

ENDBERICHT

EVALUIERUNG VON FTE-PROGRAMMEN: ZWISCHEN BEST PRACTICE ENTWICKLUNG UND ÖSTERREICHISCHEN STRUKTUREN

*STUDIE IM AUFTRAG DES
BUNDESMINISTERIUMS FÜR WISSENSCHAFT, VERKEHR
UND KUNST*

INSTITUT FÜR TECHNOLOGIE- UND REGIONALPOLITIK –InTeReg:

O.FRITZ, D. STURN

ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG – WIFO:

G. HUTSCHENREITER

RTG.96.016-01

WIEN – GRAZ FEBRUAR 1997

INHALTSVERZEICHNIS

0 EINLEITUNG	2
1 GRUNDLEGENDE THEMEN UND FRAGESTELLUNGEN BEI PROGRAMMEVALUIERUNGEN.....	4
1.1 WOZU BRAUCHT MAN EVALUIERUNGEN ÖFFENTLICHER PROGRAMME?	4
1.2 WER VERFOLGT WELCHE INTERESSEN IN EINER EVALUIERUNG?	5
1.3 WIE GEHT MAN GRUNDSÄTZLICH AN EINE EVALUIERUNG HERAN?	6
1.4 WANN WERDEN EVALUATIONEN DURCHFÜHRT? ZEITLICHE EVALUIERUNGSDIMENSION	9
1.5 WAS SOLL BEWERTET WERDEN? EVALUIERUNGSMETHODEN	11
1.6 EIN LEITFADEN DER PROGRAMMEVALUIERUNG	15
1.7 METHODEN	29
1.7.1 <i>Quantitative versus qualitative Methoden</i>	29
1.7.2 <i>Welche Methoden kommen wo zur Anwendung? Die Verknüpfung von Methoden mit Evaluierungsinhalten</i>	32
2 SPEZIFIKA VON FTE PROGRAMMEN UND VERGLEICH MIT STRUKTURPOLITISCHEN PROGRAMMEN	43
2.1 DIE EVALUIERUNG VON FTE-PROGRAMMEN.....	43
2.1.1 <i>Abschätzung von Programm-Wirkungen</i>	44
2.1.2 <i>Zieldefinition und Zielsetzung von FTE-Programmen</i>	45
2.2 VERGLEICH MIT STRUKTURFONDSEVALUIERUNG	46
2.2.1 <i>Die Regelungen von seiten der Europäischen Union</i>	47
2.2.2 <i>Die Funktionen der Evaluierung</i>	49
2.2.3 <i>Qualität und aktuelle Praxis der Evaluierungen</i>	49
2.2.4 <i>Probleme der aktuellen Praxis</i>	50
2.2.5 <i>Schlußfolgerungen und Empfehlungen</i>	53
3 INTERNATIONALE ERFAHRUNGEN MIT FTE-PROGRAMMEVALUIERUNGEN	54
3.1 DIE EVALUIERUNG DES ALVEY PROGRAMME FOR ADVANCED INFORMATION TECHNOLOGY.....	54
3.1.1 <i>Programmbeschreibung</i>	54
3.1.2 <i>Allgemeine Evaluierungsstruktur</i>	55
3.1.3 <i>Spezielle Aspekte der Evaluierung</i>	56
3.1.4 <i>Beurteilung der Alvey-Evaluierung</i>	58
3.2 EVALUIERUNG INDIREKTER F&E-FÖRDERUNG: DIE STUDIE VON B. HALL (1992).....	59
4 EVALUIERUNG VON TECHNOLOGIEPROGRAMMEN IN ÖSTERREICH.....	62
4.1 EINLEITUNG: EVALUIERUNGEN IM GESAMTSYSTEM DER FTE-FÖRDERUNG.....	62
4.2 EVALUIERUNGEN IM BEREICH DER INVESTITIONSFÖRDERUNG	63
4.3 EVALUIERUNG ÖSTERREICHISCHER TECHNOLOGIEPROGRAMME.....	66
4.4 EVALUIERUNG VON INSTRUMENTEN DER INDIREKTEN F&E-FÖRDERUNG	72
4.5 EVALUIERUNG DER ÖSTERREICHISCHEN TEILNAHME AN INTERNATIONALEN FORSCHUNGS- UND TECHNOLOGIEPROGRAMMEN	76
5 EMPFEHLUNGEN.....	80
Literatur.....	86
Anhang.....	91

0 Einleitung

Ziel der Studie ist es, grundlegende Fragen der Evaluierung technologiepolitischer Programme zu erörtern und auf dieser Grundlage die österreichische Evaluierungspraxis zu analysieren sowie Empfehlungen hinsichtlich ihrer Weiterentwicklung abzuleiten.

Im ersten Teil der Studie werden, auf Grundlage der theoretischen Evaluierungsliteratur, die verschiedenen Dimensionen einer Programmevaluierung (funktionelle, zeitliche, inhaltliche) dargestellt und miteinander verknüpft. Damit werden ein methodischer Rahmen der Programmevaluierung geschaffen und wesentliche evaluierungstheoretische Grundsätze formuliert, die im weiteren Verlauf der Studie als Orientierungshilfe dienen. Die Verknüpfung der einzelnen Dimensionen in Abschnitt 1.6 erfolgt auch mit der Absicht, ein Handbuch der Programmevaluierung hervorzubringen, anhand dessen potentielle Auftraggeber von Evaluierungen die Vorschläge zur Evaluierung konkreter Programme vergleichen können.

In Teil 2 werden Besonderheiten von FTE-Programmen vorgestellt und ein Vergleich mit strukturpolitischen Programmen vorgenommen. Diese Besonderheiten betreffen in der Regel alle Evaluierungsinhalte, also Konzept-, Design- und Prozeßevaluierung, vor allem aber die Wirkungsevaluierung. Wirkungen sind bei FTE-Programmen besonders schwer abzuschätzen, da indirekte Effekte auf verschiedenen Ebenen eine beträchtliche Rolle spielen, welche zu identifizieren und zu messen ein Problem bereitet. Der erste Abschnitt von Teil 2 diskutiert die damit verbundenen Probleme, die in erster Linie aufgrund der Schwierigkeit der Abschätzung sozialer Erträge entstehen. Es zeigt sich, daß eine Verbesserung von Verbindlichkeit und Methodik nicht ausreicht, um ein wirklich handlungsweisendes Zusammenspiel von Programmerstellung, Programmabwicklung und Programmbewertung zu erhalten. Daneben sollte in der FTE-Politik eine Diskussion über adäquate Zielsetzungen von Programmen geführt werden und neue theoretische Konzepte sowie neue Ansätze der Messung vor allem ökonomischer Effekte von Programmen berücksichtigt bzw. entwickelt werden. Der im zweiten Abschnitt von Teil 2 anschließende Vergleich mit der Evaluierungspraxis der gemeinschaftlichen Strukturpolitik soll aufzeigen, welche Vor- und Nachteile aus genauen, strikten und verbindlichen Regelungen sowohl hinsichtlich der Programmerstellung als auch hinsichtlich Durchführung, Bewertung und Begleitung entstehen und wie aufgrund dieser Erfahrungen in der FTE-Politik sinnvolle Weiterentwicklungen aussehen können.

Teil 3 beschreibt zwei internationale Evaluierungen aus dem FTE-Bereich: Die Evaluierung des Alvey-Programmes zur Förderung des Informationstechnologiesektors in Großbritannien ist ein Beispiel einer groß angelegten Programmevaluierung, die aus vielen einzelnen

Evaluierungsstudien besteht. Von diesen wird die Ex-post Evaluierung näher dargestellt. Die Studie von B. Hall aus dem Jahr 1992 beschäftigt sich mit einem Instrument der indirekten F&E-Förderung in den USA, dem "R&D tax credit", und evaluiert die Wirkungen dieses Instruments auf die F&E-Ausgaben auf Grundlage eines quantitativen (ökonometrischen) Ansatzes.

Teil 4 befaßt sich mit dem aktuellen Stand der Evaluierung von FTE-Programmen in Österreich. Einleitend wird ein Überblick über die bisherigen Evaluierungen im Gesamtsystem der österreichischen FTE-Förderung geboten. Daran anschließend werden Erfahrungen aus der Evaluierung im Bereich der österreichischen Investitionsförderung und ihre Relevanz für die Evaluierung von FTE-Programmen erörtert. Die folgenden Abschnitte sind einer detaillierten Darstellung der bisherigen und geplanten Evaluierungen von nationalen Technologieprogrammen und Überlegungen zu einer bislang noch ausstehenden Evaluierung der indirekten F&E-Förderung gewidmet. Abschließend wird auf die bisher durchgeführten Evaluierungen der österreichischen Teilnahme an internationalen FTE-Programmen und deren strategische Bedeutung für die Weiterentwicklung der Evaluierungskultur in Österreich eingegangen.

Der letzte Teil der Studie kommt auf Basis der abgeleiteten evaluierungstheoretischen Grundsätze sowie der Beispiele aus der internationalen und vor allem österreichischen Evaluierungspraxis zu Schlußfolgerungen und Empfehlungen, die auf die Weiterentwicklung der österreichischen Evaluierungskultur im FTE-Bereich abzielen.

1 Grundlegende Themen und Fragestellungen bei Programmevaluierungen

1.1 Wozu braucht man Evaluierungen öffentlicher Programme?

Die Bewertung der von öffentlichen Stellen finanzierten Programme erfüllt mehrere Funktionen, deren relative Bedeutung von den politischen und institutionellen Rahmenbedingungen abhängt, unter denen ein Programm abläuft, wie auch von den Merkmalen der jeweiligen Programme selbst. Es können folgende wesentliche Funktionen unterschieden werden:

- Die Verwendung öffentlicher Gelder ist an eine besondere Sorgfaltspflicht gebunden. Aus diesem Grund muß die Öffentlichkeit regelmäßig über die Tätigkeit öffentlicher Stellen informiert und die Verwendung der Finanzmittel gerechtfertigt werden. Beide Funktionen sollten von Evaluierungen erfüllt werden.
- Auch innerhalb der öffentlichen Verwaltung ist eine ressortinterne bzw. ressortübergreifende Kontrolle notwendig, um aufgetretene Mängel und vorhandene Schwachstellen im Verwaltungsablauf zu erkennen und zu korrigieren. Die Evaluierung ist in diesem Sinne dem laufenden Controlling eines privaten Unternehmens gleichzusetzen.
- Eine der wesentlichen Funktionen der Evaluierung betrifft die Beschaffung, Aufbereitung und Interpretation programminterner bzw. programmrelevanter Fakten. Diese Informationen sind nicht nur für das Funktionieren des laufenden Programmes, also die Erreichung der vorgegebenen Programmziele, von Nutzen (operative Informationsfunktion einer Evaluierung), sondern auch für die Gestaltung zukünftiger Programme und technologiepolitischer Entscheidungen (strategische Informationsfunktion) von Bedeutung. Nutznießer dieser Informationen sind Programmplaner und -manager, wie auch technologiepolitische Entscheidungsträger, Wissenschaftler und bestimmte am Programm interessierte Gruppen.

Evaluierungen können somit eine zielorientierte Kontrollfunktion als auch eine informationsorientierte Lernfunktion haben, wobei die meisten Evaluierungen beide Funktionen erfüllen. Die konkreten Evaluierungsinhalte sind stark von den jeweiligen Funktionen geprägt.

1.2 Wer verfolgt welche Interessen in einer Evaluierung?

Aus den unterschiedlichen Funktionen einer Evaluierung lassen sich auch unterschiedliche Interessen ableiten, die mit einer Evaluierung verbunden sind, wie auch unterschiedliche Akteure, die diese Interessen verfolgen (sogenannte "stakeholders"). Tabelle 1 stellt die Verbindungen zwischen den Akteuren einer Evaluierung (also den an dem Programm bzw. seiner Evaluierung interessierten Gruppen) und den Funktionen dar.

Tabelle 1

KREUZKLASSIFIKATION: Funktion - Akteure	
Funktion	Akteure
Programmformulierung	
Mithilfe bei der Programmplanung	Programmträger und ihm übergeordnete Stellen; Programmplaner.
Kontrolle	
Rechtfertigung der Verwendung öff. Gelder	Programmträger und ihm übergeordnete Stellen; finanzpolitische Entscheidungsträger; Öffentlichkeit.
Ressortinterne bzw. -übergreifende Kontrolle im öffentlichen Bereich.	dem Programmträger übergeordnete Stellen.
Learning	
Strategisches Controlling	Technologiepolitische Entscheidungsträger
Operatives Controlling	Programmplaner, -manager, -mitarbeiter
Andere Funktionen	
Wissenschaftliche Erkenntnisse	Wissenschaftler
Aufzeigen direkter und indirekter Effekte	Adressaten, Betroffene

Der Auftraggeber einer Programmevaluierung ist in der Regel der Programmträger oder eine übergeordnete Stelle; auch direkt oder indirekt Betroffene oder politische Interessensgruppen können eine Evaluierung initiieren bzw. sind an den Evaluierungsergebnissen interessiert.

Daneben besteht auch bei den Programmitarbeitern, Wissenschaftlern und der Öffentlichkeit Interesse an den Evaluierungsergebnissen.

Die Funktionen einer Evaluierung werden wesentlich von den Interessen des Auftraggebers bzw. der ihm übergeordneten Instanzen bestimmt, wobei jedoch darauf zu achten ist, daß genügend Spielraum für die Evaluatoren bleibt, die von ihnen als kritisch erkannten Inhalte unabhängig von den Direktiven des Auftraggebers zu verfolgen.

Eine Möglichkeit der "Arbeitsteilung" zwischen Evaluatoren und Auftraggebern der Evaluation sieht vor, daß der Auftraggeber über die grundsätzlichen Fragestellungen der Evaluation entscheidet, während die Evaluatoren die methodischen Standards (Detailfragen, Methoden, kritische Punkte) zur Beantwortung dieser Fragestellungen festlegen und die Evaluation durchführen. Der Informationsfluß zwischen Auftraggebern und Evaluatoren sollte jedoch in keiner Phase des Evaluationsprozesses abreißen, um einerseits, aus Sicht der Evaluatoren, die inhaltlichen Wünsche des Auftraggebers zu berücksichtigen (ohne jedoch dessen Erwartungen bezüglich der Evaluierungsergebnisse zu internalisieren), andererseits den Informationsbedarf der Evaluatoren zu erfüllen, soweit dieser von den Auftraggebern abgedeckt werden kann.

1.3 Wie geht man grundsätzlich an eine Evaluierung heran?

In der theoretischen Evaluierungsliteratur werden eine Reihe unterschiedlichster Evaluierungsansätze bzw. Herangehensweisen an eine Evaluierungsaufgabe diskutiert. Diese Ansätze unterscheiden sich unter anderem mit Bezug auf:

- die verwendeten Methoden,
z.B. qualitative versus quantitative Methoden,
- die Position der Evaluatoren gegenüber dem Programmträger:
distanziert, also wenig Kontakt mit den am Programm Beteiligten; oder
involviert, das heißt der Evaluator versucht, die Programmwirklichkeit aus den Erfahrungen der am Programm Beteiligten bzw. unmittelbar aus eigenem Erleben des Programmablaufs zu erfassen,
- den wissenschaftlichem Anspruch bzw. das wissenschaftliche Leitbild, dem die Evaluierung folgt:
Durchführung der Evaluation nach streng wissenschaftlichen, d.h. werturteilsfreien Kriterien
oder wertorientierte Evaluierung,
- den Hauptuntersuchungsgegenstand:

Zielorientierung, Kausalitätsbeziehungen oder Analyse der direkten und indirekten Effekte etc. können im Mittelpunkt einer Evaluierung stehen,

- die verwendeten sozialwissenschaftlichen Paradigmen: Ökonomie, Politikwissenschaft, etc.

Jeder Versuch, Methoden und Ablaufstrukturen von Evaluierungen nach nur einem einzigen Ansatz auszurichten, ist jedoch im Konflikt mit der *Verschiedenartigkeit von Evaluierungen bzw. der von ihnen zu bewertenden Programme*. Die Vielfalt der Fragestellungen, die unterschiedlichen sozio-ökonomischen Rahmenbedingungen und die unterschiedlichen Anforderungen, welche die jeweiligen Akteure an die Evaluierung stellen, erlauben keine starren Evaluierungsschemas, sondern erfordern einen methodischen Pluralismus, der aus den verschiedenen Ansätzen die für die jeweilige Evaluierung geeignetsten Instrumente und Theorien wählt.

Aus diesem Grund ist es auch problematisch bzw. schwierig, Evaluierungen zu systematisieren bzw. zu standardisieren. *Standardisierungen stehen grundsätzlich im Widerspruch zu einer notwendigerweise flexiblen Gestaltung von Evaluationen*. Die Evaluierungsergebnisse können durch Verfolgung eines starren methodischen Schemas verfälscht werden (etwa durch die Anwendung von Methoden, die nicht an die verfügbaren Daten oder die spezifischen Merkmale des Programmumfeldes angepaßt wurden) oder unbrauchbar sein, weil sie an den Bedürfnissen der Akteure vorbeigehen.

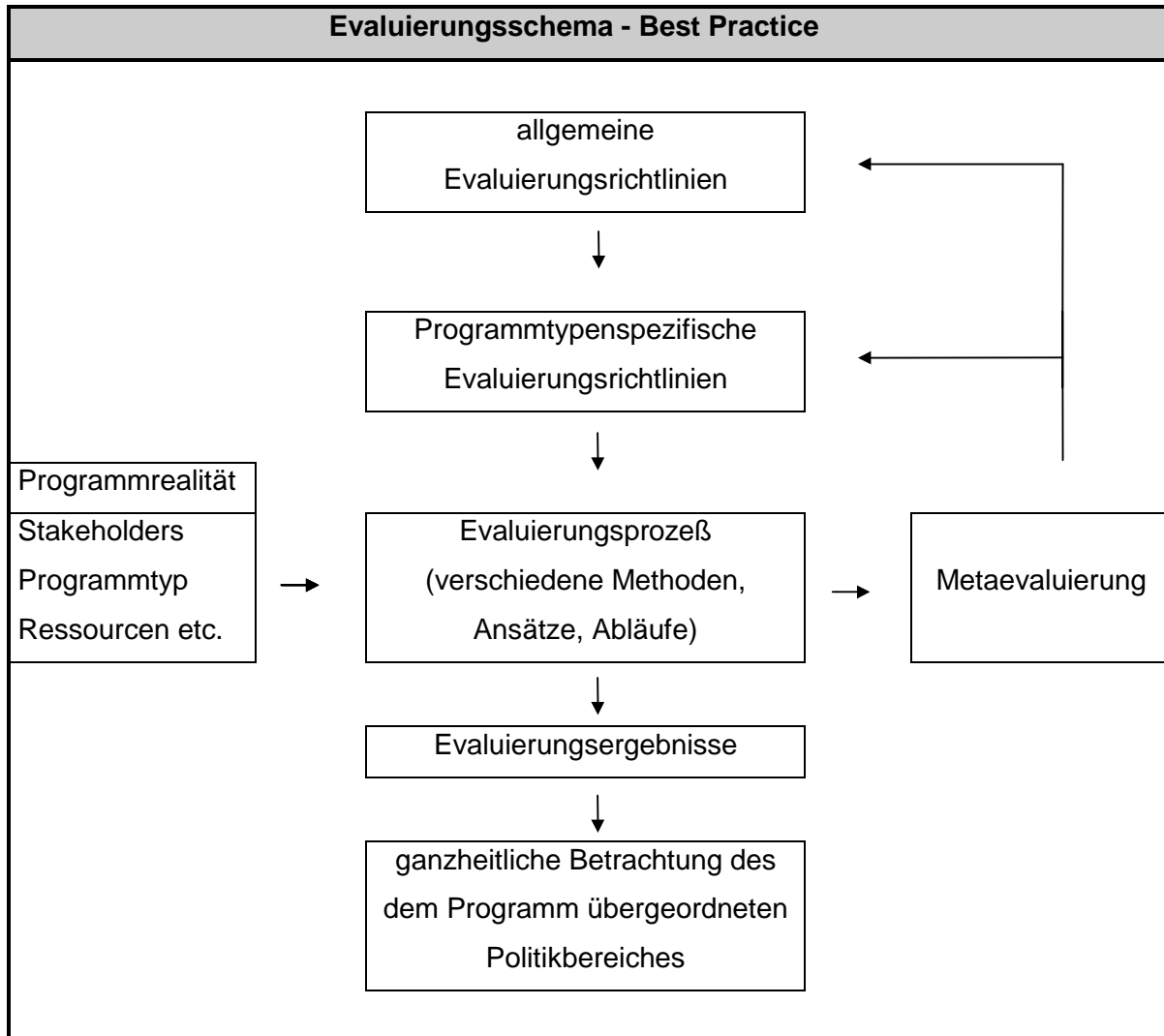
Andererseits sind *gewisse Standards, an denen sich Evaluatoren orientieren können, notwendig und sinnvoll, um die Qualität von Evaluierungen sicherzustellen, sie vergleichbar zu machen und eine Evaluierungskultur zu entwickeln*. Aus dieser Dichotomie heraus muß ein allgemeiner Evaluierungsansatz, also eine "best practice" entwickelt werden, der die Ansprüche der Flexibilisierung und Standardisierung vereint.

