWAL
- Carlo Knöpfel, Basel
- Herbert Limacher, Abteilung Lufthygiene (Emissionen), Amt für technische Anlagen und Lufthygiene ATAL, Kanton Zürich
- Herr Müller (Karl Gut), Amt für Umweltschutz, Kanton Luzern
- K. Rüegg, Marketing, Oertli Wärmetechnik, Dübendorf
- R. Rüfenacht, Abteilung Umweltschutz, KIGA, Kanton Bern
- Werner Wintsch, Immissionsschutz, Gesundheitsinspektorat der Stadt Zürich

5.3. Literaturverzeichnis

- Änderung der Luftreinhalte-Verordnung, Ergebnis des Vernehmlassungsverfahrens. September 1991. Bern: BUWAL.
- ARBEITSGRUPPE GESETZESEVALUATION (AGEVAL) (1991): "Die Wirkungen staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge. Schlussbericht an das Eidgenössische Justiz- und Polizeidepartement. Oktober 1991. Bern: EJPD/EDMZ.
- BALTHASAR, Andreas (1988): "Die gewerkschaftliche Organisation der Typographen, Metall-, Bau- und Holzarbeiter und die Geschichte der Textilarbeiterbewegung in der Schweiz." In: Balthasar, Andreas/Gruner, Erich/Hirter, Hans: Arbeiterschaft und Wirtschaft in der Schweiz 1880-1914. Band II/1. Zürich: Chronos: 255-478.
- BALTHASAR, Andreas (1992): Konzept für die Evaluation von Energie 2000. 19. Mai 1992. Luzern: Interface. (mimeo)
- BALTHASAR, Andreas (1993): "Die Wirksamkeit der Hausfeuerungskontrolle vor dem Hintergrund von Vollzugsdifferenzen. Ein quantitativer Zugang." In: Schweizerisches Jahrbuch für Politische Wissenschaft, Band 33: Vollzugsprobleme: 245-257.
- BALTHASAR, Andreas (Hrsg.) (1993): Einführung in die Evaluationsforschung für Politikwissenschaftlerinnen und Politikwissenschaftler. Mit Beispielen aus der Energiepolitik. Unterlagen des Forschungszentrums für Schweizerische Politik der Universität Bern. Bern: FSP.
- BALTHASAR, Andreas/KNÖPFEL, Carlo (1993): "Umweltvorschriften und technische Innovation." (Interview) In: Bulletin des NFP 27, Nr. 4, Juni 1993: 3-4.
- BALTHASAR, Andreas/KNÖPFEL, Carlo (1994): Umweltpolitik und technische Entwicklung. Eine politikwissenschaftliche Evaluation am Beispiel der Heizungen. Basel: Helbing & Lichtenhahn. (SB)
- BALTHASAR, Andreas/MÜLLER, Ueli (1991): Wirtschaftliche Konsequenzen einer europäischen Harmonisierung im Umweltbereich. Zürich: Rüegger.
- BECK, Ulrich (1986): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt: Suhrkamp. Bericht Luftreinhalte-Konzept, vom 10. September 1986. BBl 1986/III: 269-369.
- BIETRY, J. (1981): "Die Ölfeuerungskontrolle." In: Bundesamt für Umweltschutz (BUS) (Hrsg.): Bundesamt für Umweltschutz 1971-1981. Artikelfolge zum zehnjährigen Bestehen. Juni 1981. Bern: EDMZ.
- Botschaft zu einer Änderung des Bundesgesetzes über den Umweltschutz (USG), vom 7. Juni 1993. BBl 1993/II: 1445-1578.
- BRYMAN, Alan (1990): Quantity and quality in social research. London: Unwin Hyman.
- BUNDESAMT FÜR UMWELTSCHUTZ (Hrsg.) (1986): Mitteilungen zur Luftreinhalte-Verordnung (LRV), Nr. 1, August 1986. Verhältnis der bisherigen Richtlinien und Empfehlungen zur neuen LRV. (mimeo)
- Bundesgesetz über den Umweltschutz (Umweltschutzgesetz USG), vom 7. Oktober 1983. AS 1984: 1122 / SR 814.01.
- CAMPBELL, Donald T./FISKE, Donald W. (1959): "Convergent and discriminant validation by the multitrait-multimethod matrix." In: Psychological Bulletin, Vol. 56, No. 2 (March 1959): 81-105.
- CHATFIELD, C. (1989): The analysis of time series. An introduction. 4th edition. London: Chapman and Hall.
- COOK, Thomas D. (1985): "Postpositivist critical multiplism." In: Shotland, R. Lance/Mark, Melvin M. (eds.): Social science and social policy. Beverly Hills: Sage: 21-62.
- DENZIN, Norman Kent (1970): The research act. A theoretical introduction to sociological methods. Chicago: Aldine.