In dieser Studie wird die folgende Definition von "best practice" verwendet:

Best practice ist jene Evaluierungspraxis, die Methoden, Ansatz und Ablauf einer Evaluierung auf die Programmrealität abstimmt, ohne bestimmte Evaluierungsstandards zu vernachlässigen.

Solche Standards könne sowohl allgemein als auch programmtypenspezifisch vorgegeben und aus der Evaluierungspraxis abgeleitet werden. Dieses Evaluierungsleitbild ist in Grafik 1 dargestellt.

Grafik 1



Allgemeine Standards betreffen beispielsweise die Festlegung der für die Evaluierungsbeauftragung zuständigen Stelle, die Kriterien für die Durchführung einer internen oder einer externen Evaluierung oder einen an der Programmgröße orientierten Evaluierungsbudgetrahmen. Auch Standards für die Veröffentlichung des Evaluierungsberichts fallen in diese Kategorie. Dabei könnte gelten, daß Ex-ante Evaluationen, die eine Informationsfunktion für die Programmplaner erfüllen, nicht veröffentlicht werden müssen, solche, bei denen keine automatische Rückkoppelung mit der Programmplanung besteht, auf jeden Fall veröffentlicht werden sollten. Gleiches könnte für Interim-Evaluationen gelten, während hingegen Ex-post Evaluationen prinzipiell immer veröffentlicht werden sollten.

Programmtypenspezifische Richtlinien werden im allgemeinen auch bestimmte methodische Standards festlegen. Das könnte etwa die vorgeschriebene Durchführung einer quantitativen

Analyse der indirekten Effekte umfassen, und zwar bei solchen Programmen, deren Nutzen primär durch die Größe dieser Effekte bestimmt wird; oder auch die verbindliche Befragung der Zielgruppenmitglieder über die direkten Effekte eines bestimmten Programmes.

Die methodischen Detailfragen werden für jedes Programm einzeln festgelegt und sind von dem Programmtyp, den Wünschen des Programmträgers, dem vorhandenen Evaluierungsbudget und anderen Aspekten der "Programmrealität" festgelegt.

Die Festlegung allgemeiner und programmtypenspezifischer Richtlinien obliegt den Verantwortlichen jenes Politikbereiches, der den einzelnen Programmen übergeordnet ist. Die Standards sollten unter Mitarbeit erfahrener Programmmitarbeiter sowie Evaluierungsexperten dieses Bereichs erarbeitet werden und periodisch, nach Auswertung der Erkenntnisse aus der Durchführung neuer Programme, überprüft und gegebenenfalls revidiert werden. Die Analyse mehrerer Programmevaluierungen im Rahmen von Metaevaluierungen (also "Evaluierungen von Evaluierungen") könnten sich dabei als hilfreich erweisen. Auf jeden Fall ist aber ein Aufweichen der Richtlinien durch ständige Anpassung an die Gegebenheiten einzelner Programme zu vermeiden.

1.4 Wann werden Evaluationen durchgeführt?

Zeitliche Evaluierungsdimension

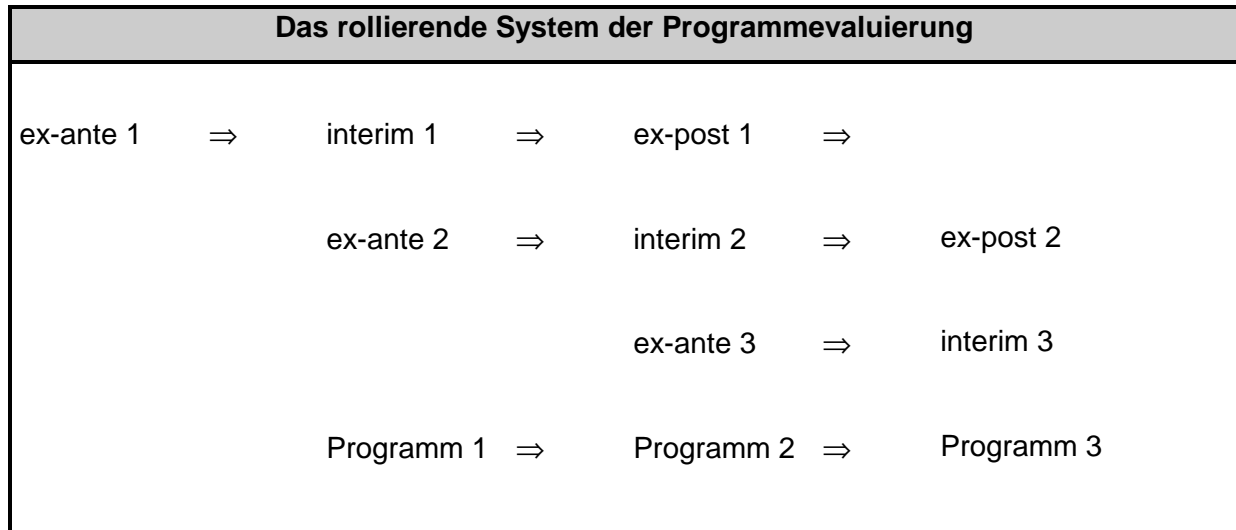
Die am weitest verbreitete Typisierung von Evaluierungen basiert auf dem Zeitpunkt ihrer Durchführung relativ zum Programmverlauf. Dabei unterscheidet man:

- **Ex-ante Evaluierungen**, die vor Programmbeginn stattfinden;
- **Interim Evaluierungen**, die während des Programms durchgeführt werden; und
- **Ex-post Evaluierungen**, die nach Programmende stattfinden.

Die innerhalb eines Politikbereiches ablaufenden Programme sollten im Rahmen eines "rollierenden" Systems einer ständigen Bewertung unterzogen werden (siehe Grafik 2). Dadurch wird die Entwicklung eines "learning system of public policy" gefördert. Innerhalb eines solchen Systems fließen die Ergebnisse und Erfahrungen aus der Durchführung der einzelnen Programme sowie deren Evaluierungen in die Gestaltung neuer Programme und deren Bewertung ein.

Ein solches System erfordert von den Evaluatoren und Programmträgern ein hohes Maß an Bereitschaft zur Kooperation, da der Erfahrungsaustausch das zentrale Element darstellt. Weiters wäre eine klare inhaltliche und auch zeitliche Abfolge von Programmen von Vorteil.

Grafik 2



Die Klassifikation in ex-ante, interim und ex-post sagt nichts über den genauen Zeitpunkt der Durchführung der Evaluierung aus, also wie lange vor bzw. nach Programmbeginn bzw. -ende evaluiert wird. Der Zeitpunkt ist jedoch für die Evaluierungsinhalte, insbesondere die Wirkungsanalyse, von Bedeutung. Dieser Punkt wird in Abschnitt 1.6 näher erläutert.

Während sich Ex-ante-Evaluierungen durch ihren zukunftsgerichteten Charakter in Inhalt und Funktion stark von Interim und Ex-post-Evaluierungen unterscheiden und die Durchführung vor dem Programmanfang auch eine eindeutige zeitliche Zuordnung erlaubt, sind nicht nur Inhalt und Funktion der beiden letzten Evaluierungstypen ähnlich, sondern ist auch deren zeitliche Trennung nicht immer exakt realisierbar. Zwar ist der Übergang von interim zu ex-post durch das Programmende theoretisch eindeutig festgelegt, doch viele Programme haben kein zeitlich eindeutig bestimmtes Ende, woraus sich ein praktisch fließender Wechsel von interim zu ex-post ergibt. Inhaltliche Überschneidungen resultieren aus der sowohl interim als auch ex-post durchgeführten Wirkungsanalyse und aus der Tatsache, daß Ex-post-Analysen meist auf den Ergebnissen der Interim-Untersuchung aufbauen.

1.5 Was soll bewertet werden?

Evaluierungsinhalte

Die folgende Tabelle 2 umfaßt allgemeine, d.h. nicht-programmspezifische Evaluierungsinhalte bzw. -fragestellungen, die Teil einer Evaluierung sein können. Die tatsächlich behandelten Inhalte hängen vom Programmtyp, den Funktionen, den zur Verfügung stehenden Methoden und nicht zuletzt von den Auftraggebern, kurzum von der Programmrealität ab. Eine ausführliche Beschreibung der Inhalte erfolgt unter Punkt 1.6.

Tabelle 2

Inhalte einer Programmevaluierung	
	Beispiel: Programm zur Erhöhung der F&E- bzw. Innovationsaktivitäten von KMUs durch einen Innovationsassistenten
Konzeptevaluierung	
Überprüfung der/des:	
Problem- und Ursachenbeschreibung	Im Vergleich zu größeren Unternehmen führen relativ weniger kleine und mittlere Betriebe F&E- und Innovationsaktivitäten durch. Welche Ursachen stecken dahinter? Kapitalaufbringung, Risikoneigung, vor allem aber zu geringes Know-how im F&E-Bereich,...
Interventionslegitimation	F&E, Innovationen werden immer wichtiger; um die Konkurrenzfähigkeit der österreichischen KMUs zu sichern, muß ihre Innovationsfähigkeit erhöht werden. Voraussetzung dafür ist auch eine Erhöhung ihrer F&E-Aktivitäten;

sozio-techno-ökonomischen Rahmenbedingungen	Österr. Industriestruktur geprägt von vielen KMUs, die vor allem im mittleren bis Niedrigtechnologiebereich produzieren, eine geringe Akademikerquote aufweisen,.....
Ziele	Erhöhung der Anzahl jener KMUs, die F&E- und Innovationsaktivitäten durchführen.
Zielgruppenfestlegung	KMUs aller Industriebranchen
potentiellen Zielkonflikte	Programm steht mit anderen Programmen in Konkurrenz um Budgetmittel. Programm verursacht Erhöhung der Personalkosten im F&E-Bereich.
Lösungsansatzes und möglicher Alternativen	Mit öffentlichen Mitteln geförderte Einstellung eines Universitäts-oder Fachhochschulabsolventen in einem KMU. Alternativen: Förderung des Wissentransfers durch Transferzentren; Förderung von kooperativen Projekten zwischen KMUs und Universitäten, Fachhochschulen.
Zeithorizont	vorläufige Programmdauer: 3 Jahre
Programmmaßnahmen	Fehlendes Know-how in KMUs ist ein wichtiger Grund für die Defizite im F&E bzw. Innovationsbereich.
Designevaluierung	
Überprüfung der/des:	
operationalen Ziele	durchschnittliche Erhöhung der Anzahl von F&E-aktiven bzw. innovationsaktiven KMUs

um 10% innerhalb von 3 Jahren.

Interventionsmodells

Förderung ist die Vermittlung und Bezahlung des Gehalts des Innovationsassistenten für ein Jahr, Einstellungsgarantie des Unternehmens für ein weiteres Jahr. Das Beschäftigungsverhältnis wird der Innovationsassistent dazu benutzen, mit dem in seiner Ausbildung erworbenen Wissen die Entwicklung neuer Produkte und Prozesse anzuregen.

Organisatorischen Struktur: Maßnahmen, Programmablauf, -kosten

Gehaltsniveau des Innovationsassistenten: Branchendurchschnitt für Akademiker bzw. Fachhochschulabgänger.
Org. Maßnahmen: Werbekampagne an Universitäten und Fachhochschulen für den Beruf eines Innovationsassistenten.
Information der Wirtschaftskammer über das Programm. Beginn: 1.1.1998, Ende der Einreichfrist von Förderanträgen: 1.1.2001.....
Programmbudget: 100 Mio ÖS
Form, Inhalt der Förderanträge. Verfahren zur Überprüfung der Anträge.

Prozeßevaluierung

Überprüfung der Implementierung:

Programmrealisierung vs. Programmplanung

Wurden die im Organisationsplan beschriebenen Maßnahmen eingehalten? Wie sieht das Gehaltsniveau der Innovationsassistenten wirklich aus?

administrative und operative Effizienz und Effektivität

Wurden die Werbe- und Informationsmaßnahmen effektiv sowie sparsam durchgeführt? Wie hoch ist der Verwaltungsaufwand des Programms, auch im

Vergleich mit anderen FTE-Programmen. Wie funktioniert die Koordination zwischen den für die Förderung zuständigen Ministerien?

Wirkungsevaluierung

Zielerreichung	Wurden die operationalen Ziele erreicht?
direkte und indirekte Effekte	
kurze und lange Frist	Wie hat sich die F&E- bzw. Innovationsquote (also der Anteil F&E- und innovationsaktiver KMUs an den gesamten KMUs) entwickelt? Welche Erfahrungen haben die Unternehmen mit ihren Innovationsassistenten gemacht? Wurden mehr Produkt- oder Prozeßinnovationen durchgeführt? Konnten die Innovationsassistenten nach Ablauf der 3 Jahresfrist ihre Anstellung verlängern? Können positive Wettbewerbs- und Beschäftigungseffekte festgestellt werden?
Kausalitäten	Gehen die Veränderungen der F&E- bzw. Innovationsquote auf die Einstellung der Innovationsassistenten zurück? Gibt es dafür statistische Evidenz?
ökonomische, soziale, technologische, wissenschaftliche und Umwelteffekte	Haben die verstärkten F&E-Aktivitäten im Bereich der KMUs zu einere verstärkten Zusammenarbeit mit den Universitäten geführt? Welche Technologie wurden durch das Programm indirekt gefördert? Wurde durch das Programm die Akademikerarbeitslosigkeit verringert?
Effizienz	Ist der Nutzen des Programms, der sich aus der zusätzlichen Beschäftigung,

Marktanteilsgewinnen im Ausland etc. ergibt,
höher als die Kosten des Programms?

1.6 Ein Leitfaden der Programmevaluierung

In einem weiteren Schritt wird nun versucht, die funktionellen, zeitlichen und inhaltlichen Dimensionen miteinander zu verknüpfen (Tabelle 3). Damit wird die Absicht verfolgt, einen inhaltlich idealtypischen Evaluierungsrahmen zu schaffen, der insbesondere potentiellen Auftraggebern von Programmevaluierungen als eine Art Leitfaden bzw. Evaluierungs-Checkliste dient, anhand dessen sie Evaluierungsvorschläge konkreter Programme vergleichen können. Weiters soll demonstriert werden, daß ein sinnvoller Leitfaden zur Durchführung einer Evaluierung nur dann entstehen kann, wenn diese Dimensionen gemeinsam betrachtet werden. Eine Umsetzung aller angeführten Punkte wird nur selten möglich sein, es kommt auch hier der "best practice"-Ansatz zur Anwendung, wonach die spezifischen Merkmale eines Programms das endgültige Evaluierungsdesign bestimmen.

Tabelle 3

KREUZKLASSIFIKATION: zeitliche/inhaltliche/funktionelle Dimension	
Ex-ante Evaluierung	
Konzeptevaluierung (strategische Informationsfunktion, Kontrollfunktion)	Ist die Durchführung eines solchen Programmes sinnvoll (Probleme/Ursachen richtig erfaßt, richtiger Lösungsansatz,...)
Designevaluierung (strategische und operative Informationsfunktion.; Kontrollfunktion.)	Sind die Maßnahmen angemessen? Muß das Design verändert werden?
Wirkungsprognose	zur Unterstützung von Konzept- und Designevaluierung

Interim Evaluierung

Prozeßevaluierung (operative, aber auch strategische Informationsfunktion; Kontrollfunktion)	Wird das Programm plangemäß umgesetzt? Ist der Programmablauf effizient / effektiv? Sind Änderungen notwendig? Soll das Programm fortgesetzt werden?
Designevaluierung, Konzeptevaluierung (strategische und operative Informationsfkt.)	Fine tuning, Designveränderung, Konzeptanpassung;
Wirkungsprognose bzw. Abschätzung der bisher erzielten Wirkungen (operative und strategische Informationsfkt., Kontrollfkt.)	Zur Unterstützung der Prozeßevaluierung; vorläufige Analyse der Effektivität und Effizienz des Programms; Welche Schlüsse können aus Prozeßevaluierung und Wirkungsprognose in Hinblick auf zukünftige Programme gezogen werden?

Ex-post Evaluierung

Wirkungsanalyse (Kontrollfunktion; strategische Informationsfkt.)	Hat das Programm die erwarteten Wirkungen gezeigt? Welche unerwarteten Wirkungen sind aufgetreten?
Designevaluierung, Konzeptevaluierung (strategische Informationsfkt.)	Waren Konzept und Design geeignet?
Prozeßevaluierung	Wurde das Programm planmäßig umgesetzt, war der Programmablauf effizient und effektiv? Welche Schlüsse können in Hinblick auf die Implementierung zukünftiger Programme gezogen werden?

Ausgehend von einer zeitlichen Gliederung der Evaluierungstypen werden die einzelnen Evaluierungsschritte und -fragestellungen näher erläutert und mit hypothetischen Beispielen angereichert, wobei auch das unter Punkt 1.5 beschriebene Beispiel des Innovationsassistenten weitergeführt wird.

Ex-ante Evaluierung

Im Rahmen der **Konzeptevaluierung** geht es darum, *die Berechtigung des Einsatzes des Programmes* (inklusive der darin enthaltenen Lösungsansätze) *zur Bekämpfung der vorher identifizierten ökonomischen, technologischen und sonstigen gesellschaftlichen Probleme zu bewerten*. Zusätzlich sollen die der Gestaltung des Programmes zugrundeliegenden Prinzipien analysiert werden.

Bei der Konzeptevaluierung besteht besonders die Gefahr eines Interessenskonflikts mit dem Auftraggeber der Evaluationsstudie, falls dieser das zu evaluierende Programm auch initiiert hat und damit für die in diesem Teil zu evaluierende grundsätzliche Ausrichtung des Programms verantwortlich ist.

Problem- und Ursachenbeschreibung

Ziel ist die Analyse jenes gesellschaftlichen Problems (z.B. geringe industrielle F&E-Aktivitäten und damit Defizite bei der Entwicklung neuer Produkte bzw. Verbesserung von Produktionprozessen), welches durch das vorliegende Programm gelöst bzw. gemildert werden soll. Welche Indizien gibt es, daß es sich tatsächlich um ein Problem handelt (z.B. geringere Anzahl von KMUs, die innovierend tätig sind)? Was sind seine Ursachen (z.B. fehlendes Know-how)? Letzteres ist kritisch für den Erfolg des Programms. Es muß hinterfragt werden, ob auch die Ursachen des Problems richtig erkannt wurden. Beruhen sie auf bloßen Annahmen oder sind sie durch empirische Indizien abgesichert? Wurde eine Befragung oder sonstige empirische Analyse der Zielgruppenmitglieder durchgeführt, um die Annahmen zu testen? Je nach Evaluierungsauftrag könnte eine solche Analyse auch durch die Evaluatoren selbst erfolgen, was besonders bei umfangreichen Programmen und Programmen mit langer Laufzeit empfehlenswert scheint. Ein auf falschen Ursachenhypothesen beruhendes Programm ist schon vor seiner Umsetzung zum Scheitern verurteilt.

Interventionslegitimation

Soll die öffentliche Hand korrigierend eingreifen? Das vorliegende gesellschaftliche Problem muß genau spezifiziert und der geplante Eingriff gerechtfertigt werden. Die Existenz eines

gesellschaftlichen Problems (z.B. zu wenig KMUs, die innovieren, forschen und entwickeln) alleine rechtfertigt noch keinen Eingriff. Die Intervention muß zu einer Wohlfahrtssteigerung führen, was a priori schwer feststellbar ist, doch zumindest argumentativ schlüssig begründet werden muß.¹ Das heißt, es muß schlüssige Indizien dafür geben, daß das Programm erstens erfolgreich sein wird und zweitens der Programmnutzen (z.B. wettbewerbsfähigere KMUs) die Programmkosten (Fördermittel, Verwaltungskosten, Kosten der durch das Programm induzierten Wettbewerbsverzerrungen etc.) übertreffen.

Sozio-techno-ökonomische Rahmenbedingungen

Auch müssen diejenigen rechtlichen, politischen, technologischen und sonstigen Rahmenbedingungen, die für das Programm relevant erscheinen, analysiert werden. Dazu gehören zum Beispiel bestimmte Rechtsvorschriften, die den Intentionen des Programms zuwiderlaufen könnten (etwa die Einhaltung von EU-Vorschriften bezüglich unerlaubter Subventionen) oder wirtschaftliche und technische Rahmenbedingungen jener Sektoren, die das Programm besonders betrifft, wie auch institutionelle oder kulturelle Barrieren. Bei einem Innovationsassistentenprogramm muß hinterfragt werden, ob das Angebot an technisch gebildeten Assistenten bzw. Forschern elastisch genug ist, um der Nachfrage einer verstärkten F&E-Aktivität kleiner und mittlerer Betriebe gerecht zu werden.

Ziele

Die grundsätzlichen Ziele eines Programmes müssen auf ihre Legitimität und Erreichbarkeit hin bewertet werden. Es ist auch besonders auf die Kompatibilität des Programms mit anderen staatlichen Maßnahmen zu achten: Dazu gehören potentielle Zielkonflikte zum Beispiel mit der Finanzpolitik (vor allem wenn es um die Verteilung knapper öffentlicher Finanzmittel geht), der Wettbewerbspolitik (so könnte etwa die Förderung von Unternehmenskooperationen im Bereich der Forschung und Entwicklung zu einer unerwünschten Einschränkung des Wettbewerbs führen) oder der Beschäftigungspolitik (wenn zum Beispiel die Einführung neuer Technologien arbeitssparende Effekte hat). In der realen Welt sind Zielkonflikte praktisch unvermeidbar. Ihre Offenlegung vor der Durchführung eines Programmes ist gerade deshalb wichtig.

Zielgruppen

¹ In den USA wurde die Gesetzgebung erst unlängst dahingehend geändert, daß jedes öffentliche Programm ex-ante einer Kosten-Nutzenanalyse unterzogen werden muß, um eben diese Wohlfahrtssteigerungen zu bestimmen.

In der Programmplanung sollte festgelegt sein, an welche Branchen, Bevölkerungsgruppen, Unternehmenstypen etc. sich das Programm primär wendet. Die Zielgruppe könnte dabei sehr eng gefaßt sein (z.B. Unternehmen einer eng gefaßten Branche/Region), oder auch sehr weit (z.B. alle österreichischen Industriebetriebe). Die Evaluierung sollte untersuchen, ob ausreichende Informationen über die Zielgruppenmitglieder vorhanden sind (so zum Beispiel über die Gründe geringerer Innovationsaktivitäten im KMU-Bereich) und das Programm an die spezifischen Merkmale der Zielgruppe (z.B. kleine und mittlerer Betriebe) angepaßt wurde, was für die Effektivität des Programms von großer Relevanz ist. Die Durchführung einer empirischen Zielgruppenanalyse ist in Betracht zu ziehen. Auch die Akzeptanz des Programmes unter den Zielgruppenmitgliedern ist eine für den Erfolg des Programms wichtige Frage, die mittels einer Befragung getestet werden könnte. Dabei ist jedoch darauf zu achten, daß es im Interesse der Zielgruppenmitglieder liegen könnte, staatliche Förderungen auf sich zu ziehen und somit die Umfrageergebnisse zu verfälschen.

Lösungsansatz

Die Evaluierung muß bewerten, ob die festgelegten Ziele mit dem gewählten Lösungsansatz erreichbar sind bzw. erreichbar scheinen, ob dieser Ansatz in sich schlüssig und mit vorhandenen theoretischen und empirischen Erkenntnissen abgesichert ist. Gleichzeitig sind mögliche Alternativen mit dem gewählten Lösungsansatz zu vergleichen. Dabei sollten Wirkungsprognosen gemacht werden und die prognostizierten Effizienzen und Effektivitätsgrade verglichen werden.

Bei einem Programm zur Förderung der F&E-Aktivitäten von KMUs, wie es etwa ein Innovationsassistentenprogramm darstellen würde, muß es also genügend Indizien für den Erfolg des Programms geben und auch der Verzicht auf die Durchführung anderer Maßnahmen der F&E-Förderung anstelle des Innovationsassistentenprogrammes schlüssig begründet werden.

Zeithorizont

Es ist zu bewerten, ob

- die geplante Programmdauer ausreichend für die Erreichung der Programmziele ist
- die Ziele auch durch eine reduzierte Programmdauer erreicht werden könnten,
- permanente Programme berechtigt sind oder die Anstoßwirkungen temporärer Programme ausreichen.

Programmannahmen

Bei der Planung einer staatlichen Intervention werden oft, implizit oder explizit, zahlreiche Annahmen getroffen, die für den Erfolg des Programms entscheidend sein können. Darunter fallen Annahmen über jene Faktoren, die ein gesellschaftliches Problem verursachen (siehe den Abschnitt über Ursachen- und Problembeschreibung), Annahmen über die Merkmale der Zielgruppenmitglieder, das zugrundeliegende theoretische Modell etc. Es ist Aufgabe der Evaluatoren, diese Annahmen, falls implizit, aufzuzeigen und implizite und explizite Annahmen auf ihre Richtigkeit zu überprüfen. Die Überprüfung der Annahmen sollte idealerweise empirisch erfolgen. Falls dies nicht möglich ist, kann auf andere Studien bzw. die vorhandene Literatur in diesem Bereich verwiesen werden.

Während im Rahmen der Konzeptevaluierung die theoretischen und empirischen Grundlagen eines Programmes bewertet werden sollen, geht es in der **Evaluierung des Programmdesigns** darum, die konkrete Gestaltung eines Programmes zu bewerten.

Operationale Ziele

Im Rahmen der Evaluierung sollte überprüft werden, inwieweit operationale Ziele im Programm enthalten sind.

Operationale Ziele sollen im Idealfall

- explizit, das heißt in der Programmplanung festgehalten, statt nur impliziert oder angenommen sein,
- spezifisch ("Erhöhung des Anteils jener KMUs, die innovieren", nicht jedoch: "...mit diesem Programm sollen Betriebe gestärkt werden...")
- meßbar ("Erhöhung des Anteils jener KMUs, die innovieren, um 10%.....")
- und zeitlich fixiert sein ("....bis zum Jahr 1999")

Eine Zieloperationalisierung wird vielfach erschwert, da

- die Wirkungen des Programmes a priori nur schwer eingeschätzt werden können, weshalb Ziele allgemein statt spezifisch formuliert werden (z.B. "deutliche Erhöhung der F&E-Ausgaben");
- viele Ziele schwer meßbar sind (z.B. Beschleunigung des technologischen Wandels).

Erst die Operationalisierung von Zielen macht eine spätere Bewertung der Programmeffektivität möglich. Vom Standpunkt der Evaluierung sind operationale Ziel deshalb unbedingt anzustreben, während vom Standpunkt der Programmplanung die Zieloperationalisierung oft als willkürlich und der Programmrealität nicht entsprechend angesehen wird.

Interventionsmodell

Das Interventionsmodell beschreibt die *Verbindung der im Rahmen des Programmes geplanten Maßnahmen mit den Reaktionen der Zielgruppe* (direkte Effekte, z.B. höhere F&E-Ausgaben in Betrieben, die einen Innovationsassistenten beschäftigen) *und anderen indirekten Effekten* (z.B. Entwicklung neuer Produkte als Folge der intensiveren F&E-Aktivitäten, Ausweitung der Produktion, Aufnahme neuer Beschäftigter, Stärkung der Wettbewerbsfähigkeit österreichischer KMUs). Das Interventionsmodell spezifiziert also die erwartete Maßnahmen-Wirkungskette des Programms. Alle dabei gemachten Annahmen (z.B. ein Innovationsassistent wird im Unternehmen vor allem mit F&E-Aufgaben betraut werden) müssen auf ihre Schlüssigkeit hin überprüft werden, wobei die theoretischen und empirischen Erkenntnisse

Organisatorische Struktur

Die organisatorische Struktur des Programms, das heißt alle im Rahmen des Programmes geplanten Aktionen und ihre Abfolge müssen unter anderem hinsichtlich folgender Kriterien bewertet werden:

- die geplanten Aktionen müssen mit dem gewählten Lösungsansatz in Einklang stehen,
- sie müssen effizient, effektiv und durchführbar sein,
- sie müssen in zeitlich konsistenter Abfolge zueinander stehen,
- die Programmkosten müssen nachvollziehbar sein.

Im Rahmen einer Ex-ante Evaluierung muß auch eine **Wirkungsprognose**, also eine Abschätzung der zu erwartenden direkten und indirekten Effekte, durchgeführt werden. Eine solche Prognose sollte idealerweise unter Zuhilfenahme quantitativer Methoden erfolgen. Dazu eignen sich vor allem mathematisch-statistische Prognosemodelle, ergänzt durch Szenariotechniken. Ist die Anwendung quantitativer Methoden aufgrund fehlender Daten, unzulänglicher Modelle etc. nicht möglich bzw. erlaubt die Art der Fragestellung keine sinnvolle Anwendung solcher Methoden, sollte trotzdem eine Abschätzung der möglichen Programmwirkungen vorgenommen werden. Dazu sind alle verfügbaren Informationen

heranzuziehen, die aus Expertenbefragungen, nationalen und internationalen Erfahrungen mit ähnlichen Programmen etc. gewonnen werden können.

Das Ergebnis einer solchen Wirkungsprognose sollen die zu erwartenden, direkten und indirekten, positiven sowie negativen Effekte sein, die gegeneinander abgewogen werden müssen. Die Durchführung des Programmes kann nur empfohlen werden, wenn der erwartete Programmnutzen die erwarteten Programmkosten übertrifft, was von den Evaluatoren in quantitativer oder qualitativer Form überprüft werden muß

Interim-Evaluierung

Die begleitende Evaluierung erfüllt folgende Aufgaben:

Erstens muß im Rahmen der **Prozeßevaluierung** die Implementierung des Programms bewertet werden. Die Prozeßevaluierung umfaßt folgende Fragestellungen:

Vergleich Programmplanung - Programmumsetzung

Es gilt festzustellen, ob die Programmrealisierung von der Programmplanung abweicht, und wenn ja in welchen Bereichen. Sind die Abweichungen bloße sinnvolle Anpassungen des Programmdesigns als Folge der seit Programmbeginn gemachten Erfahrungen ("learning and adjusting by doing"), und tragen damit zur Erreichung der Programmziele bei, oder gibt es keine rationale Begründung für die Nichtbefolgung des Programmplanes? Welche Auswirkungen haben diese negativen Abweichungen?

Analyse der bisherigen Umsetzungserfahrungen

Die administrativen und operativen Abläufe müssen auf ihre Effizienz und Effektivität geprüft werden. Die Implementierung ist als effektiv anzusehen, wenn die vorgegebenen organisatorischen Aufgaben erfüllt werden; sie ist gleichzeitig effizient, wenn diese Aufgaben mit kleinstmöglichem Mitteleinsatz durchgeführt werden. Vorhandene Schwachstellen in der Programmumsetzung und ihre Ursachen sind zu identifizieren, wobei auch die Bedeutung der aufgetretenen Probleme hinsichtlich der Zielerreichung beurteilt werden sollte. Die Evaluatoren sollten auch Vorschläge zur Behebung aufgedeckter Schwachstellen machen.

Zweitens muß die bisher erreichte *Effektivität* des Programmes, also der Zielerreichungsgrad, beurteilt werden. Ist das Ziel, wie im Innovationsassistentenbeispiel, eine Erhöhung der F&E-Quote im KMU-Bereich um 10% innerhalb von 3 Jahren, dann sollte ein Jahr nach

Programmbeginn schon eine Erhöhung der F&E-Quote bzw. eine Tendenz zur Erreichung der 10%-Marke feststellbar sein. Die Effektivitätsprüfung setzt natürlich die Existenz operationaler (das heißt: expliziter, spezifischer, meßbarer und zeitlich fixierter) Ziele voraus. Fehlen operationale Ziele, kann der Programmoutput nur gemessen (Erhöhung der F&E-Quote um 0.5%), aber nicht (mit den 10%) verglichen werden. In diesem Fall sollten die Evaluatoren von den Programmplanern bzw. den Programmverantwortlichen eine nachträgliche Zieldefinition verlangen, vor allem wenn die Interim-Evaluierung früh im Programmablauf durchgeführt wird.

Im Rahmen der Effektivitätsbeurteilung sollte auch eine qualifizierte Prognose im Hinblick auf den endgültigen Zielerreichungsgrad abgegeben werden.

Drittens muß die *Effizienz* des Programmes bewertet, also Programminputs (alle aufgetretenen Kosten) mit direkten und indirekten sowie erwünschten und unerwünschten Programmoutputs verglichen werden. Besonderes Augenmerk ist dabei auf die Frage der Kausalität zwischen Inputs und Outputs zu legen, das heißt, es muß analysiert werden, ob es einen tatsächlichen Zusammenhang zwischen den im Rahmen des Programms gesetzten Aktionen und den als Programmoutputs definierten Beobachtungen gibt. Haben beispielsweise die Innovationsassistenten die gemessene Erhöhung der F&E-Quote bewirkt oder hängt diese Entwicklung mit ganz anderen Faktoren, etwa EU-Förderungen, zusammen?

Eine negative Programmeffizienz (das heißt die Kosten sind höher als die bewerteten Programmoutputs) vor Abschluß des Programmes kann jedoch nur bedingt als Indikator für den Programmerfolg herangezogen werden, vor allem wenn relativ viele Kosten am Beginn des Programms auftreten, die Wirkungen jedoch eher am Programmende oder sogar danach.

In diesem Zusammenhang muß die Verbindung zwischen dem Auftreten der Programmwirkungen und dem Zeitpunkt ihrer Messung von den Evaluatoren berücksichtigt werden

(siehe Grafik 3).

Je weiter dieser Zeitpunkt vom Programmanfang bzw. Programmende entfernt ist, um so mehr entfalten sich die durch das Programm ausgelösten (vor allem indirekten) Effekte. Wird die Interim-Wirkungsanalyse sehr bald nach Programmanfang durchgeführt, wird nur ein geringer Teil der gesamten Effekte erfaßt werden können. Mit zunehmendem zeitlichem Abstand der ihnen zugeordneten Programminterventionen verlieren diese Effekte aber auch ihre Schärfe, das heißt, die zugrundeliegenden Kausalitäten werden weniger und weniger identifizierbar, das Finden einer verlässlichen Benchmark-Schätzung schwieriger. Der Zusammenhang Programmintervention - Programmwirkung geht nach und nach verloren.

Eventuelle Beschäftigungseffekte eines Innovationsassistentenprogrammes werden wahrscheinlich erst mit mehrjähriger Verzögerung eintreten, während ein Ansteigen der F&E-Quote schon früher sichtbar werden sollte.

Je früher Interim-Wirkungsanalysen durchgeführt werden, um so eher können erste Aussagen über die Programmeffizienz getroffen werden, doch um so weniger Effekte werden erfaßt. Je nach Programmlänge sollten eine oder mehrere Interim-Wirkungsanalysen durchgeführt werden, deren Ergebnisse dann unter Berücksichtigung des Zeitpunkts ihrer Durchführung interpretiert werden müssen.

Programmanpassungen

Sind im ersten Schritt der Prozeß- und Wirkungsanalyse Schwachstellen im Programm identifiziert worden, ist es auch Aufgabe der Evaluatoren, deren Bedeutung für den Programmerfolg zu beurteilen und Strategien zur Behebung dieser Mängel vorzuschlagen. Dabei können geringfügige Anpassungen, etwa in der Programmimplementierung, ausreichend sein („fine tuning“). Mitunter ist es aber auch notwendig, schwerwiegende Mängel im Rahmen einer Design- und/oder Konzeptevaluierung weiter zu analysieren und Lösungskonzepte zu entwickeln. Werden schwerwiegende und schwer korrigierbare Defizite im Konzept oder Design sichtbar, kann auch der Abbruch des Programms empfohlen werden, wobei die Kosten des Programmabbruchs (etwa der teilweise Verlust der bisher in das Programm eingeflossenen Mittel) und der Nutzen eines solchen Schritts (Verhinderung der nicht sinnvollen Verwendung weiterer Mittel) gegeneinander abgewogen werden müssen.

Die Erkenntnisse, die aus der Prozeßevaluierung sowie der vorläufigen Beurteilung des Programmerfolgs (Effektivität, Effizienz) gewonnen werden, haben, wie fast alle anderen Evaluierungsergebnisse, auch strategischen Wert, d.h. sie sollen in die Gestaltung zukünftiger Programme einfließen.

Ex-post Evaluierung

Im Rahmen der Ex-post Evaluierung geht es darum,

- die Erreichung der Programmziele zu bewerten (Effektivitätsanalyse);
- alle durch das Programm direkt oder indirekt ausgelösten Effekte zu identifizieren und soweit wie möglich zu messen und zu bewerten;
dabei ist vor allem auf die unerwarteten Effekte Bedacht zu nehmen;

- und schließlich eine Beurteilung der Programmeffizienz durchzuführen, also einen Vergleich des Programmnutzens, unter Einschluß aller erfaßbaren direkten und indirekten Effekte, mit den Programmkosten.

Effektivitätsanalyse

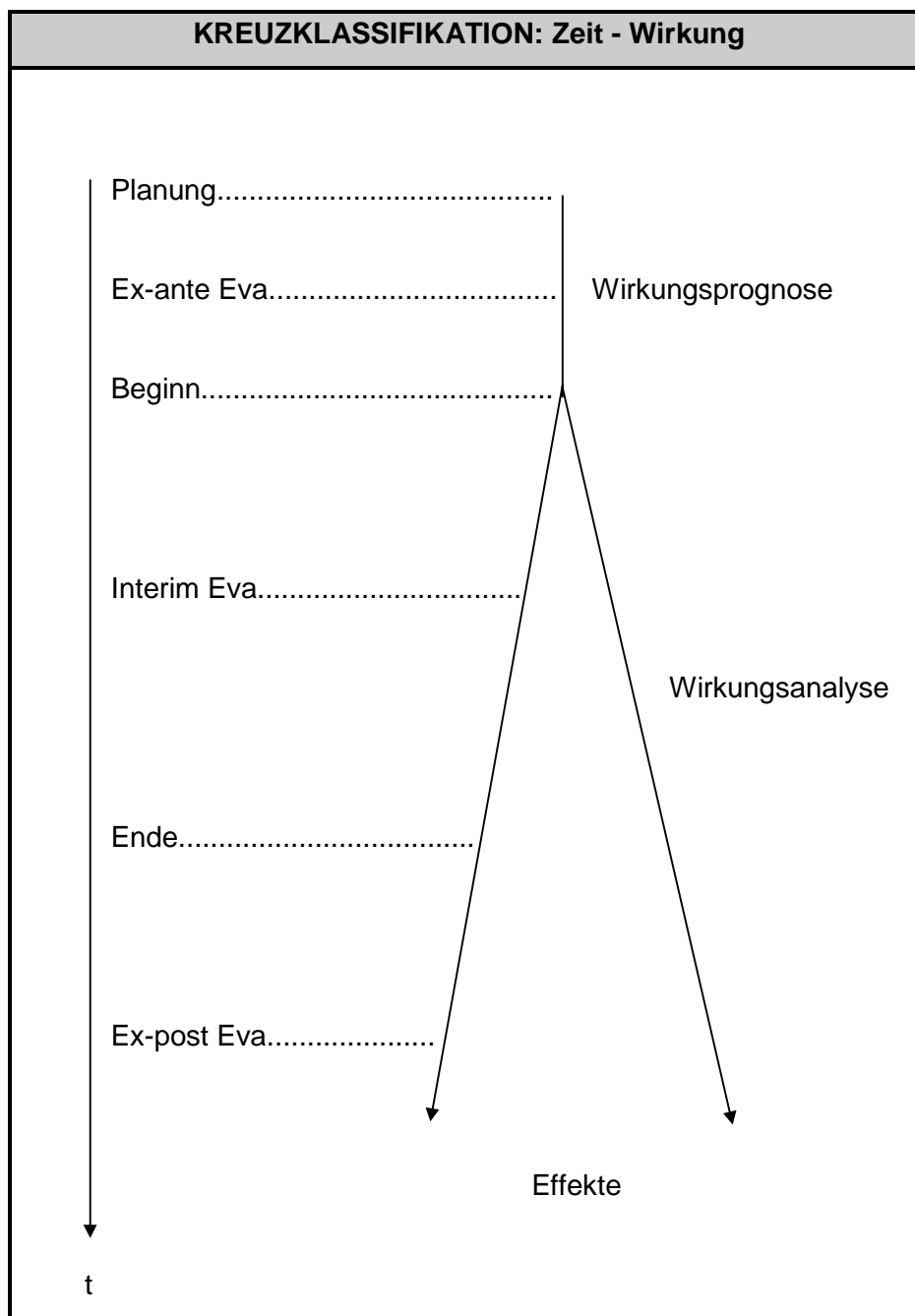
Um die Effektivität eines Programms bewerten zu können, sind operationale, also explizite, spezifische, meßbare und zeitlich fixierte Ziele eine unbedingte Voraussetzung. Fehlen operationale Ziele, können zwar Aussagen über den erreichten Programmoutput gemacht werden, eine Beurteilung der Programmeffektivität ist jedoch nicht möglich.