- DENZIN, Norman Kent (1978a): The research act. A theoretical introduction to sociological methods. 2nd edition. New York: McGraw-Hill.
- DENZIN, Norman Kent (1989): The research act. A theoretical introduction to sociological methods. 3rd edition. Englewood Cliffs: Prentice Hall.
- DENZIN, Norman Kent (ed.) (1978b): Sociological methods: A sourcebook. New York: McGraw-Hill.
- ECKSTEIN, Harry (1966): Division and cohesion in democracy. A study of Norway. Princeton: Princeton University Press.
- ECKSTEIN, Harry (1975): "Case study and theory in political science." In: Greenstein, Fred I./POLSBY, Nelson W. (eds.): Handbook of political science, Vol. 7: Strategies of inquiry. Reading: Addison-Wesley: 79-137. Reprint in: Eckstein 1992: 117-176.
- ECKSTEIN, Harry (1992): Regarding politics. Essays on political theory, stability, and change. Berkeley: University of California Press.
- ECKSTEIN, Harry/GURR, Ted Robert (1975): Patterns of authority: A structural basis for political inquiry. New York: John Wiley & Sons.
- EGGLI, Ruedi/HESS, Walter (1985): "Luftverschmutzung und Luftreinhaltung in Zürich." In: Weidner, Helmut/Knoepfel, Peter (Hrsg.): Luftreinhaltungspolitik in städtischen Ballungsräumen. Internationale Erfahrungen. Frankfurt: Campus: 149-162.
- EGGLI, Ruedi/HESS, Walter/KUSTER, Robert/SUTER-WEIDER, Peter (1979): Ölfeuerungskontrolle in der Stadt Zürich. Zusammengefasste Resultate der Heizperioden 1976/77 und 1977/78. Schriftenreihe der SVG, Nr. 71. Zürich: SVG.
- EIDGENÖSSISCHE KOMMISSION FÜR LUFTHYGIENE (Hrsg.) (1973): Probleme der Lufthygiene. Unterlagen zur Umweltschutzgesetzgebung. Oktober 1973. Bern/Zürich: EKLH.
- EIDGENÖSSISCHES DEPARTEMENT DES INNERN (EDI) (1984): Luftreinhalte-Verordnung (LRV) - Entwurf. Mai 1984. Bern: EDI.
- EIDGENÖSSISCHES DEPARTEMENT DES INNERN (EDI) (1985): Ergebnisse des Vernehmlassungsverfahrens über den Entwurf zur Luftreinhalte-Verordnung (LRV). September 1985. Bern: EDI.
- EIDGENÖSSISCHES DEPARTEMENT DES INNERN (EDI) (1990): Entwurf zur Änderung der Luftreinhalte-Verordnung (LRV). März 1990. Bern: EDI.
- Erläuterungen zur Luftreinhalte-Verordnung. Texte zum Umweltschutzgesetz. Juni 1989. Bern: EDMZ.
- ESTERMANN, Stefan/FELDMANN, Sven (1990): Les incidences de la politique de protection de l'air sur l'innovation technologique de l'industrie domestique et sur la création de nouveaux marchés. Support de cours et dossier de travail, No. 7, juin 1990. Lausanne: IDHEAP. (mimeo)
- EUW, René von (1990): "Blaue Flamme, saubere Luft." In: Brückenbauer, 31. Januar 1990: 32-34.
- FIELDING, Nigel G./FIELDING, Jane L. (1986): Linking data. Beverly Hills: Sage.
- FITZE, Urs (1994): "Staatlicher Umweltschutz fördert Technikfortschritt." TA vom 20. April 1994: 84.
- FREIBURGHAUS, Dieter (1988): Methoden der Evaluation von Verwaltungshandeln aus der Sicht der Politik- und Verwaltungswissenschaft. Vortrag vor der Arbeitsgruppe Verwaltungskontrolle der Geschäftsprüfungskommission am 7. September 1988. Cahiers de l'IDHEAP, No. 46. Lausanne: IDHEAP.
- FREIBURGHAUS, Dieter (1989): Le développement des moyens de l'action étatique. Cahiers de l'IDHEAP, No. 59. Lausanne: IDHEAP.
- FREIBURGHAUS, Dieter/BALTHASAR, Andreas/ZIMMERMANN, Willi/KNÖPFEL, Carlo (1991): Technik-Standort Schweiz. Von der Forschungs- zur Technologiepolitik. Bern: Haupt.
- FREIBURGHAUS, Dieter/ZIMMERMANN, Willy/BALTHASAR, Andreas (1990): Evaluation der Förderung praxisorientierter Forschung, Studie Nummer 12 des Bundesamtes für Konjunkturfragen (BfK). Bern: BfK.

- FREY, Bruno S. (1977): *Moderne Politische Ökonomie. Die Beziehungen zwischen Wirtschaft und Politik*. München: Piper.
- FREY, Bruno S. (1981): *Theorie demokratischer Wirtschaftspolitik*. München: Vahlen.
- GESUNDHEITSDIREKTION DER STADT BERN, AMT FÜR UMWELTSCHUTZ UND LEBENSMITTELKONTROLLE (Hrsg) (1980): *Jahresbericht 1980*. Bern: Amt für Umweltschutz und Lebensmittelkontrolle.
- GUBA, Egon G./LINCOLN, Yvonna S. (1989): *Fourth generation evaluation*. Newbury Park: Sage.
- HAKIM, Catherine (1992): *Research design. Strategies and choices in the design of social research*. London: Routledge.
- HESS, Walter (1980): *Erste Resultate von Wirkungsgradermittlungen an den Ölfeuerungsanlagen in der Stadt Zürich*. Schriftenreihe der SVG, Nr. 77. Zürich: SVG.
- HESS, Walter (1982): *Erfolgsstatistik der Ölfeuerungskontrolle aus der Sicht des Umweltschutzes und der Energieeinsparung*. Schriftenreihe der SVG, Nr. 84. Zürich: SVG.
- HESS, Walter (1985): "Geschichtlicher Überblick und Konsequenzen bezüglich der Ölfeuerungskontrolle in der Stadt Zürich." In: Weidner, Helmut/Knoepfel, Peter (Hrsg.): *Luftreinhaltepolitik in städtischen Ballungsräumen. Internationale Erfahrungen*. Frankfurt: Campus: 163-181.
- JICK, Todd D. (1979): "Mixing qualitative and quantitative methods: Triangulation in action." In: Maanen, J. van (ed.): *Qualitative methodology*. London: Sage.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981): *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994): *The program evaluation standards. How to assess evaluations of educational programs*. 2nd edition. Thousand Oaks: Sage.
- KNOEPFEL, Peter/IMHOF, Rita/ZIMMERMANN, Willi (1994): *Massnahmenpläne zur Luftreinhaltung. Wie sich Behörden beim Umweltschutz arrangieren*. Bericht 57 des NFP 'Stadt und Verkehr'. Zürich: NFP 25.
- KNÖPFEL, Carlo (1988): *Der Einfluss der politischen Stabilität auf die internationale Wettbewerbsfähigkeit der Schweiz. Kritische Analyse eines Standortfaktors*. Basler Dissertation. Grösch: Rüegger.
- KNÖPFEL, Carlo (1989): "Nichttarifäre Handelsbarrieren und Integration: Die Schweiz und die europäische Gemeinschaft." In: *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 125, Jahrgang, Heft 3: 327-354.
- Luftreinhalte-Verordnung (LRV), vom 16. Dezember 1985 (Stand am 1. Juli 1993). SR 814.318.142.1. (LRV 92)
- Luftreinhalte-Verordnung (LRV), vom 16. Dezember 1985. AS 1986: 208. (LRV 86)
- MÄDLER, Robert W. (1988): *Sanierung der Feuerungen - ergiebigste Massnahme zur Verbesserung der Stadtluft*. Studie im Auftrag der City Vereinigung Zürich in Zusammenarbeit mit P. Suter, Labor für Energiesysteme an der ETH und Oertli AG. Zürich: City Vereinigung. (mimeo)
- MAYNTZ, Renate (1990): "Entscheidungsprozesse bei der Entwicklung von Umweltstandards." In: *Die Verwaltung*, Band 23, Heft 2: 137-151.
- MENARD, Scott (1991): *Longitudinal research*. Newbury Park: Sage.
- NZZ vom 16./17. April 1994: 24: "Wirtschaftliche Vorteile durch Umweltgesetze."
- NZZ vom 5. Januar 1994: 18: "Überfordert der Staat sich selbst?"
- PATRY, Jean-Luc (1989): "Evaluationsstudien zu Forschungszwecken: Ein Beispiel von 'kritischem Multiplizismus'." In: *Unterrichtswissenschaft*, 17: 359-374.
- PATTON, Michael Quinn (1990): *Qualitative evaluation and research methods*. 2nd edition. Newbury Park: Sage.