Effizienzanalyse

Die Beurteilung des Programmerfolgs wird letztlich im Rahmen der Effizienzanalyse durchgeführt, wenn die Programmoutputs den Programminputs gegenübergestellt werden. Nur wenn der Wert der gesamten (direkt und indirekt aufgetretenen) Outputs die gesamten Kosten des Programms (einschließlich aller unerwünschten Outputs) überwiegt, kann das Programm positiv bewertet werden. Das gilt auch dann, wenn die Programmziele (also der vorgegebene Output) erreicht oder sogar übertroffen wurden, wenn also die Programmeffektivität positiv beurteilt wurde.

Dabei muß der Zusammenhang zwischen dem Auftreten der Programmwirkungen und dem Zeitpunkt ihrer Messung von den Evaluatoren berücksichtigt werden (siehe Grafik 3).

Grafik 3



Im allgemeinen gilt: Je weiter dieser Zeitpunkt vom Programmende entfernt ist, um so mehr entfalten sich die durch das Programm ausgelösten (vor allem indirekten) Effekte. Mit zunehmendem zeitlichem Abstand der ihnen zugeordneten Programminterventionen verlieren diese Effekte aber auch ihre Schärfe, das heißt die zugrundeliegenden Kausalitäten werden weniger und weniger identifizierbar, das Finden einer verlässlichen Benchmark-Schätzung schwieriger. Der Zusammenhang zwischen Programmintervention und Programmwirkung geht nach und nach verloren.

Der optimale Zeitpunkt für die Durchführung einer Ex-post Wirkungsanalyse ist dann gegeben, wenn keine wesentlichen direkten und indirekten Programmwirkungen mehr zu erwarten sind. Dieser Zeitpunkt hängt jedoch auch von den Fragestellungen ab, die durch die Evaluierung beantwortet werden sollen. Betreffen diese die direkten Effekte, die durch eine Befragung der Programmteilnehmer erhoben werden sollen (etwa die Aufnahme von F&E-Aktivitäten in Unternehmen, die einen Innovationsassistenten beschäftigen), so kann eine solche Befragung auch relativ bald nach dem Programmende durchgeführt werden, ohne die Aussagekraft der Evaluierungsergebnisse zu schmälern. Stehen jedoch die durch ein Programm ausgelösten indirekten Effekte im Vordergrund (etwa die Stärkung der Wettbewerbsfähigkeit oder die Erhöhung der Exportchancen der teilnehmenden Betriebe als Ergebnis ihrer Forschungsanstrengungen), so würde eine Evaluierung sofort nach Programmende zu unbrauchbaren oder verfälschten Ergebnissen führen.

Die gemessenen Wirkungen müssen schließlich mit dem Programmkonzept und dem Programmdesign in Zusammenhang gebracht werden, um eine abschließende Design- und Konzeptevaluierung durchzuführen und Schlüsse für die Gestaltung zukünftiger Programme zu ziehen. Kritisch für diese Bewertung ist es, festzustellen, ob die beobachteten Wirkungen tatsächlich auf das Programm zurückzuführen sind oder mit anderen, außerhalb des Programms stehenden Faktoren (etwa der Förderung privater Forschungsprojekte durch EU-Gelder) in Zusammenhang stehen. Eine solcher Kausalitätsbeweis, der praktisch die Frage: "Was wäre ohne Programm passiert" schlüssig beantworten muß, ist schwierig zu führen und verursacht beachtliche methodische Probleme, da eine hypothetische Wirklichkeit ohne Programmeinflüsse simuliert werden muß. Dieses Problem tritt auch in vielen anderen ökonomischen Analysen auf: Vor dem Eintritt Österreichs in die Europäische Gemeinschaft wurde etwa von den Forschern verlangt, die ökonomischen Folgen dieses Schrittes ex-ante zu evaluieren. Als erste Stufe dieser Evaluierung mußte ein Szenario entworfen werden, das die zukünftige Entwicklung der österreichischen Wirtschaft ohne Beitritt darstellt. Ein solches Szenario ist klarerweise mit großen Unsicherheiten behaftet, unter anderem deshalb, weil die österreichische Entwicklung nicht unabhängig von der anderer Volkswirtschaften ist.

1.7 Methoden

1.7.1 Quantitative versus qualitative Methoden

Die hier getroffene Unterscheidung zwischen quantitativen und qualitativen empirischen Methoden basiert erstens auf der Herangehensweise an Datenbeschaffung und

Datenauswertung, und zweitens auf der wissenschaftstheoretischen Herangehensweise, also der Unterscheidung zwischen deduktiven und induktiven Methoden.

Quantitative Methoden weisen nach diesen Ordnungskriterien folgende Merkmale auf:

- standardisierte Meßinstrumente werden angewandt, wie statistische, ökonometrische oder mathematische Modelle und Verfahren zur Bestimmung der Beziehung zwischen zwei oder mehreren Variablen oder der Entwicklung von Indikatoren;
- es wird versucht, Meßergebnisse zu vergleichen und zu generalisieren,
- es wird eine deduktive Herangehensweise verfolgt.

In den Bereich der quantitativen Methoden fallen

- die beschreibende und schließende Statistik,
- ökonometrisch-mathematische Verfahren,
- standardisierte Befragungen,
- standardisierte offene und verdeckte Beobachtungen;

Nach den obengenannten Kriterien weisen **qualitative Methoden** die folgenden Merkmale auf:

- sie liefern eine Fülle von Detailinformationen über kleine Anzahl von Akteuren bzw. Sachverhalten,
- Programm und Programmresultate werden auf der Grundlage der Perspektiven einzelner Akteure analysiert,
- und sie verfolgen eine induktive Herangehensweise.

Als qualitative Methoden gelten:

- qualitative Interviews,
- Gruppengespräche,
- Inhaltsanalysen von schriftlichen Unterlagen und Aussagen,
- teilnehmende Beobachtungen, und
- die Aktionsforschung;

Welcher Methodentyp bei welchem Programm bzw. zur Beantwortung welcher Fragestellung herangezogen werden sollte, kann generell nicht gesagt werden, sondern hängt von den

spezifischen Merkmalen des jeweiligen Programms ab. Folgende allgemeine **Vor- und Nachteile** quantitativer bzw. qualitativer Methoden lassen sich jedoch feststellen:

Quantitative Methoden

- quantifizierbare Bewertungen liefern oft eindeutigere Aussagen, die jedoch einer qualifizierten Interpretation bedürfen;
- quantitative Methoden sind zum Test von Hypothesen geeignet;
- und die Wirkungen verschiedener Programme sind durch ihre Standardisierbarkeit miteinander vergleichbar;
- andererseits sind viele Phänomene nicht eindeutig beobachtbar und meßbar und können damit durch quantitative Methoden nicht erfaßt werden;
- die Ergebnisse mathematisch-ökonomischer Modelle sollten jedenfalls auf ihre Robustheit hin getestet werden, das heißt es muß untersucht werden, ob nicht geringfügige Änderungen der Modellparameter oder Daten große Auswirkungen auf die Ergebnisse haben;
- die Qualität solcher Modelle ist stark von Datenverfügbarkeit und -qualität abhängig;
- und sie beruhen oft auf schwer zu verifizierenden Hypothesen oder Annahmen;
- um mathematisch-ökonomische Modelle anwenden zu können, muß die Realität auf wesentliche Zusammenhänge verkürzt werden; durch den daraus folgenden Verlust an Detailinformationen kann die Komplexität der Programmwirklichkeit nicht voll erfaßt werden, wodurch falsche Schlüsse resultieren könnten;
- andererseits ist gerade und nur durch diese Vereinfachung der Wirklichkeit im Rahmen eines Modells die Erfassung wichtiger Zusammenhänge und Kausalitäten möglich;
- vor allem für Fragestellungen, die eine große Zahl interdependenter Variablen umfassen, deren wechselseitige Verbindungen (feedback loops) nicht offensichtlich sind, ist eine modellhafte Darstellung notwendig;

Qualitative Methoden

- qualitative Methoden hingegen versuchen die volle Komplexität zu erfassen;
- sind deshalb aber nicht in der Lage, generalisierbare Erkenntnisse und damit eindeutige Aussagen und Bewertungen abzuleiten; dadurch wird die Entscheidungsfindung erschwert;
- schwer meßbare Größen können durch ein qualitativ orientiertes Herangehen besser erfaßt werden;

Innerhalb eines Programmes werden, je nach Fragestellung, meist beide Methodentypen zur Anwendung kommen bzw. einander ergänzen.

1.7.2 Welche Methoden kommen wo zur Anwendung?

Die Verknüpfung von Methoden mit Evaluierungsinhalten

Die Selektion der für den jeweiligen Evaluierungsinhalt optimalen Methode hängt neben den oben genannten Faktoren auch von zahlreichen anderen Umständen ab, wie etwa

- der Programmgröße (Fördervolumen, Zielgruppengröße etc.),
- dem Evaluierungsbudget,
- dem Programmtyp,
- dem methodischer Ansatz der Programmplanung,
- der Datenverfügbarkeit,
- und vor allem von der Art der erforderlichen Information.

Daraus folgt, daß es *keine allgemein gültige Verknüpfung von Inhalten und Methoden* gibt, sondern daß Programmeigenschaften und -rahmenbedingungen, also die Programmrealität, zusammen mit den methodenspezifischen Merkmalen die optimale Inhalt-Methoden Kombination bestimmt.

Im folgenden wird versucht, mögliche methodische Ansätze für einzelne Evaluierungsinhalte zu beschreiben und zu bewerten, wobei verschiedene hypothetische Beispiele die Ausführungen erläutern sollen.

Ursachenbeschreibung (ex-ante)

Die Hinterfragung bzw. Klärung der Ursachen eines gesellschaftlichen Problems kann von den Evaluatoren durch eine standardisierte oder qualitative Befragung der Betroffenen unterstützt werden. Prinzipiell gilt, daß standardisierte Befragungen bei großen Zielgruppen, qualitative Befragungen bei kleinen Zielgruppen eingesetzt werden sollte.

Ein Beispiel: Klein- und Mittelbetriebe könnten etwa danach befragt werden, welche Faktoren sie bei der Durchführung von F&E-Aktivitäten behindern bzw. sie davon abhalten (z.B. hohe Fixkosten der dazu benötigten Geräte, hohe F&E-Personalkosten, keine F&E-Notwendigkeit etc.). Daraus kann eine auf die am häufigsten auftretenden Probleme abgestimmte Förderung konzipiert werden.

Auch die Ergebnisse vorliegender empirischer Studien bzw. theoretischer Ansätze im Fachbereich des zu evaluierenden Programmes sollten wenn möglich zur Ursachenanalyse herangezogen werden.

Zielgruppenanalyse (ex-ante)

Die in der Programmplanung enthaltenen Annahmen bezüglich der relevanten Charakteristika der Zielgruppenmitglieder sollten von den Evaluatoren überprüft werden. Als Methode eignen sich dafür vor allem standardisierte oder qualitative Befragungen. Auch hier gilt: Je mehr Informationen die Programmplaner über die von einem Programm anvisierten Unternehmen, Institutionen etc. zur Verfügung haben, um so effektiver kann das Programm gestaltet werden.

Bewertung alternativer Lösungsansätze (ex-ante)

Für jedes gesellschaftliche Problem, das des Eingreifens der öffentlichen Hand bedarf, gibt es meist nicht nur einen, sondern mehrere Lösungsansätze. Die Selektion des im Rahmen des Programmes verwendeten Ansatzes sollte erst nach Durchführung einer empirischen Wirkungsprognose erfolgen, die alle jene Lösungsansätze überprüft, die theoretisch plausibel und praktisch durchführbar scheinen. Dies kann in einem Vorauswahlverfahren - etwa dem Studium der theoretischen und empirischen Literatur, der Auswertung von Erfahrungen aus ähnlichen Programmen, der Evaluierung des Interventionsmodells - bewertet werden.

Die Wirkungsprognose muß auf der Grundlage des oben beschriebenen Interventionsmodells erfolgen, in dem die geplanten Maßnahmen des Programmes und die ihnen zugeschriebenen Wirkungen dargestellt sind. Häufig kommt dabei die Szenariomethode zur Anwendung, um der Unsicherheit schwer bestimmbarer Einflußfaktoren Rechnung zu tragen. So kann die Profitabilität und damit die Adaptionswahrscheinlichkeit einer bestimmten Technologie, die im Rahmen eines Diffusionsprogrammes gefördert werden soll (z.B. alternative Energieerzeugungstechnologien), von der internationalen Preisentwicklung eines Rohstoffes (z.B. Erdöl) abhängen. Durch Simulation der Programmwirkungen auf der Grundlage verschiedener Preisszenarien können die Auswirkungen dieser Unsicherheit zumindest demonstriert und ihre Relevanz mit Bezug auf den Programmerfolg beurteilt werden.

Der Einsatz mathematisch-ökonometrischer Modelle ist im Rahmen der prognostischen Wirkungsanalyse nützlich, wenn es um die Analyse von Externalitäten oder komplexer interdependenter Zusammenhänge geht. So ist es zum Beispiel durch eine Input-Output Analyse möglich, die komplexe Vernetzung der Wirtschaftssektoren anhand der intersektoralen

Güterströme darzustellen. Die Wirkungen eines branchenspezifischen Programms auf die vor- und nachgelagerten Sektoren können somit beurteilt werden, was ohne quantitatives Modell kaum möglich ist. Mit komplexeren Modellen können zusätzlich die Verbindungen zwischen dem Konsum- und dem Produktionssektor oder auch Preiseffekte (z.B. die Konsequenzen niedrigerer Energiekosten für das Konsumverhalten) analysiert werden.

Auf der Grundlage der für die einzelnen Ansätze geschätzten Wirkungen kann dann, unter Umständen mit Hilfe von entscheidungstheoretischen Verfahren, das optimale Vorgehen bestimmt werden. Die Entscheidungstheorie liefert Kriterien, auf Basis derer die alternativen Ansätze verglichen werden können (z.B. Quest-Methode, Multicriteria Analysis u.a.).

Die Wirkungsprognosen können auch zur Analyse der Interventionslegitimation und damit der Wohlfahrtseffekte herangezogen werden. Dabei muß der gesellschaftliche Nutzen der Programmwirkungen beurteilt und gegen die Kosten des Programms abgewogen werden. Quantitative Modelle sowie die ökonomische Kosten-Nutzen-Analyse werden dabei zur Quantifizierung bzw. Monetarisierung des Nutzens herangezogen, was die Voraussetzung für einen direkten Kostenvergleich ist. Bei der Interpretation der Ergebnisse der Kosten-Nutzen-Analyse muß berücksichtigt werden, daß der Nutzen vieler Wirkungen nur schwer in Zahlen bzw. Geldeinheiten ausgedrückt werden kann (z.B. verbesserte soziale Kontakte als Folge neuer Telekommunikationstechnologien).

Prozeßevaluierung (interim)

Einsicht in die organisatorischen Abläufe eines individuellen Programmes gewinnt man leichter mit qualitativen Verfahren, mit denen die Eindrücke und Erfahrungen der am Programm Beteiligten unmittelbar und in allen Details erfaßt und somit die spezifischen Aspekte und Probleme der Implementierung genau analysiert werden können. Ausführliche Gespräche mit Programmmitarbeitern und Befragungen von Programmteilnehmern fallen in diese Kategorie. Fragestellungen inkludieren die administrative und operationale Effizienz und Effektivität, Abweichungen von der Programmplanung und aufgetretene Probleme der Programmimplementierung.

Wirkungsanalyse (ex-ante, interim, ex-post)

Zweck der Wirkungsanalyse ist es, die direkten und indirekten Effekte (Outputs) eines Programmes zu erfassen. Dabei kommen die folgenden Methoden zur Anwendung:

- quantitative Methoden

Zur Erfassung der indirekten, vor allem ökonomischen Programmwirkungen werden oft mathematische und statistische/ökonometrische Modelle verwendet, die das gesamtwirtschaftliche System bzw. Teile davon mathematisch abbilden oder eine statistische Beziehung zwischen Wirkungsindikatoren und Programmindikatoren messen. Zusätzlich zu diesen Modellen bzw. als Grundlage zur ihrer Entwicklung werden auch standardisierte Befragungen, Beobachtungen und Experteninterviews durchgeführt und quantitative Indikatoren konstruiert.

- qualitative Methoden

Auf Basis von nicht-standardisierten Befragungen, Beobachtungen und Interviews von Programmteilnehmern, Personen/Institutionen, auf die das Programm unmittelbar oder mittelbar Wirkung ausgeübt hat, und Experten werden Rückschlüsse auf die Programmwirkungen gezogen.

Wichtig dabei ist die *Feststellung der Kausalität zwischen Effekten und Programmaktionen*. Können bestimmte Beobachtungen, Messungen, Entwicklungen (etwa eine beschleunigte Diffusion einer neuen Technologie) wirklich auf die Programmeinflüsse (also etwa die finanziellen Förderungen) zurückgeführt werden oder sind andere Faktoren im Spiel, die außerhalb des Programms stehen (etwa Preisentwicklungen auf den Rohstoffmärkten, dem Arbeitsmarkt etc.)? Es ergibt sich also die Notwendigkeit der Bildung einer Referenzbasis ("benchmark"), was nichts anderes als die Fortschreibung der Wirklichkeit ohne die Existenz des Programmes ist.

Es gibt zwei grundsätzlich verschiedene Ansätze zur Erlangung einer Referenzbasis.

Kontrollgruppenkonzept

Das Kontrollgruppenkonzept ist eine in den Naturwissenschaften und bestimmten Sozialwissenschaften (z.B. Psychologie) besonders gebräuchliche Methode, die den systematischen Vergleich von Programmteilnehmern und Nichtteilnehmern beinhaltet und die beobachteten Differenzen beider Gruppen bezüglich ausgewählter Indikatoren als Programmeffekte definiert. Die Beobachtungen sollten unter kontrollierten „Laborbedingungen“ stattfinden, d.h. beide Gruppen sollten abgesehen von den Programmeinflüssen denselben Bedingungen ausgesetzt werden und zudem homogen sein, das heißt die gleichen Merkmale aufweisen.