- PERROCHET, Jean-François (1989): Le contrôle des chauffages à combustion dans le cadre de la protection de l'air. Comparaison de quatre villes suisses. Avril 1989. Lausanne: IDHEAP. (mimeo)
- POMMEREHNE, Werner W./FREY, Bruno S. (Hrsg.) (1979): Ökonomische Theorie der Politik. Berlin: Springer.
- PRESSESTELLE NFP/SPP: Strenge Grenzwerte führten zu technischen Fortschritten in der Heizungsbranche. Pressemitteilung vom 12. April 1994.
- Richtlinien über die Auswurfbegrenzung bei Haus- und Industriefeuerungen, vom 7. Februar 1972. BBl 1972/I: 1089-1095.
- Richtlinien über die Konstruktion und den Betrieb von Heizkesseln und Zerstäuberbrennern, vom 21. März 1978. BBl 1978/I: 1403-1408.
- Richtlinien über die Typenprüfung von Heizkesseln und Zerstäuberbrennern bezüglich der luft-hygienischen Anforderungen, vom 1. Mai 1984. BBl 1984/II: 1230-45.
- Richtlinien über die Typenprüfung von Heizkesseln und Zerstäuberbrennern, vom 2. Juli 1980. BBl 1980/III: 273-281.
- Richtlinien zur Prüfung der Abgase von Ölf Feuerungen, vom 21. März 1978. BBl 1978/I: 1396-1402.
- SCHUMPETER, Joseph (1912): Theorie der wirtschaftlichen Entwicklung. Leipzig: Duncker & Humblot. (Ausgabe 1988: Düsseldorf: Wirtschaft und Finanzen)
- SCHWEIZERISCHE VEREINIGUNG FÜR GESUNDHEITSTECHNIK (SVG) (Hrsg.) (o.J.): Energie und Umwelt, Heft 1. Grundlagen der Ölf Feuerungskontrolle und Wirkungsgradermittlung an Feuerungsanlagen. Schriftenreihe der SVG, Nr. 69. Zürich: SVG.
- SCHWEIZERISCHE VEREINIGUNG FÜR GESUNDHEITSTECHNIK (SVG) (Hrsg.) (1982): Erfolgstatistik der Ölf Feuerungskontrolle aus der Sicht des Umweltschutzes und der Energieeinsparung. Schriftenreihe der SVG, Nr. 84. Zürich: SVG.
- SCHWEIZERISCHE VEREINIGUNG FÜR GESUNDHEITSTECHNIK (SVG) (Hrsg.) (1983a): Luftreinhaltung in der Gemeinde. Seminar über den Vollzug von Luftreinhaltmassnahmen auf Gemeindeebene und die damit zusammenhängenden Probleme und Schwierigkeiten. Schriftenreihe der SVG, Nr. 85. Zürich: SVG.
- SCHWEIZERISCHE VEREINIGUNG FÜR GESUNDHEITSTECHNIK (SVG) (Hrsg.) (1983b): Die Ölf Feuerungskontrolle in der Schweiz. Schriftenreihe der SVG, Nr. 87. Zürich: SVG.
- SCHWEIZERISCHER BUNDESRAT (1992): Technologiepolitik des Bundes. Bern: BfK.
- SCHWEIZERISCHER NATIONALFONDS (SNF) (Hrsg.) (1989): Nationales Forschungsprogramm 27. Wirksamkeit staatlicher Massnahmen. Ausführungsplan. Oktober 1989. Bern: SNF.
- SMITH, L.M./KLEINE, P.G. (1986): "Qualitative research and evaluation: Triangulation and multi-methods reconsidered." In: Williams, David D. (ed.): Naturalistic evaluation. New Directions for Program Evaluation, No. 30. San Francisco: Jossey-Bass.
- TREND, M.G. (1978): "On the reconciliation of qualitative and quantitative analyses. A case study." In: Human Organization, 37: 345-354. Reprint in: Cook, Thomas D./Reichhardt, Charles S. (1979): Qualitative and quantitative methods in evaluation research. Beverly Hills: Sage: 68-86.
- Verordnung über Luftreinhalte-Massnahmen bei Feuerungen (LMFV), vom 10. Dezember 1984. AS 1984: 1516.
- WIDMER, Thomas (1991): Evaluation von Massnahmen zur Luftreinhaltspolitik in der Schweiz. Zürich: Rüegger.
- WIDMER, Thomas (1996a): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Bern: Haupt.
- WIDMER, Thomas (1996b): Fallstudien zur Meta-Evaluation, Teil II: 'Unternehmerische Innovationsprozesse' (Studie Nr. 2), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.
- WIDMER, Thomas (1996c): Fallstudien zur Meta-Evaluation, Teil III: 'Wohneigentumsförderung durch den Bund' (Studie Nr. 4), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.

- WIDMER, Thomas (1996d): Fallstudien zur Meta-Evaluation, Teil IV: 'Evaluation der Zürcher Gemeinschaftszentren' (Studie Nr. 5), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.
- WIDMER, Thomas (1996e): Fallstudien zur Meta-Evaluation, Teil V: 'Evaluation Kehrichtsackgebühr' (Studie Nr. 7), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.
- WIDMER, Thomas (1996f): Fallstudien zur Meta-Evaluation, Teil VI: 'Heizkostenabrechnung und erneuerbare Energien' (Studie Nr. 8), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.
- WIDMER, Thomas (1996g): Fallstudien zur Meta-Evaluation, Teil VII: 'Evaluierung der EWR-Informationenkampagne' (Studie Nr. 10), Reihe "Schlussberichte" NFP 27. Bern: NFP 27.
- WÜEST & PARTNER (Hrsg.) (1992): Konzepte zur Förderung des Einsatzes emissionsarmer Feuerungen. Im Auftrag von BUWAL und BEW. Schlussbericht, September 1992. Zürich: Wüest & Partner. (mimeo)
- 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)

5.4. Abkürzungsverzeichnis

a.a.O.	am angegebenen Ort
Abs.	Absatz
Abt.	Abteilung
Amtl.	Amtlich
Anm.	Anmerkung
ARIMA	Autoregressiv-integriertes-moving-average-Modell
Art.	Artikel
AS	Amtliche Sammlung des Bundesrechts
BBl	Bundesblatt
bes.	besonders
BEW	Bundesamt für Energiewirtschaft
BfK	Bundesamt für Konjunkturfragen
BL	Kanton Basel-Landschaft
bspw.	beispielsweise
BUWAL	Bundesamt für Umwelt, Wald und Landschaft
CVP	Christlich-Demokratische Volkspartei
EAM	Eidgenössisches Amt für Messwesen
EDI	Eidgenössisches Departement des Innern
EDMZ	Eidgenössische Druck- und Materialzentrale
EK	Evaluationskontrakt
EMPA	Eidgenössische Materialprüfungs- und Forschungsanstalt
ESB	Entwurf zum Schlussbericht
F&E	Forschung und Entwicklung
FDP	Freisinnig-Demokratische Partei
FG	Forschungsgesuch
FSB	Finanzieller Schlussbericht
FSP	Forschungszentrum für schweizerische Politik an der Universität Bern
FZB	Finanzieller Zwischenbericht
IDHEAP	Institut de hautes études en administration publique
KWF	Kommission für wissenschaftliche Forschung
LGBL	Landesgesetzblatt
LMFV	Verordnung über Luftreinhaltmassnahmen bei Feuerungen (AS 1984: 1516)
LNN	Luzerner Neueste Nachrichten
LRV 86	Luftreinhalte-Verordnung 1986 (AS 1986: 208)
LRV 92	Luftreinhalte-Verordnung 1992 (SR 814.318.142.1)
MW	Megawatt
NFP	Nationales Forschungsprogramm
NFP 27	Nationales Forschungsprogramm 'Wirksamkeit staatlicher Massnahmen'
NO ₂	Stickstoffdioxide
NO _x	Stickoxide
Nr.	Nummer
NZZ	Neue Zürcher Zeitung
OF	Offerte
OLS	Ordinary-least-squares
s.	siehe