Solche Bedingungen entsprechen jedoch nicht der Realität der meisten sozialwissenschaftlichen Programme, bei denen externe Faktoren nicht ausgeschlossen und

die beobachteten Effekte deshalb nicht eindeutig auf das Programm zurückgeführt werden können; Personen, Unternehmen etc., die freiwillig nicht am Programm teilnehmen, weisen außerdem andere Merkmale auf als Programmteilnehmer, weshalb die Homogenität der Gruppen nicht gegeben ist.

In der Praxis unternehmensbezogener Programme hieße das zum Beispiel, daß am Programm teilnehmende und nichtteilnehmende Unternehmen in zwei Gruppen geteilt werden und die Entwicklung bestimmter Indikatoren (z.B. Produkt-/Prozeßinnovationen pro Unternehmen) aufgezeichnet werden. Dabei sollte sichergestellt sein, daß Unterschiede in der durchschnittlichen Innovationsrate der beiden Gruppen auf das Programm zurückzuführen sind statt etwa auf einen unterschiedlichen Branchen- oder Unternehmensgrößemix der beiden Gruppen. Eine Voraussetzung für die Vergleichbarkeit der Gruppen ist also, sie bezüglich bestimmter Merkmale zu homogenisieren. Diese Merkmale können Branche und Geschäftsfeld, Unternehmensgröße, regionaler Standort, Alter des Unternehmens etc. sein. Allein die Tatsache, daß ein Unternehmen am Programm teilgenommen hat, während sich ein anderes gegen die Teilnahme entschied, läßt jedoch auf wesentliche Unterschiede in bestimmten Merkmalen (z.B. der Risikobereitschaft) schließen, die der Homogenitätsbedingung widersprechen.

Vorher-Nachher Vergleich

Im Rahmen eines Vorher-Nachher Vergleichs werden, wie schon im Namen zum Ausdruck kommt, die Werte bestimmter Indikatoren vor Beginn des Programms mit den Werten dieser Indikatoren nach Programmende bzw. einige Zeit nach Programmbeginn verglichen und aus den festgestellten Unterschieden Rückschlüsse auf die Programmwirkungen gemacht. Man unterscheidet den:

Ist-Ist Vergleich:

Die Differenz zwischen den Indikatorenwerten vor und nach Programmbeginn wird als Programmeffekt betrachtet. Dieser Ansatz nimmt somit das völlige Fehlen jeglicher anderer Einflußfaktoren an, wodurch die Programmwirkungen tendenziell überschätzt werden. Dient etwa die Innovationsrate als Erfolgsindikator eines branchenspezifischen Innovationsförderungsprogrammes, wird jeder Anstieg dieser Rate unter den teilnehmenden Betrieben als Programmwirkung interpretiert, obgleich sich während der Laufzeit des Programms der weltweite Konkurrenzdruck in dieser Branche erhöht hat und zumindest teilweise für die zusätzlichen Innovationsaktivitäten verantwortlich sein könnte.

Ist-Status-quo Vergleich:

Die hypothetischen Entwicklung der Indikatoren ohne Programm wird simuliert und mit der tatsächlichen Entwicklung verglichen. Die Differenz der hypothetischen mit den tatsächlichen Indikatorenwerten wird wiederum als Programmwirkung interpretiert. Diese Methode versucht also, den Einfluß programmexterner Faktoren auf die Indikatoren zu berücksichtigen. Der Einfluß zunehmenden Konkurrenzdrucks auf die Innovationsrate, wie in dem obigen Beispiel, könnte bei der Berechnung der hypothetischen Rate miteinbezogen werden. Es bleibt jedoch die Unsicherheit, ob dieser Einfluß auch richtig eingeschätzt wurde, was unter anderem von der Qualität des Modells und der vorhandenen Daten abhängt.

Soll-Ist Vergleich:

Hier wird die tatsächlich gemessene Veränderung des Indikators mit der geplanten verglichen, woraus die Effektivität des Programmes ersehen werden kann.

Effektivität (interim, ex-post)

Die Effektivität eines Programmes wird daran gemessen, ob das Programm seine Ziele bzw. seinen Zweck erreicht. Dabei ist zu berücksichtigen:

- Die spezifischen Programmziele sind die Vergleichsgrundlage ("benchmark"), an welcher der Programmoutput gemessen werden kann;
- Beim Vorhandensein mehrerer Ziele müssen diese gewichtet werden, um zu einer globalen Zielabweichung zu kommen;
- Fehlen spezifische Ziele, kann ein Indikator für den Programmoutput in Relation zu anderen Variablen gesetzt werden (z.B. Zahl der geförderten Unternehmen in Relation zur Gesamtzahl aller Unternehmen, die am Programm teilnehmen konnten) und diese Relation mit den identischen Relationen ähnlicher (etwa internationaler) Programme verglichen werden; allerdings müssen die auf jeden Fall vorhandenen Unterschiede in den Programmstrukturen, Rahmenbedingungen etc. berücksichtigt und gute Indikatoren gefunden werden
- Der Effektivitätsgrad mißt das Ausmaß der Zielerreichung bzw. der Zielabweichung und wird wie folgt definiert:

Effektivitätsgrad = erreichter Output / Zieloutput, oder

Effektivitätsgrad = erreichter relativer Output / erreichter relativer Output des Vergleichsprogramms;

- Die *Kausalität* ist, wie schon öfter angesprochen, *in diesem Zusammenhang kritisch*. Es muß gründlich hinterfragt werden, ob der erreichte Output auf das Programm zurückgeführt werden kann oder auch ohne jenes zustande gekommen wäre.

Effizienz (interim, ex-post)

Bei der Effizienzbestimmung werden Programminputs, also jene Mittel und Ressourcen, die in das Programm einfließen, und Programmoutputs verglichen.

Die Inputs bestehen aus:

- administrativen und operativen Kosten
- Mitteleinsatz für Förderungen.

Auch die Opportunitätskosten eines Programmes sollten miteinbezogen werden. Dies sind jene (hypothetischen) Kosten, die entstehen, da die Mittel zur Finanzierung eines Programms für alternative Verwendungen (alternative FTE-Programme, Programme in anderen öffentlichen Bereichen, Transferleistungen etc.) nicht mehr zur Verfügung stehen.

Die Outputs sind die direkten und indirekten, real oder monetär bewerteten Wirkungen, die sich aus dem Programm ergeben. Diese Effekte können positiv oder negativ sein: Der positive Effekt eines Diffusionsprogrammes für flexible, computergesteuerte Produktion, als ein Beispiel, liegt in der verstärkten Anwendung dieser Technologie und den damit verbundenen Wettbewerbsvorteilen, während als negativer Effekt Beschäftigungsverluste eintreten könnten. Negative Wirkungen müssen bei der Berechnungen der Kosten-Nutzen-Relation als Kosten behandelt werden und sind damit Inputs gleichgestellt.

Um zu einer nützlichen Definition der Programmeffizienz zu gelangen, muß als erstes die Frage geklärt werden, ob nur geplante oder auch ungeplante Outputs (also unerwünschte Nebenwirkungen des Programmes) bei der Effizienzbestimmung miteinbezogen werden.

Im Rahmen dieser Studie wird als **Effizienz im engeren Sinn** jene Effizienz verstanden, die nur geplante Outputs einbezieht; bei der **Gesamteffizienz des Programms** (=Effizienz im weiteren Sinn) hingegen werden die gesamten Outputs einbezogen.

Effizienz kann wie folgt definiert werden:

- a) Minimierung der Mittel zur Erreichung eines vorgegebenen Ziels.
- b) Maximierung des Programmoutputs bei vorgegebenem Mitteleinsatz
- c) Maximierung des Output-Input-Verhältnisses.

Zusätzlich soll hier eine Unterscheidung in interne und externe Effizienz getroffen werden:

Die **interne Effizienz** (administrative und operative Effizienz) ist die Effizienz der Umsetzung des vorgegebenen Programmkonzepts/-designs;

Die **externe Effizienz** (strategische Effizienz) beinhaltet den Vergleich der Programmeffizienz (einschließlich der Effizienz des Konzepts und Designs) mit der Effizienz alternativer Lösungsansätze.

Zur Definition (a) ist zu sagen:

- Sie setzt operationale Ziele voraus, die oft nicht vorhanden sind;
- ist für die Effizienz im engeren und weiteren Sinn geeignet, da negative Outputs als Mittel der Zielerreichung angesehen werden können;
- ist auch nur unter experimentellen Bedingungen brauchbar, die allerdings in der Programmrealität kaum gegeben sind;
- und setzt die Zielerreichung voraus.

Da die Formulierung operationaler Ziele oft große Schwierigkeiten aufwirft und die Zielerreichung nicht Voraussetzung der Effizienzmessung sein sollte, sondern unabhängig davon zu untersuchen ist, erscheint dieser Begriff für die Ermittlung der Programmeffizienz wenig geeignet.

Zur Definition (b) ist zu sagen:

- Sie setzt ein fixes Programmbudget voraus;
- erlaubt variable Ziele (z.B. maximal erreichbare Diffusion);
- ist für die interne und externe Effizienz geeignet;
- Kostensenkungen sind jedoch ausgeschlossen.

Durch den Ausschluß von Kostensenkungen ist auch diese Definition für die Ermittlung der Programmeffizienz nur bedingt bzw. nicht für alle Programme geeignet.

Zur Definition (c) ist zu sagen:

- Sie erlaubt variable Ziele und variable Inputs;
- und ist sowohl für die interne als auch externe Effizienz geeignet.

Diese Definition ist durch ihre Flexibilität für die Ermittlung der Programmeffizienz gut geeignet.

Für die Programmevaluierungspraxis bedeutet dies, daß ein Programm dann als effizient betrachtet werden kann, wenn maximale (positive) Wirkungen mit minimalem Mitteleinsatz erreicht werden. *Der Evaluator sollte somit versuchen, Potentiale der Wirkungsverbesserung aufzuzeigen und gleichzeitig auf Kosteneinsparungspotentiale hinzuweisen.* In vielen Fällen wird es dabei zu Konflikten kommen, wenn zusätzliche Mittel nur dann eingespart werden können, wenn auch Wirkungseinschränkungen hingenommen werden bzw. wenn die Programmwirkungen nur durch zusätzlichen Mitteleinsatz verbessert werden können. In diesen Fällen müssen die Kosten und Nutzen gegeneinander abgewogen werden.

Methoden der Effizienzmessung

Die Bestimmung der Effizienz eines Programmes beinhaltet also die Erfassung und Bewertung von Inputs und Outputs. Im Rahmen der Inputbewertung wird vor allem die Schätzung der Opportunitätskosten Schwierigkeiten bereiten, da diese Kosten nur eine hypothetische Größe darstellen.

Die Programmoutputs werden im Rahmen der Wirkungsanalyse erfaßt. Für ihre Bewertung stehen mehrere Methoden zur Verfügung:

Kosten-Nutzen Analyse:

Die KNA beinhaltet eine monetäre Bewertung des aus dem Programm resultierenden Nutzens (z.B. der Wert der Verwendung neuer Umwelttechnologien in einer bestimmten Anzahl von Betrieben) sowie die Umrechnung der Kosten und Nutzen auf einen einheitlichen Bezugszeitpunkt (Diskontierung), da die Programmkosten in der Regel früher anfallen als der Programmnutzen, der dann abdiskontiert werden muß.

Kosten-Wirkung Analyse:

Die KWA vermeidet die Monetarisierung des Programmnutzens, die oft große Schwierigkeiten verursacht, sondern versucht, diesen Nutzen auf andere Arten zu beschreiben oder zu bewerten.

In einem Programm zur Verringerung betrieblicher Emissionen beispielsweise erfordert die KNA eine monetäre Bewertung der nicht emittierten Schadstoffe (z.B. 1 Gramm Blei, das nicht in den Nahrungskreislauf gelangt, ist x Schillinge wert), während die KWA als Nutzen einfach die nicht emittierte Schadstoffmenge definiert (1 Gramm Blei). Allgemein kann gesagt werden, daß die monetäre Bewertung von Programmoutputs vor allem dann problematisch ist, wenn die Outputs

sich aus Gütern ohne Marktpreisen zusammensetzen (z.B. saubere Luft). Die nicht-monetäre Bewertung hingegen erschwert die Vergleichbarkeit von Kosten und Nutzen.

Auch die Wahl einer geeigneten Diskontrate ist nicht eindeutig, sondern beruht auf Annahmen über zeitliche Präferenzen. Dabei gilt es festzustellen, um wieviel ein Schilling in seinem Wert sinkt, wenn er nicht heute, sondern morgen oder nächstes Jahr ausbezahlt wird.

2 Spezifika von FTE Programmen und Vergleich mit strukturpolitischen Programmen

Die in Abschnitt 1 gezeigten Evaluierungsschemata zeigen eine Trennung in allgemeine Evaluierungsstandards bzw. Richtlinien und programmtypenspezifische Richtlinien. In letztere fließen die Besonderheiten von FTE-Programmen ein. Diese Besonderheiten betreffen in der Regel alle Evaluierungsinhalte, also Konzept-, Design- und Prozeßevaluierung, vor allem aber die Wirkungsevaluierung. Wirkungen sind bei FTE-Programmen besonders schwer abzuschätzen, vor allem weil indirekte Effekte auf verschiedenen Ebenen eine beträchtliche Rolle spielen. Diese zu identifizieren und zu messen bereitet ein Problem.

Der folgende Abschnitt zeigt die damit verbundenen Probleme, die vor allem aufgrund der schwierigen Abschätzung sozialer Erträge entstehen. Der anschließende Vergleich mit der Evaluierungspraxis der gemeinschaftlichen Strukturpolitik soll aufzeigen, welche Vor- und Nachteile aus genauen, strikten und verbindlichen Regelungen sowohl hinsichtlich der Programmerstellung als auch hinsichtlich Durchführung, Bewertung und Begleitung entstehen und wie aufgrund dieser Erfahrungen in der FTE-Politik sinnvolle Weiterentwicklungen aussehen können.

2.1 Die Evaluierung von FTE-Programmen

Die Evaluierung von FTE-Programmen wurde mit der wachsenden Bedeutung und Differenziertheit von Technologiepolitik zu einem wichtigen Teil der technologiepolitischen Diskussion, da das Interesse an der Erforschung direkter und indirekter Wirkungen von Programmen aber auch an einer systematischen Bewertung immer größer wurde. Die öffentlichen Entscheidungsträger fragen sich - und werden gefragt - wie die kostenaufwendigen und folgenschweren technologiepolitischen Interventionen angesichts der Zielsetzung eines schlankeren und effizienteren Staates zu legitimieren sind. In der Tat gibt es einen Zusammenhang zwischen der Haushaltslage eines politisch-administrativen Systems und seiner Evaluierungspraxis: in Zeiten der Budgetkonsolidierung werden besonders differenzierte und strenge Praktiken entwickelt.

Neben den unten aufgeführten Bewertungsproblemen, welche die Erstellung von Evaluierungen mit allen Dimensionen erschweren, gibt es auch zwei grundlegende theoretische Probleme:

- Erstens fehlt es an einer umfassenden, allgemein akzeptierten **Theorie** des technischen Wandels einschließlich der Bestimmung der Rolle staatlichen Handelns;
- Zweitens werden für die Ableitung staatlicher Interventionen eine ganze Reihe von Annahmen über die Entstehung und Verbreiterung von neuen Techniken, über das Innovationsverhalten von Firmen, über Wirkungsketten verwendet, für die **es keine empirische gesicherte Basis** gibt. Es handelt sich in der Regel um reine Plausibilitätsüberlegungen.

2.1.1 Abschätzung von Programm-Wirkungen

Wirkungen sind bei FTE-Programmen besonders schwer abzuschätzen, vor allem weil indirekte Effekte auf verschiedenen Ebenen eine beträchtliche Rolle spielen, welche zu identifizieren und zu messen ein Problem bereitet. Vor allem aufgrund dieser Besonderheiten werden in der Regel bei FTE-Programmen auch spezifische Methoden zum Tragen kommen, die eine adäquate Berücksichtigung indirekter Effekte erlauben.

Dabei erschweren gerade jene Elemente die Bewertung der Programme, die letztlich die Legitimation und Motivation der Intervention sind: Die hohe Unsicherheit bei forschungsintensiven Projekten, Externalitäten, öffentliche-Gut-Probleme, Lerneffekte etc..

Das zentrale Problem bei der Abschätzung von Programmwirkungen ist die Ermittlung der aus den FTE-Projekten resultierenden **sozialen Erträge**:

- **Externe Effekte von F&E ("F&E Spillovers")** lassen private und soziale Erträge auseinanderklaffen. Eine Kalkulation des Ertrags der Projekte allein auf Grundlage der Angaben involvierter Firmen und Einrichtungen greift daher - selbst wenn diese ihre privaten Erträge exakt angeben können - zu kurz. Erschwerend kommt hinzu, daß die externen Erträge an ganz verschiedenen Stellen der Volkswirtschaft (oder sogar im Ausland) auftreten können (Hutschenreiter, 1995).
- **Forschungskooperationen** können zu einer Internalisierung dieser externen Erträge führen, sind aber häufig mit anderen Formen des Marktversagens verbunden (Geroski, 1995)
- Neben den - intendierten - positiven externen Effekten können auch **nicht-intendierte Wirkungen** auftreten, die den sozialen Ertrag mindern: Auswirkungen auf den Wettbewerb, auf Arbeitsbedingungen, auf die Umwelt können mögliche nicht-intendierte Effekte sein, die wiederum schwer faßbar sind.
- Hohe **Unsicherheiten** vor allem bei forschungsintensiven Projekten führen zu ex-ante schwer abschätzbaren Erträgen. Eine Festlegung genauer Ziele, gewünschter Wirkungen und vermuteter Erträge bereits beim Entwurf des Programms ist daher nahezu unmöglich.

Andererseits führt der völlige Verzicht auf diese Dimension auch auf den Verzicht einer echten Kontrolle der Zielerreichung des Programms.

- FTE-Programme zeichnen sich durch sehr unterschiedliche und teilweise sehr lange **Zeithorizonte** der Wirkungen aus. Das erschwert die Festlegung des optimalen Zeitpunkts der Ex-post-Evaluierung. Eine Evaluierung direkt nach Programmende wird für die Design- und Prozeßevaluierung vorzuziehen sein, für die Wirkungsevaluierung hingegen deutlich zu früh. Wählt man hingegen einen viel späteren Zeitpunkt, so sind zwar Wirkungen, die erst nach sehr langer Zeit auftreten vorhanden, aber von anderen Einflüssen nur schwer zu trennen.
- Unteilbarkeiten und "**Netzwerk-Externalitäten**" können zu Diskontinuitäten bei der Programmwirkung führen. Besonders deutlich zeigen sich diese Effekte bei Informationstechnologien oder Telematik-Applikationen. Ist z.B. die Verwendung von EDI gering verbreitet, so ergeben sich nur geringe Produktivitätsgewinne oder Effizienzverbesserungen beim einzelnen Nutzer; diese treten erst auf, wenn ein großer Teil der Partner eines Unternehmens auch EDI anwendet. Entsprechend werden die Wirkungen eines Programmes zur Unterstützung der Einführung solcher Technologien erst ab einer gewissen Größe und Verbreitung des Programmes spürbar, ist diese Mindestgröße einmal erreicht, sind kumulative Effekte zu erwarten.
- Mit FTE-Programmen sind darüber hinaus zahlreiche **Lerneffekte** verbunden. Diese können sich in der leichteren Abwicklung von Folgeprojekten, in der Erhöhung der Arbeitsproduktivität oder anderen Effizienzsteigerungen z.B. aufgrund verbesserter strategischer Kompetenz äußern, wobei die Zurechnung solcher Effekte auf konkrete Programme naturgemäß schwerfällt. Schließlich sind solche Effekte wie die Erhöhung des allgemeinen Wissensstands oder der Problemlösungskapazität überhaupt nicht mehr meßbar, können aber sehr wohl wichtige intendierte Effekte von FTE-Programmen sein.

Nur von wenigen Maßnahmen gibt es Anhaltspunkte dafür, wie sie wirken, über die Wechselwirkung verschiedener Maßnahmen ist nahezu nichts bekannt. Je mehr die Komplexität der Zielsetzungen von FTE-Politik zunimmt, desto schwieriger wird es auch, ihre tatsächlichen Wirkungen, Wechselwirkungen und ihren Einzel- und Gesamtnutzen zu erkennen.