SB	Schlussbericht
sFr.	Schweizer Franken
SGU	Schweizerische Gesellschaft für Umweltschutz
SNF	Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung
SPS	Sozialdemokratische Partei der Schweiz
SR	Systematische Sammlung des Bundesrechts
SVG	Schweizerische Vereinigung für Gesundheitstechnik
TA	Tages-Anzeiger
Tab.	Tabelle
USG	Bundesgesetz über den Umweltschutz (Umweltschutzgesetz) (SR 814.01)
vgl.	vergleiche
Vol.	Volume
WeWo	Die Weltwoche
ZB	Zwischenbericht
ZG	Kanton Zug
ZH	Kanton Zürich
Ziff.	Ziffer
zit. n.	zitiert nach

6. Stellungnahme der Evaluatoren

Die Durchführung von Metaevaluationen hat in der Schweiz noch keine Tradition. Wir haben die Arbeit von Thomas Widmer deshalb sehr begrüsst und erhoffen uns davon einen Beitrag zur Qualitätssicherung in einer noch jungen Disziplin. Dass der Autor bei der Metaevaluation unserer Untersuchung "Umweltpolitik und technische Entwicklung" zu einem grundsätzlich positiven Gesamturteil (vgl. Kapitel 4.) gekommen ist, freut uns noch zusätzlich. Obwohl uns die Gewissenhaftigkeit und Kompetenz, mit welcher der Autor an die Arbeit gegangen ist, beeindruckte, hat die Lektüre der Fallstudie bei uns einige Skepsis ausgelöst. Diese hängt einerseits mit der Vorgehensweise des Autors zusammen. Andererseits scheinen uns auch grundsätzliche Vorbehalte gegenüber dem beigezogenen Kriterienraster, bzw. seiner Anwendung angebracht. Im folgenden werden wir die Gelegenheit nutzen, unsere Ueberlegungen zu den beiden genannten Punkten kurz darzulegen.

Bemerkungen zur Vorgehensweise des Autors

Die Lektüre der Fallstudie macht deutlich, dass sich der Autor im allgemeinen mit Unvoreingenommenheit unserer Arbeit genähert hat. Er hat sich, ohne den beachtlichen Arbeitsaufwand zu scheuen, in die Einzelheiten unseres Themas eingearbeitet und sachkundig gemacht. Dadurch weist er sich uns gegenüber als kompetenter Gutachter aus. Auf dieser Grundlage ist es ihm gelungen, Schwachstellen unserer Arbeit zu identifizieren und präzise zu benennen. Wir denken insbesondere an Hinweise betreffend die Validität und die Reliabilität einzelner Variablen der Regressionsanalyse (Kapitel 3.4.5. und 3.4.6.). Seine Arbeit hat uns auch vor Augen geführt, dass eine ausführlichere Darlegung der Vorgehensweise und unserer methodischen Ueberlegungen zweckmässig gewesen wäre. Bei der Abfassung unseres Berichts haben wir in weiten Teilen darauf verzichtet, weil wir die anvisierte Lesergruppe nicht mit den Bürden sozialwissenschaftlicher Forschungspraxis belasten wollten. Dass dies aus wissenschaftlicher Sicht unbefriedigend ist, war uns bewusst. Für die interessierten Wissenschaftler haben wir darum das entsprechende Material in unserem Archiv bereit gehalten. Es stand dem Metaevaluator vollumfänglich zur Verfügung.

Hier setzt die erste Kritik an, welche wir an der Vorgehensweise des Autors anbringen möchten. Trotz verschiedener textlicher Hinweise (z.B. Kapitel 4.3 und Anhang 5. zur Fallstudie) bleibt unklar, welchen Stellenwert er internen Dokumenten und Gesprächsinhalten bei der Berichtsabfassung beigemessen hat. Wir veranschaulichen dies an zwei Beispielen:

- In Kapitel 3.4.3 "Beschreibung von Zielen und Vorgehen" weist der Autor darauf hin, dass in unserem Schlussbericht zwar die grundlegenden Ziele und Vorgehensweisen ausreichend genau beschrieben sind, dass hingegen keine präzise Darstellung des Evaluationsprozesses er-

folgt. Auf den folgenden Seiten stellt er über 50 konkrete Fragen dazu. Ein grosser Teil dieser Themen wurde in ausführlichen Gesprächen zwischen dem Metaevaluator und dem Projektleiter thematisiert und - aus unserer Sicht - auch geklärt. Anlässlich eines Besuches in unserem Büro bekam der Autor beispielsweise Einsicht in die Liste der befragten Expertinnen und Experten, in die Gesprächsleitfaden und in die Gesprächsprotokolle. Es bleibt uns verborgen, wieso die Fragen, die wir ausführlich diskutiert und durch interne Unterlagen erhellt haben, nach wie vor als ungeklärt bezeichnet werden (vgl. Punkt e) im genannten Kapitel).

- Das zweite Beispiel stammt aus Kapitel 3.4.11, das sich mit der "objektiven Berichterstattung" befasst. Dort vertritt der Metaevaluator die Ansicht, dass wir gegenüber quantitativen Verfahren befangen wären und unsere Berichterstattung in dieser Sache "durch persönliche Gefühle oder Vorlieben" verzerrt sei. Wir haben in verschiedenen Gesprächen nachdrücklich darauf hingewiesen, dass diese Voreingenommenheit nicht besteht. Andernfalls hätten wir uns in unserer Evaluation nicht intensiv mit verschiedenen quantitativen Verfahren auseinandergesetzt. Unsere Hinweise zu diesem Punkt hat der Autor jedoch nicht in seinen Bericht aufgenommen.

Kommen wir zum zweiten Kritikpunkt an der Vorgehensweise. Wir haben oben deutlich gemacht, dass wir die fachliche Kompetenz des Metaevaluators als hoch einschätzen. Dennoch glauben wir, dass die Metaevaluation an Glaubwürdigkeit gewinnen würde, wenn bei der Beurteilung inhaltlicher Aspekte grössere Zurückhaltung geübt würde. Grundsätzlich kann es unserer Ansicht nach nicht Aufgabe der Metaevaluation sein, sich zu ungenutzten Datenpotentialen zu äussern. Dies auch dann nicht, wenn kurzfristig eigene Auskünfte eingeholt worden sind. Auch diese Problematik möchten wir an zwei Beispielen aus der Fallstudie, die unsere Arbeit betrifft, verdeutlichen:

- An verschiedenen Stellen vertritt der Autor die Ansicht, dass wir im Rahmen der Marktanalyse (Kapitel 4. unseres Schlussberichtes) zugängliche Informationen nur teilweise genutzt hätten (Kapitel 3.1.3. und 3.4.9.). Es wird angetönt, dass man - mit wenig Aufwand - mehr hätte leisten können. Dies wirkt sich bei verschiedenen Kriterien negativ aus (z.B. "Informationsauswahl" und "Analyse qualitativer Informationen"). Gerade in diesem Untersuchungsteil hat sich aber die sorgfältige Prüfung einzelner Datensätze und Informationen als äusserst wichtig erwiesen. Gemäss der üblichen sozialwissenschaftlichen Praxis haben wir auf den Beizug fragwürdiger Daten systematisch verzichtet.
- Das zweite Beispiel betrifft die Diskussion der Operationalisierung von Variablen für die Regressionsanalyse in Kapitel 3.4.5. Verschiedentlich weist der Metaevaluator auf kleinere und grössere Unzulänglichkeiten bei der Auswahl der Variablen hin. Auch wenn es für uns unbefriedigend ist, so ist es doch sein Recht, Schwächen zu benennen, ohne Lösungsvorschläge

präsentieren zu müssen. Dass diese nämlich schwieriger zu finden sind, als es auf den ersten Blick erscheint, wird dort deutlich, wo der Metaevaluator selbst einen Vorschlag unterbreitet. Im Abschnitt cd) des erwähnten Kapitels schlägt er vor, die Amtsdauer von Kontrolleuren intervall-skaliert zu erfassen, da damit die Freiheitsgrade der Regressionsanalyse erhöht werden könnten. Auf ein solches methodeninduziertes Vorgehen haben wir bewusst verzichtet, weil wir damit die Erinnerungsfähigkeit und die Auskunftsbereitschaft unserer Informanten überstrapaziert hätten. Wir haben es willentlich unterlassen, Daten zu erzeugen, welche zwar methodengerecht gewesen wären, die aber eine ungerechtfertigte Genauigkeit supponiert hätten.