2.1.2 Zieldefinition und Zielsetzung von FTE-Programmen

Schwierig ist auch die Zieldefinition von FTE-Programmen. Wie bereits erwähnt, verbessert eine klare Zieldefinition des Programms die Aussagekraft von Evaluierungen erheblich, da Effektivität - in Sinn der Zielerreichungskontrolle - und Effizienz beurteilt werden können.

Klare und operationable Zielsetzungen finden sich selten, obwohl die Auswahl der geeigneten Instrumente wie auch Programmabwicklung und -design in enger Abstimmung mit den Programmzielen erfolgen sollten. *Doch auch - oder gerade - hinter nicht ausgesprochenen Zielen befinden sich bestimmte Erwartungen hinsichtlich der Wirkung des Programms*, diese Erwartungen sind oft vieldimensional und komplex. Es geht um die Erhöhung der Wettbewerbsfähigkeit, um die Sicherung hochwertiger Arbeitsplätze, um die Sicherung der Umweltqualität und um ähnliche hoch-komplexe Themata.

Bei spezifischen Schwerpunktprogrammen finden sich allerdings auch sehr enge und operationable Zielsetzungen wie z.B. die Verdopplung des Exportanteils der Halbleiterindustrie oder ähnliches.

Diese verschiedenen, *unterschiedlich ambitionierten und unterschiedlich expliziten Zielsetzungen* erschweren die Wirkungsanalysen der Evaluierungen. Was z.B. erwartet man sich von einer Steigerung der FTE-Quote? Welche Bewertungen gehen damit einher?

Diese zentralen Fragestellungen werden meist sowohl auf der Ebene der Programmerstellung als auch auf der Ebene der Evaluierung ausgespart. Vielmehr konzentrieren sich die meisten Studien auf Fragen der Programmabwicklung, auf Design und Management des Programms und die Mittelallokation auf einzelne Projekte. Fragen nach der Wirksamkeit oder der Effizienz der Programme werden oft ausgespart oder mit rein deskriptiven Indikatoren illustriert, die keine wirklichen Schlüsse auf die direkten und indirekten Effekte des Programms zulassen.

Aus diesem Bild läßt sich erkennen, daß eine Verbesserung von Verbindlichkeit und Methodik von Evaluierungen nicht ausreicht, um ein wirklich handlungsweisendes Zusammenspiel von Programmerstellung, Programmabwicklung und Programmbewertung zu erhalten. Daneben sollte

- eine *Diskussion über adäquate Zielsetzungen von Programmen* geführt werden und
- neue *theoretische Konzepte* sowie
- neue Ansätze der *Messung vor allem ökonomischer Effekte* von Programmen berücksichtigt bzw. entwickelt werden.

2.2 Vergleich mit Strukturfondsevaluierung

Die Gegenüberstellung der bisher behandelten Evaluierung von FTE-Programmen mit der Evaluierungspraxis regionaler Strukturpolitikprogramme erfolgt an dieser Stelle vor allem aus

einer Motivation heraus: Bei der gemeinschaftlichen Strukturpolitik handelt es sich um einen Politikbereich, der sowohl hinsichtlich Programmerstellung als auch hinsichtlich Durchführung, Bewertung und Begleitung strikten und verbindlichen Regelungen unterliegt. Vor- und Nachteile dieser Praxis lassen sich klar herausarbeiten und daher Überlegungen anstellen, wie in Politikbereichen mit einer bisher vergleichsweise geringer entwickelten Evaluierungspraxis - zum Beispiel eben FTE-Politik - sinnvolle Weiterentwicklungen zu unterstützen sind.

2.2.1 Die Regelungen von seiten der Europäischen Union

Regelung und Praxis der Evaluierung entwickelte sich seit den 80er Jahren ganz parallel zu den großen Zäsuren und Reformen der regionalen Strukturpolitik. Vor 1988 gab es nur sporadische und unvollständige Kontrollen, wofür und in welcher Weise Gelder verwendet wurden, auch mangelte es den meisten Mitgliedsländern an Evaluierungs-know-how und -erfahrung. Mit der Strukturfonds-Reform 1988 gab es zunächst ein deutliches Mehr an Evaluierungen ohne spürbare Verbesserungen in Qualität, Vergleichbarkeit und Effizienz. Entsprechend kritisch äußerten sich auch verschiedene Institutionen der Gemeinschaft über enorme Bürokratie, exzessiven Detaillierungsgrad und Inflexibilitäten.

1993 gab es mit der grundlegenden Reform und geänderten Mittelallokation einige entscheidende Änderungen, sowohl Verbindlichkeit als auch Funktion und Qualität der Evaluierung betreffend. Die entsprechenden Verordnungen sehen vor:

- Eine klare Unterscheidung zwischen Ex-ante-Evaluierung, Begleitung und Ex-post-Bewertung (appraisal, monitoring, evaluation)
- Die verpflichtende Durchführung der Evaluierungen in gemeinsamer Verantwortung von Kommission und Mitgliedsländern.
- Die Verpflichtung der Kommission, das Parlament und die Ausschüsse in regelmäßigen Abständen über die Durchführung der Aktionen zu unterrichten.

Die Rahmenbedingungen für Beurteilung und Bewertung (appraisal und evaluation) sind (Artikel 26 der Strukturfonds-Verordnung 2082/93):

- *"Appraisal and evaluation was to be carried out by contrasting the goals with the results obtained ... and by reference to macro-economic and sectoral objectives and indicators based on regional or national statistics, to information yielded by descriptive and analytical studies and to qualitative analyses"*

- “*Appraisal and evaluation shall take account of the socio-economic benefits expected or achieved commensurate with the resources deployed, conformity with Community policy and rules ... and the conditions in which the measures are implemented.*”

Für **Begleitung und Zwischenbewertung** (Monitoring) legte die Kommission für die Programmplanungsperiode 1994-1999 gemeinsame Leitlinien fest:

- Die Begleitung erfolgt durch die zu diesem Zweck eingerichteten Begleitausschüsse (in Österreich mit Vertretern der Kommission, der Bundesministerien und der entsprechenden Bundesländer koordiniert von der ÖROK). Die mit der Durchführung der Strukturfondsinterventionen betrauten Stellen erstatten den Begleitausschüssen regelmäßig Bericht, diese wiederum legen den entsprechenden Stellen der Mitgliedsstaaten und der Kommission Jahresberichte sowie einen Schlußbericht.
- Die Begleitung umfaßt erstens die operationelle Begleitung (Monitoring im engeren Sinn), welche sich gliedert in:
 - Begleitung der Ausführung, der Ergebnisse und der Auswirkungen von Programmen (möglichst in quantitativer Form mit Hilfe von Indikatoren zum Vollzug, zur Wirksamkeit und zur Effizienz)
 - Begleitung der verwaltungsmäßigen Abläufe (wie z.B. die Arbeitsweise der Verwaltungs- und Finanzstrukturen)
 - und die Begleitung des operationellen Kontexts (wie z.B. die Kohärenz);
- zweitens die Zwischenbewertungen, welche den Begleitmechanismus verstärken und insbesondere die vom Begleitsystem gelieferten Informationen analysieren und systematisieren soll;
- drittens die Begleitung des globalen Kontexts, welche eine regelmäßige Analyse der Entwicklung des wirtschaftlichen und sozialen Umfelds (konkretisiert nur für Ziel 1 Gebiete) sowie die Begleitung des Grundsatzes der Additionalität umfaßt.

Dies bedeutet eine relativ genaue Festlegung des Ablaufs und der institutionellen Verankerung von Evaluierungen. Dies betrifft insbesondere das Monitoring-System. Festgelegt sind weder die genauen Inhalte der Evaluierungen - obwohl es dort von Seiten der Kommission bestimmte Erwartungen und Qualitätsvorstellungen gibt, die dann in den jeweiligen Ausschreibungen konkretisiert werden - noch die *Methodik*.

Allerdings gibt es in den hochformalisierten Programmplanungsdokumenten bereits eine Festlegung der Evaluierungsinhalte sowie der relevanten Indikatoren. Diese Festlegung

wiederum spiegelt sich in der Vorausbeurteilung (appraisal) wider. Sowohl die Programmplanungsdokumente selbst als auch die Vorausbeurteilung befolgen im wesentlichen die ROAME technique (Rationale, Objectives, Appraisal, Monitoring, Evaluation).

2.2.2 Die Funktionen der Evaluierung

Zentrale Motivation für die Neugestaltung der Evaluierungen war die mit den beiden Reformen 1988 und 1993 eingegangene Verpflichtung, auf Grundlage umfassender Evaluierungen über die Effektivität wie auch über die Wirkung der Strukturfonds-Interventionen Bericht zu legen. Damit steht die **Legitimationsfunktion** - gerade auch angesichts der beträchtlichen Mittel, welche in regionale Strukturpolitik fließen - im Vordergrund.

Auch nehmen die Evaluierungen eine gewisse **Informationsfunktion** ein, da die Kommission weder über Detailkenntnis bezüglich der regionalen Ausgangssituation und Entwicklung noch bezüglich einzelner Instrumente und Politiken verfügt.

Der hohe Stellenwert, der dem gesamten Evaluierungssystem zukommt, ergibt sich allerdings aus der nicht unbeträchtlichen **Lenkungsfunktion**. Zwar ist die Mittelvergabe bei den Zielgebietsprogrammen nicht unmittelbar an Evaluierungsergebnisse geknüpft, wohl aber die Allokation der Mittel auf einzelne Ziele, Prioritäten und Maßnahmen. Das Design des Monitoring-Systems erlaubt auch Umgewichtungen während der laufenden Programmplanungsperiode. Diese möglichen Feinanpassungen werden als wichtiges Element des gesamten Evaluierungssystems angesehen.

2.2.3 Qualität und aktuelle Praxis der Evaluierungen

Mit der neuen Regelung 1993 wurden auch Maßnahmen ergriffen, die Qualität der Evaluierungen zu erhöhen:

- Verbesserung der den Evaluatoren zur Verfügung stehenden Daten (EUROSTAT wie auch nationale Erhebungen detaillierter sozio-ökonomischer Daten auf regionaler Ebene).
- Gründung einer koordinierenden Evaluierungs-Einheit in der DG XXII (später DG XVI), welche Management-Verantwortung für Ex-ante, Interim und Ex-post Evaluierungen in Ziel 1 und 2 Regionen trägt.
- Verbesserung der Evaluierungsmethodologie: Gründung des MEANS Programms (Methods for Evaluating Actions of Structural Nature), beim Centre for European Evaluation Expertise

(C3E) in Lyon. Aufgabe des MEANS Programms ist die Entwicklung einer "Europäischen Evaluierungskultur".

Diese Bemühungen waren insofern erfolgreich, als daß auf seiten der Evaluatoren das Wissen um bestimmte Techniken und ihre Anwendbarkeit wie auch der Erfahrungsaustausch untereinander enorm anwuchs, auf seiten der Programmverantwortlichen wiederum eine kritische und avancierte Nachfrage entstand.

Trotz dieser spürbaren Verbesserungen liegt methodisch noch einiges im argen. Obwohl einige methodisch anspruchsvolle modell-basierte Studien existieren, enthalten die meisten Evaluierungen einer Kombination aus:

- groben Makro-Abschätzungen mit einer "before-and-after"- oder "with-and-without"-Abschätzung, numerischen Analysen mit einer nur unzureichenden Abschätzung der Interventions-Wirkungen.
- einer ad-hoc Interpretation von Mikro-Indikatoren, die bei Interim- bzw. ex-post-Evaluierungen dem Monitoring-System entnommen sind. Dabei kommen sowohl quantitative als auch qualitative Informationen zum Tragen, die in Interviews, Expertenbefragungen und Fallstudien gewonnen werden.

2.2.4 Probleme der aktuellen Praxis

Weshalb diese Diskrepanz zwischen einer anspruchsvollen methodischen Diskussion in Kombination mit sich entwickelnden Evaluierungskulturen und einer gängigen Evaluierungspraxis, nach der echte Wirkungsanalysen kaum durchgeführt werden sondern in schematischer Weise Arbeitsplätze zusammengezählt, Investitionen aufaddiert und Quadratmeter erschlossener Flächen aufgelistet werden? Häufig handelt es sich um beschreibende ad-hoc-Ansätze, die weder zur Effizienzkontrolle noch zur Wirkungskontrolle wirklich brauchbar sind, sondern vor allem den Vollzug der Intervention beschreiben und bewerten.

Wenn auch die Unzufriedenheit aller Beteiligten mit dieser "alten" Art der Evaluierungserstellung (vgl. z.B. verschiedene Veröffentlichungen des Bremer Ausschusses für Wirtschaftspolitik) wächst, gibt es doch ein Grunddilemma, das qualitative Weiterentwicklungen aktuell eher behindert als unterstützt:

Der Bedarf von seiten der Kommission an einfachen, vergleichbaren Zahlen, welcher sich vor allem aufgrund der Berichtspflicht auch dort ergibt, wo diese seriöserweise nicht oder nur ungenau zu ermitteln sind. Am deutlichsten wird dies bei der Messung von Beschäftigungseffekten:

Es beginnt mit der Pflicht zur Quantifizierung der Ziele in den Programmplanungsdokumenten, was zu genauen Zahlenangaben (Maßnahme x schafft in y Jahren z neue Arbeitsplätze) führt, welche dann eben auch in der Evaluierung überprüft werden müssen. Dabei sind die antizipierten Wirkungen einzelner Programme und Maßnahmen in den Planungsdokumenten in der Regel mehr oder weniger Hausnummern. Angesichts der enormen formalen Erfordernisse an die Dokumente, die meist noch in sehr kurzer Zeit erstellt werden müssen, bleibt selten Zeit für eine seriöse Abschätzung der Wirkungen.

Die Evaluatoren stehen vor dem Dilemma, diese Zahlen überprüfen zu müssen, ohne auf echte Wirkungsindikatoren für die einzelnen Maßnahmen zurückgreifen zu können. Verwendet werden daher in der Regel einfache Output-Indikatoren, die seriöserweise nicht wirklich mit der in Frage stehenden Intervention in Verbindung gebracht werden können (z neue Arbeitsplätze sind in der Region y entstanden, also war der Beschäftigungseffekt von Maßnahme x etwa z Arbeitsplätze). Weshalb sind solche wenig brauchbaren Aussagen dennoch gewünscht? Die Berichtspflicht in Kombination mit der nahezu unbewältigbaren Fülle an Informationen führt zu dem verständlichen Wunsch nach leicht und schnell lesbaren Planungsdokumenten sowie einfach vergleichbaren Evaluierungen. Dazu die Kommission:

“Taken literally, all these requirements would generate hundreds, if not thousands of studies. Given the productive capacity of European professionals and the time that political and administrative leaders have for reading reports, this is an unrealistic ambition” (Results of the 1995 MEANS programme, DG XVI Brussels, p.4).

In einer solchen Situation sind einfach vergleichbare, schematische und katalogisierende Elemente in den Evaluierungen hilfreich, qualitative Analysen über Wirkungszusammenhänge können nur schwer gewürdigt werden. Dies führt zu den oben erwähnten Pseudo-Quantifizierungen, zu unzulässigen Verallgemeinerungen von Fallstudien sowie zu falschen Kausalzurechnungen.

Der eher einfache, stark beschreibende Ansatz hat sich auch deshalb durchgesetzt, da modellbasierte Studien deutlicher mit prinzipiellen Problemen konfrontiert sind, von denen in einfachen ad-hoc Modellen nur wenig zu spüren ist:

- das Problem mangelhafter und unzureichender Daten wiegt bei quantitativen Modellen aufgrund von Interdependenzen doppelt schwer. Für kleine Zielgebietsregionen und insbesondere für Gemeinschaftsinitiativen sind geeignete Daten in der Regel nicht erhältlich.
- die Festlegung des Referenzpfads (was wäre ohne die zu evaluierende Maßnahme passiert?)
- Quantifizierung von Sekundäreffekten (Spill-overs und Externalitäten, Multiplikatoreffekte, Linkage-Effekte, Crowding-out Effekte), die z.T. auf Makro-, z.T. aber auf Meso- oder Mikro-Ebene zum Tragen kommen.
- die Ermittlung des Zusatznutzen weiterer Geldeinheiten im Projekt (zentrale Frage bei Additionalität)
- der unterschiedliche Wirkungszeitraum der Einzelmaßnahmen, die in einem Programm enthalten sind

Zusammenfassend liegt das Grundproblem darin, daß

- angesichts der Höhe der Interventionssummen quantitative Makro-Modelle nur begrenzt einsetzbar sind (am ehesten in großen Ziel 1 Gebieten).
- aufgrund der Erfordernisse der Programmplanungsdokumente (d.h. indirekt aufgrund des Wunschs nach einer einfachen Vergleichbarkeit verschiedener Interventionsformen, den die Kommission vor allem aufgrund ihrer Berichtspflicht hat) quantitative Indikatoren auch dort verlangt werden, wo sie seriöserweise nicht erhoben werden können, was häufig zu Pseudo-Quantifizierungen führt (am deutlichsten bei den Arbeitsplatzwirkungen)
- eine echte Wirkungskontrolle wegen mangelnder Indikatoren oder Modelle oft nicht möglich ist und der Schwerpunkt der Evaluierung daher auf einer - in der Regel schematisch erstellten - Vollzugskontrolle liegt.

Um dieses Problem zu mindern, müßten auf der einen Seite Mechanismen implementiert werden, die es erlauben, auch qualitative Aussagen systematisieren und vergleichen zu können. Auf der anderen Seite müßten quantitative Modelle in der Weise modifiziert werden, daß sie auch Anhaltspunkte über die Wirkungen kleinerer Interventionen geben (beispielsweise über den verstärkten Einsatz mikroanalytischer Modelle, die auf Daten des Monitoring-Systems zurückgreifen). Drittens müßten im Rahmen des Monitorings Mechanismen gefunden werden, nicht nur Output-Indikatoren, sondern auch echte Wirkungsindikatoren zu erheben.

2.2.5 Schlußfolgerungen und Empfehlungen

- Es können folgende verallgemeinerbare Schlußfolgerungen gezogen werden:
- Sinnvoll und auf andere Evaluierungsbereiche übertragbar ist die genaue Festlegung von Regelmäßigkeit, Ablauf und institutioneller Verankerung, nicht aber von Inhalt und Methodik. Aufgrund der hochformalisierten Programmplanungsdokumente ergibt sich allerdings indirekt eine Vorgabe der Inhalte.
- Damit im Zusammenhang steht die enge Interdependenz zwischen Programmerstellung, Umsetzung, Begleitung und Bewertung: Es zeigt sich, daß eine mangelhafte Spezifikation von Ziel, Intention und erwünschten Wirkungen der Programme zu wenig hilfreichen Evaluierungen führen. Gute Evaluierungen benötigen eine avancierte, informierte Nachfrage und umgekehrt.
- Die Implementierung eines programmbegleitenden Monitoring-Systems kann sehr hilfreich sein. Die Mikro-Informationen des Monitoring Systems sollten sinnvoll in Bewertungen einfließen. Dazu bedarf es der Entwicklung einer neuer mikrobasierten Methodik zusätzlich zu der gängigen, rein deskriptiven Interpretation der Indikatoren. Zweitens sollte das Monitoring-System auch darauf abstellen, echte Wirkungsindikatoren erheben zu können.
- Der Einsatz quantitativer Modelle vor allem zur Abschätzung von Sekundäreffekten sollte weiterentwickelt werden. Modellbasierte Aussagen über Programmwirkungen sollen auch dort möglich sein, wo die Interventionssummen im Vergleich zu den Makro-Größen relativ klein sind.
- Die Verpflichtung zur Berichtslegung ist prinzipiell zu begrüßen, da sie zu mehr Verbindlichkeit führt. Allerdings wird am Beispiel der regionalen Strukturpolitik deutlich, daß ein zuviel an Informationspflicht zu methodischen Versteinerungen führen kann. Gefragt ist ein aussagekräftiger Mix an qualitativen und quantitativen Argumenten, die Vergleiche verschiedener Interventionsformen ermöglichen, ohne in starre Pseudo-Quantifizierungen zu führen.

3 Internationale Erfahrungen mit FTE-Programmevaluierungen

3.1 Die Evaluierung des Alvey Programme for Advanced Information Technology

3.1.1 Programmbeschreibung

Nach der Ankündigung Japans, eine fünfte Computergeneration zu entwickeln, beschloß die britische Regierung, ein Programm zur Förderung des international nur bedingt konkurrenzfähigen inländischen Informationstechnologiesektors zu initiieren. Unter dem Titel Alvey wurde schließlich ein fünfjähriges Programm zur Förderung der vorwettbewerblichen, kollaborativen Forschung und Entwicklung im Bereich der Informationstechnologien durchgeführt, in dem zwischen 1983 und 1992 nicht weniger als 200 Mio Pfund an direkten Förderungen und zusätzlich noch 150 Mio an Mitteln der teilnehmenden Industriebetriebe aufgewandt wurden.

Die öffentlichen Gelder für Alvey stammten aus dem Ministry of Defence (MoD), dem Department of Trade and Industry (DTI) und dem Science and Engineering Research Council. Die Förderungen waren mit 50% der industriellen Gesamtkosten limitiert. Als Programmträger wurde eigens für Alvey ein Direktorat gegründet, nur die vertraglichen Abschlüsse mit den Firmen und Universitäten blieben im Verantwortungsbereich der oben erwähnten Institutionen.

Alveys Ziele, die qualitativ definiert und in diesem Sinne nicht operational waren, beinhalteten technische, strukturelle sowie strategische Vorgaben, die in einem hierarchischen Verhältnis zueinander standen.

Die technischen Ziele für die einzelnen Subprogramme und Projekte, die alle in vier a priori selektierten technischen Bereichen lagen, wurden vom Direktorat bzw. den Projektteams festgelegt und bildeten die unterste Stufe der Zielhierarchie. Ihnen übergeordnet waren die strukturellen Ziele, die einerseits eine Verbesserung der Kooperationsbasis zwischen dem universitären und dem industriellen Sektor vorsahen, andererseits die Stärkung der britischen Informationstechnologiebasis beinhalteten. An der Spitze der Zielhierarchie standen die strategischen und im wesentlichen sozio-ökonomischen Ziele: Die Wettbewerbsfähigkeit der britischen Industrie sollte gestärkt sowie die Leistungsfähigkeit in für militärische Zwecke wichtigen Informationstechnologie-Bereichen erhalten werden. Konkretere strategische Ziele umfaßten Marktanteilsgewinne für britische Unternehmen am Markt für Informationstechnologie (IT), die Erhöhung der britischen technologischen Leistungsfähigkeit auf dem Gebiet der IT-

Technologie, die Verbesserung der IT-Handelsbilanz und die Erhöhung des Anteils inländischer IT-Güter unter den britischen Anwendern.

3.1.2 Allgemeine Evaluierungsstruktur

Alvey wurde in zahlreichen begleitenden Einzelevaluierungen sowie einer Ex-post Evaluierung bewertet, was aufgrund der Größe des Fördervolumens angemessen erscheint. Trotzdem wurde auf eine Ex-ante Evaluierung verzichtet, Programmkonzept und -design wurden erst nach Programmbeginn im Rahmen der laufenden Evaluierungen untersucht. Das Evaluierungsteam setzte sich aus Experten der akademischen Institute PREST (Programme of Policy Research in Engineering Science and Technology) sowie SPRU (Science Policy Research Unit) zusammen.

Der erste veröffentlichte Evaluierungsbericht stammt aus dem Jahr 1985, also zwei Jahre nach Programmbeginn. Der abschließende Bericht wurde 1991 veröffentlicht, vier Jahre nach dem offiziellen Programmende. Zu diesem Zeitpunkt flossen jedoch noch immer Alvey-Gelder in IT-Projekte.

Die inhaltliche Dimension der Alvey-Evaluierung umfaßte die Bewertung folgender Punkte:

- Implementierungseffizienz
- Programmwirkungen und -effektivität
- Angemessenheit des Programms zur Erreichung der Ziele und Angemessenheit der Ziele selbst, was im Rahmen der in Teil 1 definierten inhaltlichen Gliederung einer Konzept- und Designevaluierung entspricht.

Zeitlich betrachtet wurde zuerst die Angemessenheit des Programms bewertet, also eine verspätete „Ex-ante“ Evaluierung durchgeführt; danach wurde die Implementierung, dann die Wirkungen und die Effektivität bewertet. Der Evaluierungsprozeß schloß mit einer weiteren Bewertung der Angemessenheit, diesmal auf Grundlage der im Laufe des Programms gewonnenen Erkenntnisse, ab.

Die von den Evaluierungsteams angewandten Methoden waren tendenziell qualitativ orientiert und inkludierten:

- eine Analyse der Programmdokumentation
- Befragungen, Interviews von Forschern, Managern, Administratoren standardisiert und nicht-standardisiert)
- Indikatoren des Projekterfolgs (inkl. Patentdaten)

- „Network“-Analyse (Konominationstechnik, Beobachtung von Sitzungen)
- Analyse internationaler IT-Programme

3.1.3 Spezielle Aspekte der Evaluierung

Zielidentifizierung

Die Evaluatoren waren der Meinung, die offiziellen Programmziele wären unter „Marketinggesichtspunkten“ formuliert worden: Um das Programm für die Öffentlichkeit attraktiver erscheinen zu lassen und somit seine politische Akzeptanz zu erhöhen, wurden die Ziele unrealistisch hochgeschraubt. *Die Evaluatoren sahen es deshalb als ihre Aufgabe an, die „wahren“ Ziele in einem komplexen Prozeß der Zielidentifizierung zu finden. Dadurch sollten es möglich sein, die Programmeffektivität unter realistischeren Voraussetzungen zu beurteilen.*

Diese neuen, realistischeren Ziele wurden aus dem offiziellen Programmvorschlag, Gesprächen mit verschiedenen Programmakteuren und jenen Zielen, die sich im Laufe des Programmes ergaben, entwickelt.

Bewertung der Implementierungseffizienz

Die Bewertung der Programmeffizienz basiert nicht auf einem quantifizierten Kosten-Nutzen Verhältnis, sondern geht von der Feststellung offensichtlicher Effizienzverbesserungs Potentiale aus. Die Evaluierung orientiert sich daher eher an dem in Teil 1 beschriebenen Effizienzbegriff (a), wonach die Effizienz als Minimierung der Mittel zur Erreichung eines vorgegebenen Ziels definiert wird.

Da als Akteure der Programmimplementierung nicht nur das Alvey-Direktorat, sondern auch die partizipierenden Unternehmen und Forschungsinstitutionen betrachtet werden (was der starken Betonung des kooperativen Elements im Programm entspricht), beschränkt sich die Bewertung der Implementierungseffizienz nicht nur auf den eigentlichen Programmträger (das Direktorat), sondern umfaßt auch die Aktivitäten der Programmteilnehmer.

Im Rahmen der Prozeßevaluierung wurden folgende Bereiche analysiert:

Programmadministration

Das Direktorsmodell wurde mit anderen Projektträgerstrukturen verglichen; unter diesem Punkt wurde auch das Ausmaß sowie die Art der Förderung diskutiert. Weiters

wurden die Aufgaben und Tätigkeiten des Direktorats und seine Beziehungen mit den Sponsorinstitutionen bewertet sowie das Verhalten des Direktorats in Krisenzeiten, also bei Änderungen in laufenden Projekten oder bei Nichterreichung der Projektziele.

Programmmanagement

Dieser Teil schloß unter anderem das Finanzmanagement und das Management-Informationssystem mit ein.

Technologische Ausrichtung der Subprogramme

Bewertet wurden die Festlegung strategischer Richtlinien, die Aufgabe des Alvey-Direktorats war; weiters das Projektauswahlverfahren und die Übereinstimmung der selektierten Projekte mit den festgelegten Richtlinien.

Programminternes Monitoring

Zu den Aufgabenbereichen des Direktorats gehörte auch die begleitende Kontrolle aller Aktivitäten, die wiederum von den externen Evaluatoren bewertet wurde.

Kollaborationsprozesse

Die Evaluatoren analysierten auch die Struktur der verschiedenen Kooperationen zwischen Unternehmen und Universitäten.

Evaluierung des Programm- bzw. Projektfortschritts

Auf der Grundlage qualitativer Methoden, das heißt der Auswertung von Berichten laufender und schon abgeschlossener Projekte sowie von Berichten interner Evaluatoren, Befragungen und Interviews mit Projektteilnehmern etc., wurde versucht:

- den Fortschritt jedes Programmes/Projekt es zu evaluieren, das heißt den (technischen) Output qualitativ zu beurteilen;
- jene Faktoren zu diskutieren, die für den Erfolg bzw. Mißerfolg der einzelnen Programme/Projekte verantwortlich waren;
- die Nutzung bzw. Umsetzung der Forschungsergebnisse zu bewerten.

Evaluierung der Programmeffektivität und -effizienz

Die Beurteilung der Zielerreichung und der Effizienz im Sinne eines Kosten-Nutzen-Verhältnisses wurde auf der Basis von Umfragen durchgeführt. Dabei mußten die Programmteilnehmer selbst beurteilen, ob ihre anfänglichen Erwartungen erfüllt wurden und der Programmnutzen die Kosten übertraf.

Die Teilnehmer wurden weiters befragt, welche Bedeutung Alvey für ihre Forschungstätigkeit hatte, welche Qualifikationseffekte sich ergeben haben, ob das Programm die Bildung nationaler und internationaler Forschungsk Kooperationen erleichterte und auch wie die gewonnenen Erkenntnisse diffundierten. Auch ein Vergleich mit internationalen IT-Programmen wurde durchgeführt.

Programmeffekte, die über die Wirkungen auf die Teilnehmer hinausgingen, etwa die gesamtwirtschaftliche Bedeutung eines konkurrenzfähigeren IT-Sektors oder mögliche Spillover-Effekte auf andere Sektoren wurden jedoch nicht explizit berücksichtigt. Bei der Interpretation der Umfrageergebnisse muß auch berücksichtigt werden, daß nur die Einschätzungen der Teilnehmer selbst erhoben wurden. Eine Beurteilung der Bedeutung Alveys durch programmexterne Experten fehlte.

Weiters wurde auch eine systematische Identifikation der erfolgsbeeinflussenden Faktoren durchgeführt, wobei zwischen technischen, programmspezifischen und teilnehmerspezifischen Faktoren unterschieden wurde.

3.1.4 Beurteilung der Alvey-Evaluierung

Mehrere Aspekte der Alvey-Evaluierung fallen besonders auf: Erstens die Tatsache, daß die offiziellen Programmziele nicht als Evaluierungsgrundlage herangezogen wurden, sondern andere, inoffizielle und weniger ambitionöse Ziele. Ob dieses Vorgehen der Evaluatoren im Rahmen der Bewertung der Programmeffektivität richtig ist, bleibt fraglich, da die Finanzierung Alveys auf der Basis der offiziell formulierten Ziele legitimiert wurde und damit jene Ziele der Evaluierungsmaßstab sein sollten.

Zweitens fällt auf, daß die Evaluierung Alveys sehr auf den Erfahrungen der Programmteilnehmer beruht und somit eher den Charakter einer Selbstkritik als den einer unabhängigen Evaluierung hat.

Drittens wurde eine relativ genaue Erfassung der direkten Effekte durchgeführt, während die indirekten, gesamtwirtschaftlichen Effekte jedoch kaum untersucht wurden. Dies, obwohl der Anstoß für die Durchführung eines IT-Programms die Bedrohung gesamtwirtschaftlicher Interessen war.

3.2 Evaluierung indirekter F&E-Förderung: Die Studie von B. Hall (1992)

Prinzipiell gibt es zwei Ansätze zur Quantifizierung der Effekte indirekter F&E-Förderungsinstrumente (Cordes, 1989):

- ökonometrische Studien.
- direktere Evidenz, z.B. Befragungen.

Einen Überblick über eine Reihe empirischer Studien bieten Cordes (1989), Harhoff (1994) und OECD (1995A). Die hier vorgestellte Studie von Hall bedient sich eines ökonometrischen Ansatzes.

Die "klassische" theoretische Begründung für die Förderung privater F&E-Investitionen besteht darin, daß die gesamtwirtschaftlichen ("sozialen") Erträge von F&E die privat aneignbaren Erträge übersteigen. Theoretisch sollte daher jener Förderungssatz bestimmt werden, bei dem die Kluft zwischen der privaten und sozialen Ertragsrate von F&E-Ausgaben geschlossen wird. In der Praxis ist die Bestimmung der Ertragsverläufe kaum möglich. Die einschlägigen ökonometrischen Studien beschränken sich daher in der Regel auf eine *Art Kosten-Nutzen-Analyse, wobei die geschätzten, durch die indirekte Förderung zusätzlich induzierten F&E-Ausgaben den damit verbundenen Steuerausfällen gegenübergestellt werden*. Diese Gegenüberstellung ist zwar aus theoretischer Sicht nicht völlig zufriedenstellend, ist jedoch intuitiv einleuchtend und wohl auch für die Förderungspolitik eine sinnvolle erste Annäherung an das Problem.

Frühere Studien (bis etwa Mitte der achtziger Jahre) beurteilen die Anreizwirkung indirekter F&E-Förderungen eher skeptisch. Typischerweise bewegen sich die in diesen Studien geschätzten Preiselastizität von F&E im Intervall -0,3 bis -0,4. Dies bedeutet daß eine Reduktion des Preises (der Kosten) von F&E eine Erhöhung der F&E-Ausgaben im Ausmaß von 0.3-0.4% zur Folge hat (siehe etwa den Überblick bei Cordes, 1989, Hall, 1992B oder Harhoff, 1994). Einige neuere Untersuchungen gelangen zu deutlich höheren Schätzungen für die Preiselastizität von F&E, darunter auch die Studie von Hall (1992B). Die Untersuchung von Hall ist der bislang wahrscheinlich anspruchsvollste Versuch, die Wirkungen eines Instruments der steuerlichen F&E-Förderung, des "R&D tax credit" in den USA, auf die F&E-Ausgaben mit ökonometrischen Methoden zu quantifizieren.

Der "R&D tax credit" in den USA wurde 1981 (Economic Recovery Tax Act) eingeführt und war ursprünglich bis Ende 1985 befristet. In der zweiten Hälfte der achtziger Jahre und in den neunziger Jahren wurde die Laufzeit mehrfach verlängert, wobei zumeist auch Änderungen in den Bedingungen vorgenommen wurden. Siehe dazu den Überblick bei Hall (1992B). In jedem Fall wurde der "R&D tax credit" ermittelt, indem die F&E-Ausgaben in der jeweils gültigen Definition, die einen bestimmten Basiswert überschreiten, mit dem gültigen Satz ("credit rate") multipliziert wird. Der daraus sich ergebende Betrag wurde von der Körperschaftssteuer ("corporate income tax") abgezogen. Es bestand eine dreijährige Rücktrags- und 15jährige Vortragsmöglichkeit, falls im laufenden Jahr kein steuerbares Einkommen vorhanden war.

Ein erstes bemerkenswertes Faktum stellt der Umstand dar, daß - obwohl sich der Satz für den "R&D tax credit" zumeist in der Höhe von 20-25% bewegte - der effektive Satz viel niedriger lag, nämlich bei rund 4% oder etwas darüber. Dies hatte eine Reihe von Ursachen. Erstens erhöhten zusätzliche F&E-Ausgaben im laufenden Jahr die Basis zur Berechnung der steuerbegünstigten zusätzlichen F&E-Ausgaben in den drei Folgejahren. Zweitens gab es eine Obergrenze für die zusätzlichen F&E-Ausgaben und drittens komplexe Wechselwirkungen zwischen dem Steuersystem der USA und anderen nationalen Steuersystemen.

Den analytischen Rahmen für die Analyse von Hall bildet ein dynamisches Investitionsmodell mit Anpassungskosten. Diese Anpassungskosten werden als proportional zur Intensität der F&E-Investitionen angenommen. Diese werden minimiert, wenn die F&E-Investitionen den Ersatzinvestitionen entsprechen.

Die empirische Grundlage der Untersuchung besteht aus einem großen Sample von Unternehmen des Manufacturing-Bereichs (rund 1.000 pro Jahr) im Zeitraum 1980-1991. Die darin enthaltenen Unternehmen tätigen etwa 85% der industriellen F&E-Ausgaben. Ein - durch Rückgriff auf ältere Schätzungen bzw. per Annahmen behelfsmäßig gelöstes - Problem der Studie besteht darin, daß weder der Anteil der steuerlich qualifizierten F&E-Ausgaben an den gesamten F&E-Ausgaben der Unternehmen bekannt war noch Information über den exakten steuerlichen Status der Unternehmen verfügbar war.

Auf Basis ökonometrischer Schätzungen des zugrundeliegenden Modelle kommt Hall zu dem *Ergebnis, daß die Preiselastizität der F&E-Ausgaben in der Nähe von 1 liegt*. Diese - im Vergleich zu früheren Studien - relativ starke Reaktion der F&E-Ausgaben auf Preisänderungen hat zur Folge, daß nach Halls Schätzungen die durch den "R&D tax credit" zusätzlich induzierten F&E-Ausgaben höher sind als die damit verbundenen Einnahmehausfälle. Weiters

legen Halls Ergebnisse den Schluß nahe, daß die (allerdings kaum präzise schätzbaren) langfristigen Effekte des "tax credit" groß sind. Aus der verzögerten Anpassung kann umgekehrt die Empfehlung abgeleitet werden, mit kurzfristigen Adjustierungen bzw. kurzfristigen Laufzeiten der steuerlichen Förderung - im Gegensatz zur tatsächlichen Geschichte des "R&D tax credit" in den USA - vorsichtig umzugehen. Auf die Risiken der Fehlspezifikation der Referenzgröße für "inkrementelle" Förderungen wurde bereits hingewiesen.

4 Evaluierung von Technologieprogrammen in Österreich

4.1 Einleitung: Evaluierungen im Gesamtsystem der FTE-Förderung

Die vorliegende Untersuchung der bisherigen Evaluierungspraxis in Österreich konzentriert sich auf die Evaluierungen von FTE-Förderungsprogrammen und intendiert nicht, das weite Feld der Evaluierung von FTE-Institutionen (Förderinstitutionen, universitäre und außeruniversitäre Forschungseinrichtungen etc.) in seiner Gesamtheit abzudecken. Aus dieser Orientierung ergibt sich implizit eine Konzentration auf die in Hinblick auf Schwerpunktsetzungen am stärksten ausdifferenzierte Institution der österreichischen FTE-Förderungslandschaft, den Innovations- und Technologiefonds (ITF). Diese Sicht ist insofern jedoch nicht verzerrend als - wie sich zeigen wird - ein großer Teil der Evaluierungsaktivitäten im FTE-Förderungsbereich in der Realität auf die Schwerpunkte des ITF konzentriert war.

Bevor die Evaluierung österreichischer FTE-Programme im Detail betrachtet wird, erscheint es sinnvoll, den derzeitigen Stand der Evaluierungspraxis im Gesamtsystem der österreichischen FTE-Förderung zu dokumentieren. Dieser Stand der Praxis ist überblicksmäßig in Tabelle 4 dargestellt.

Vergleichsweise umfangreiche Evaluierungstätigkeiten fanden im Bereich der ITF-Schwerpunkte statt. Diese werden weiter unten eingehender behandelt. An dieser Stelle sei erwähnt, daß der ITF als Institution (im Unterschied zu seinen Schwerpunktprogrammen) bisher nicht evaluiert wurde. Über ihre Rolle bei der Abwicklung der ITF-Programme hinaus spielen die beiden großen Fonds, ERP-Fonds und Forschungsförderungsfonds der Gewerblichen Wirtschaft (FFF), jeweils eigenständig eine herausragende Rolle in der österreichischen FTE-Förderlandschaft. Eine Evaluierung der ERP-Kredite im industriell-gewerblichen Bereich wurde von Hahn (1993) vorgenommen. Eine spezielle Evaluierung von FTE-Förderaktivitäten des ERP-Fonds, etwa des ERP-Technologie-Programms, liegt derzeit nicht vor. Im Auftrag des FFF führt das Institut für Gewerbe- und Handwerksforschung interne Nachuntersuchungen der Ergebnisse geförderter Projekte (einschließlich der Berechnung von "Forschungs-" bzw. "Förderungsmultiplikatoren") durch. Eine umfassende, internationalen Standards entsprechende Evaluierung des FFF ist geplant.