Bemerkungen zur Anwendung des Kriterienkatalogs

Der Kriterienkatalog des "Joint Committee" hat den Metaevaluator zu einer schematischen Abarbeitung jedes einzelnen der rund 30 Punkte veranlasst. Gesamthaft gesehen hat dieses schematische Vorgehen aus unserer Sicht den Vorteil gehabt, dass alle Untersuchungen mit der gleichen, umfassend erscheinenden Gründlichkeit angegangen werden. Zahlreiche wichtige Fragen werden auf diese Weise thematisiert. Dies betrachten wir als wichtige Stärke des Ansatzes. Gleichwohl halten wir die Art, wie der Katalog umgesetzt wurde, für die Beurteilung von Evaluationen für nicht besonders ergiebig. Wie kommen wir zu diesem Urteil?

Der Katalog führt zu einer buchhalterischen Abarbeitung der einzelnen Beurteilungskriterien. Punkt für Punkt werden die Schwachstellen der untersuchten Fälle mit grosser Ausführlichkeit dargelegt. Zum Teil werden sogar sprachliche Unsauberkeiten (z.B. in Kapitel 3.1.3.) oder gar Tippfehler aufgelistet (z.B. in Kapitel 3.4.1). Jede Schwachstelle wird notiert, um am Schluss zu einer Bewertung zwischen "erfüllt" und "nicht erfüllt" zu gelangen. Dieser Ansatz trägt den Zeit- und Kostenrestriktionen, welche die praktische Arbeit prägen, nur ungenügend Rechnung. Auch sucht er vor allem nach Schwächen und weniger nach den Stärken. Noch wichtiger ist aber der Umstand, dass bei diesem Ansatz die Kriterien, welche die Feinplatzierung auf der Bewertungsachse bestimmen, weitgehend im Dunkeln bleiben. Ob der Erfüllungsgrad bei einem Kriterium als "ganz" oder nur als "teilweise" bezeichnet wird, erscheint als Ergebnis einer weitgehend subjektiven Einschätzung.

Doch damit ist nur ein Teil des Problems, das der gewählte Ansatz mit sich bringt, benannt. Die Hauptschwäche der gewählten kriteriengeprägten Betrachtungsweise besteht darin, dass sie den Blick für integrative Aspekte nicht öffnet. Selbstverständlich müssen methodische und ethische Kriterien bei der Durchführung von Evaluationen eingehalten werden. Die Qualität einer Untersuchung zeigt sich unseres Erachtens aber weniger darin, als in der Art, wie Rahmenbedingungen, praktische Ansprüche und methodische sowie ethische Anforderungen miteinander kombi-

niert werden. Bei der Beurteilung dieser für uns massgeblichen Dimension der Qualität einer Evaluation versagt die vorliegende Art der Anwendung des Kriterienkatalogs. Wie hängen beispielsweise Breite der Fragestellung und praktischer Nutzen zusammen? Wie wirken sich unterschiedliche Methoden auf Kriterien wie "Lesbarkeit", "Akzeptanz" oder "Kosten" aus? Wie lösen Evaluatoren methodische Probleme angesichts beschränkter finanzieller Mittel? Wie gelingt es ihnen, in einem von Insidern geprägten Feld unabhängige Beurteilungen vorzunehmen? Wie und mit welchem Nutzen werden Ergebnisse qualitativer und quantitativer Untersuchungsteile verknüpft? Von den Antworten auf solche und ähnliche Fragen sind unseres Erachtens die wesentlichen Beiträge der Metaevaluation auf dem Weg zu einer qualitativ hochstehenden und praktisch nützlichen Evaluationsforschung zu erwarten. Diese Fragen liegen jedoch quer zur angewandten Vorgehensweise. Darum wurden sie in der vorliegenden Metaevaluation leider weder gestellt noch beantwortet.

Luzern, 20. September 1995

A. Balthasar und C. Knöpfel