Keine Evaluierung gibt es bislang auch im Bereich der indirekten F&E-Förderung, insbesondere des Forschungsfreibetrags, obgleich diese - wie gezeigt werden wird - gemessen am Einnahmehausfall - eine quantitativ durchaus gewichtige Rolle im System der österreichischen FTE-Förderung spielt.

Die Gesamtschau (Tabelle 4) weist darauf hin, daß die Evaluierung des Gesamtsystems der österreichischen FTE-Förderung zur Zeit nur als lückenhaft zu bezeichnen ist. Tragende Elemente des Systems wurden bisher keiner systematischen Evaluierung unterzogen oder wurden bisher kaum in eine umfassende Reformdiskussion einbezogen.

Tabelle 4

Übersicht über Programmevaluierungen			
	ex-ante	begleitend	ex-post
Direkte Förderung:			
ITF		Schwerpunkte: siehe Tabelle 5	
ERP-TECH	-	-	-
FFF	-	Institut für Gewerbe- und Handwerksforschung (intern) Evaluierung geplant	
Indirekte Förderung	-	-	-

4.2 Evaluierungen im Bereich der Investitionsförderung

Die entwickelten Industrieländern haben in den achtziger Jahren ihre Bemühungen um Transparenz im Förderungswesen aus verschiedenen Gründen - wie z.B. den Übergang von strukturerhaltenden Subventionen zur "Strukturanpassung" und wegen budgetärer Restriktionen - deutlich verstärkt (siehe dazu etwa OECD, 1983). Auch in Österreich wurde in den achtziger Jahren in einer Reihe von Arbeiten das System der Wirtschaftsförderungen und einzelne Segmente bzw. Instrumente dieses Systems dargestellt und analysiert. Einen Überblick über die in Österreich durchgeführten Arbeiten der achtziger Jahre bieten Aiginger - Hutschenreiter (1991).

Szopo - Aiginger - Lehner (1985) untersuchten in einer Gesamtdarstellung Ziele, Instrumente und Effizienz des Systems der Investitionsförderung. In weiterführenden Arbeiten beschäftigte sich Szopo mit den Subventionen in Österreich auf Basis der Volkswirtschaftlichen Gesamtrechnung (Szopo, 1986) sowie mit der direkten Wirtschaftsförderung des Bundes (Szopo, 1990). Dabei ging es insbesondere darum, mittels der - damals noch nicht allgemein etablierten - Bewertung der Subventionsströme zu Barwerten zu einer konsistenten Quantifizierung zu gelangen und damit mehr Transparenz im Förderungswesen zu schaffen.

Eine Evaluierung von Förderungsprogrammen im strikten Sinn des Wortes wurde höchstens ansatzweise vorgenommen.

Was die Evaluierung auf Programmebene betrifft, spielte rückblickend betrachtet die TOP-Aktion, ein regelgebundenes Verfahren qualifizierter Investitionsförderung, eine wichtige Rolle. Zur konzeptionellen Grundlage siehe Aiginger - Bayer (1982) sowie - im Rückblick - Aiginger (1996). Die Zielsetzungen der TOP-Aktion wurden ex-ante durch großteils quantitative Kriterien operationalisiert.

Die Kriterienerfüllung wurde mittels eines entsprechenden Bewertungsschemas gemessen: Auf Basis vorgegebener (empirisch ermittelter) Vergleichswerte (Benchmarks) wurden je Kriterium Punkte aus einer vorgegebenen "Notenskala" vergeben, die gemäß dem ebenfalls vorgegebenen Gewicht des entsprechenden Kriteriums in die Gesamtbeurteilung des Förderungsantrags einging. Zur Gewährung einer Förderung mußte eine fixierte Mindestanzahl von Punkten erreicht werden.

Die Evaluierung der Regionalen Innovationsprämie (RIP) - eines bedeutenden regionalpolitischen Instruments der qualifizierten Investitionsförderung - knüpfte unmittelbar an die Erfahrungen mit der TOP-Aktion an. Insbesondere wurde - siehe Hutschenreiter (1993A) sowie Beirat für Wirtschafts- und Sozialfragen (1994) - ein den Zielsetzungen der Regionalen Innovationsprämie entsprechend modifiziertes Bewertungsverfahren entwickelt. Mit diesem sollten

- einerseits strukturpolitische Anforderungen explizit in die Projektbeurteilung einbezogen und
- andererseits der mit der Regionalen Innovationsprämie - gemäß den unterschiedlichen regionalwirtschaftlichen Ausgangs- und Problemlagen - intendierten Differenzierung der Förderungspraxis nach Regionstypen (alte Industriegebiete - periphere Gebiete) Rechnung getragen werden.

Das vorgeschlagene Verfahren sollte in der folgenden Weise eine differenzierte Bewertung der eingereichten Projekte nach dem Projektstandort unterstützen:

- Erstens wird - bei quantitativen Kriterien - die Kriterienerfüllung anhand regionsspezifischer Vergleichskennzahlen gemessen. Die (quantitativen und qualitativen) Kriterien sind folgenden Obergruppen zugeordnet: 1. Unternehmen I (Dynamik), 2. Unternehmen II (strukturpolitische Einordnung), 3. Projekt, 4. Produkt und Markt, 5. Regionale Aspekte, 6. Nebenaspekte.
- Zweitens gehen die Kriterien mit unterschiedlichem Gewicht in die Gesamtbeurteilung ein. Dem innovativen Gehalt eines Projektes wurde etwa in alten Industriegebieten ein höheres Gewicht zugewiesen als in Randgebieten.

Es ist jedoch festzuhalten, daß - im Unterschied zur TOP-Aktion - das im Rahmen der Evaluierung der Regionalen Innovationsprämie vorgeschlagene Bewertungsverfahren in der Praxis nicht im Prozeß der Vergabeentscheidung angewandt wurde.

Soll ein *regelgebundenes Verfahren* wirksam sein, so muß es den Entscheidungsprozeß notwendigerweise "strukturieren", d.h. aber auch ihm bindende Beschränkungen auferlegen. An eben dieser Eigenschaft knüpft auch die Kritik an regelgebundenen Verfahren an. Die Einwände betonen die "Willkürlichkeit" der Kriterien, die Starrheit des Verfahrens in bezug auf atypische Fälle und Eventualitäten etc.

Dem kann entgegengehalten werden, daß *ein regelgebundenes Verfahren wie das beschriebene geeignet ist, die Rationalität von Förderungsprogrammen zu erhöhen*, indem es

- die (ansonsten erfahrungsgemäß oft unscharf gehaltenen) Ziele des Programms klar festlegt,
- Argumentationszwänge im Entscheidungsprozeß schafft und
- diskretionäre Eingriffe hintanhält,
- Transparenz herstellt und
- die Evaluierung erleichtert.

Aus den Erfahrungen mit der TOP-Aktion können einige wichtige Lehren für die Evaluierungspraxis gezogen werden. Dazu zählen die folgenden:

- Die Ex-ante-Operationalisierung der Programmziele erleichtert die Programmevaluierung.
- Für eine Evaluierung (interim oder ex-post) notwendige Daten fallen quasi als Nebenprodukt im Prozeß der Vergabeentscheidung an.
- Periodische Evaluierungen im Sinne einer vergleichenden Gegenüberstellung der Daten für die geförderten Unternehmen mit entsprechenden Benchmark-Werten ergeben sich als logische Konsequenz des Bewertungsverfahrens.

Auch für die Evaluierungspraxis im Bereich der FTE-Programme sind diese Erfahrungen wertvoll. So etwa ist zu überlegen, bei der Programmformulierung bzw. im Rahmen einer Ex-ante-Evaluierung die Programmziele zu operationalisieren und die Bewertung zu formalisieren. Das heißt einerseits, daß die Programmziele klar und widerspruchsfrei definiert werden und möglichst an beobachtbare und verifizierbare Variable geknüpft werden. (Dies schließt Zielformulierungen der Qualität "Erhaltung und Erhöhung der Wettbewerbsfähigkeit" aus.) Andererseits ist damit eine Strukturierung der Projektbewertung anhand von Variablen und Kriterien gemeint, die für die Erreichung der Programmziele zentral sind.

Für Zwecke der Ex-post-Evaluierung von FTE-Programmen ist der aus der Praxis der Investitionsförderung stammende Ansatz der TOP-Aktion allerdings aus verschiedenen Gründen ergänzungsbedürftig. Insbesondere "marktferne" FTE-Projekte sind in der Regel mit hoher Unsicherheit behaftet. Ihre Resultate sind daher ex-ante - nicht nur von den externen Begutachtern sondern auch von den in der Regel besser informierten Akteuren selbst - schwer abschätzbar. Zwischen den Einschätzungen ex-ante und Ergebnissen ex-post ist daher mit dem Auftreten großer Abweichungen zu rechnen. In Zusammenhang mit der Ex-post-Evaluierung von FTE-Programmen ist es daher einerseits dringlicher (und gleichzeitig auch schwieriger) - mit entsprechendem Zeitabstand - die tatsächlich eingetretenen Effekte der Programme zu messen. Andererseits sind bei der Evaluierung von FTE-Programmen zusätzliche oder andere Indikatoren heranzuziehen, die den spezifischen Charakteristika und Zielsetzungen von FTE-Programmen Rechnung tragen.

4.3 Evaluierung österreichischer Technologieprogramme

Technologieschwerpunktprogramme haben in Österreich eine relativ kurze Geschichte. Erst Mitte der achtziger Jahren liefen - quasi als Pilotversuch - die Technologieförderungsprogramme 1985-1987 der Bundesregierung an. Im Anschluß an diese wurde im Jahr 1987 der Innovations- und Technologiefonds (ITF) eingerichtet, aus dessen Mitteln Schwerpunktprogramme finanziert wurden.

Gegen Ende der achtziger Jahre wurden die ersten Schritte zur Evaluierung im Bereich der FTE-Programme eingeleitet. Einen Überblick geben Steiner - Sturn (1995). In dieser ersten Phase (Ende der achtziger und Anfang der neunziger Jahre) waren dabei zwei Schwerpunkte der Evaluierungstätigkeit erkennbar:

1. die Ex-post-Evaluierung von Programmen, die schon ausgelaufen waren bzw. bereits einige Zeit liefen. Dazu zählt die Ex-post-Evaluierung der Technologieschwerpunktprogramme der Bundesregierung 1985-1987 sowie die - weiter unten behandelte - Evaluierung der österreichischen ESA-Beteiligung.
2. die Aufbereitung von Entscheidungsgrundlagen für die Schwerpunktfindung im Rahmen des ITF.

Die erste umfassende FTE-Programmevaluierung in Österreich stellt die Evaluierung der Technologieförderungsprogramme der Bundesregierung 1985-1987 mit den Schwerpunkten Mikroelektronik und Informationstechnik, einschließlich CAD/CAM, sowie Biotechnologie und Gentechnik (Hutschenreiter et al., 1991) dar.

Die Evaluierung der Technologieschwerpunktprogramme der Bundesregierung stellt nach wie vor die umfangreichste abgeschlossene Evaluierung eines nationalen Technologieprogramms in Österreich dar. Bei der Untersuchung der Technologieförderungsprogramme handelte es sich um eine Ex-post-Evaluierung. Sie ist aber auch im Kontext der Weiterführung der Schwerpunktförderung im Rahmen des 1987 eingerichteten ITF zu sehen. Das Evaluierungsprojekt war interdisziplinär angelegt und umfaßte volkswirtschaftliche, betriebswirtschaftliche, technische, forschungspolitische und soziale/qualifikationsbezogene Aspekte der Programme. Das mit der Projektleitung betraute WIFO untersuchte insbesondere volks- und betriebswirtschaftliche Wirkungen der Programme. Die technische Seite wurde durch das Forschungszentrum Seibersdorf abgedeckt. Darüber hinaus wurde die Rolle der wissenschaftlichen Forschungseinrichtungen durch das Interdisziplinäre Forschungszentrum Sozialwissenschaften und soziale und qualifikationsbezogene Auswirkungen durch das Institut für sozio-ökonomische Entwicklungsforschung der Österreichischen Akademie der Wissenschaften untersucht.

Insgesamt wurden im Rahmen der Technologieförderungsprogramme nahezu 400 Projekte mit einem Förderungsvolumen von 730 Mill. S gefördert. Dazu kam noch die Förderung ausgewählter Schwerpunktinstitute im Bereich des Schwerpunkts Mikroelektronik und Informationsverarbeitung. Die Technologieförderungsprogramme waren aber nicht nur von ihrem - damals doch vergleichsweise beträchtlichen - Volumen her von Interesse. Ein Novum stellte für Österreich vor allem auch der Versuch dar, durch die Förderung von Unternehmen einerseits und Forschungsinstituten andererseits ein höheres Maß der Integration von Technologie- und Forschungspolitik zu erreichen. Unter diesem Aspekt stellen die Technologieförderungsprogramme der achtziger Jahre ein nach wie vor interessantes Experiment dar. Die Evaluierungsstudie bewertete gerade die konzeptionelle Vorbereitung und Einbindung in die Forschungspolitik als Vorzüge der Technologieschwerpunktprogramme. Wie bei allen Innovationen traten dabei - zum Teil unvorhergesehene - Umsetzungsprobleme auf, die in der Evaluierung eine prominente Rolle spielten.

Was die Wirkung der Evaluierungsstudie betrifft, so hat deren generell positive Tenor sicherlich das Vertrauen in die Machbarkeit und Effektivität der damals noch neuen technologiepolitischen Förderungsprogramme gestärkt. Die Wirkung der Studie war jedoch u.a. dadurch beschränkt, daß gerade der grundsätzlich positive bewertete Ansatz zur Umsetzung integrierter technologiepolitischer Programme nicht weiterverfolgt wurde.

Die Evaluierung der Technologieförderungsprogramme der Bundesregierung stand - gemeinsam mit der Evaluierung des österreichischen Engagements im Rahmen der ESA -

längere Zeit für sich allein. Erst heute findet sie gewissermaßen eine Fortsetzung, etwa in den in Arbeit befindlichen oder anlaufenden Ex-post-Evaluierungen von Schwerpunktprogrammen des ITF.

Die Erarbeitung von Entscheidungsgrundlagen für die Schwerpunkte des ITF stellte die zweite - in dieser Phase auch dominierende - Entwicklungslinie zu Beginn der neunziger Jahre dar. Im Lauf der Zeit hat sich die Nachfrage des ITF nach Evaluationen bzw. evaluationsnaher Expertise weiter entfaltet. Tabelle 5 vermittelt einen Überblick über die in Zusammenhang mit ITF-Schwerpunkten beanspruchte "externe Expertise" (ex-ante, interim und ex-post). Die darin enthaltenen Aktivitäten sind nicht auf Evaluierungen im strikten Sinn beschränkt, sondern umfassen die gesamte Palette "externer Expertise" (Vorstudien, Programm-Management etc.), da darin auch gewisse Evaluierungs-Elemente enthalten sind und die Grenzen zu Evaluierungen im engeren Sinn fließend sind. Evaluierungen im engeren Sinn sind in Tabelle 5 durch Fettdruck hervorgehoben.

Tabelle 5

ITF-Schwerpunkte	Externe Expertise		
	ex-ante	begleitend	ex-post
Biotechnologie	-	-	-
Mikroelektronik	-	-	-
Neue Werkstoffe	ATMOS	-	-
Umweltverfahrenstechnik	ATMOS	ÖFZS, Öko-Consult	ISI/WIFO, in Arbeit, O.Ö. Umweltakademie (Öko-Audit)
Internationale Projekte	-	-	-
Industrielle Lasertechnik	ATMOS	-	-
Weltraumtechnologie	-	-	WIFO
Flexible computerintegrierte Produktion	ATMOS, ÖFZS	tip	demnächst geplant
Energietechnik	ÖFZS/ Energieverwertungs-agentur /TU Wien/WIFO	Energieverwertungs-agentur/ JOANNEUM RESEARCH	Zivilingenieur Bruck, in Arbeit
Industrial Design	ÖFZS/TU Wien/WIFO	-	-
Verkehrstechnik	ATMOS	AMC	demnächst geplant
Softwaretechnik	ÖFZS/TU Wien	-	-
Informationsgesellschaft	tip, N. Knoll	JOANNEUM RESEARCH /AustriaPro	-

			Evaluierung geplant
Technologietransfer	-	-	-
Seed-financing	ÖFZS	TRICONSULT	Institut f. Soziologie, Univ. Wien, in Arbeit

Ex ante wurden - wie aus Tabelle 5 hervorgeht - für die überwiegende Zahl der ITF-Schwerpunkte Entscheidungsgrundlagen durch externe Expertenteams erarbeitet. Die Aufbereitung von Entscheidungsgrundlagen für die Schwerpunkte des ITF erfolgte Anfang der neunziger Jahre zunächst im Rahmen der Projekts ATMOS - einer Arbeitsgemeinschaft bestehend aus der von Prof. Götschl und Prof. Stepan initiierten Arbeitsgemeinschaft für Technologiemonitoring "Atmo", ÖFZS und WIFO - und später überwiegend durch variable Arbeitsgemeinschaften. Die technologiepolitischen Empfehlungen der umfangreichen Untersuchungen in sechs ausgewählten Technologiefeldern (Verfahrenstechnik, Computer Integrated Manufacturing, Werkstofftechnik, Medizintechnik, Lasertechnik und Verkehrstechnik) sind in einem Optionen-Band (ATMOS, 1990) zusammengefaßt. Vorstudien gab es in der Folge auch zu den Schwerpunkten Energietechnik (Ohler - Hutschenreiter - Mühlberger, 1992), Industrial Design (Ohler - Leo et al., 1992) und Softwaretechnik (Ohler - Katzenberger, et al., 1992) und zum Schwerpunkt Informationsgesellschaft die tip-Telekommunikations-Studie (Leo et al., 1994).

Die genannten Untersuchungen sind generell nicht als Ex-ante-Evaluierungen im engeren Sinn zu bezeichnen, wenngleich sie jeweils auch Evaluierungs-Elemente enthalten. Einerseits fehlen ihnen Merkmale einer (idealtypischen) Evaluierung, andererseits waren sie breiter angelegt als Ex-ante-Programmevaluierungen im herkömmlichen Sinn. Die genannten Studien geben in der Regel einen Überblick über die Ausgangssituation und Potentiale in den jeweiligen Schwerpunkt-Bereichen und spannen einen Raum möglicher Aktionslinien auf, die zum Teil weit über die unmittelbaren Möglichkeiten und Instrumente des ITF hinausgehen. Dieser Umstand entsprach auch dem Selbstverständnis des ITF als Forum technologiepolitischer Diskussion, Koordination und Strategieentwicklung.

Im Interim-Bereich wurde bei einigen ITF-Schwerpunkten eine externe Programmbegleitung eingerichtet, so in den Schwerpunktprogrammen Umwelttechnik, Energietechnik und Verkehrstechnik. Eine Interim-Evaluierung im engeren Sinn wurde im Bereich Seed-financing (Triconsult, 1991) und des ITF-Schwerpunkts Flexible computerintegrierte Produktion (FlexCIM) vorgenommen. Die Evaluierung der Pilotphase dieses Förderprogramms wurde im Rahmen des Programms tip, einer Arbeitsgemeinschaft bestehend aus WIFO und ÖFZS (Polt et al., 1994) durchgeführt. Eine Interim-Evaluierung ist im Bereich des Schwerpunkts Informationsgesellschaft geplant.

Die Entwicklung im Bereich der Ex-post-Evaluierungen zeigt, daß wir uns derzeit an einem Knotenpunkt in der Geschichte der Evaluierungen im Bereich des ITF befinden. Nach der frühen WIFO-Evaluierungsstudie im Bereich Weltraumtechnologie (Leo, 1991), die eine

Bewertung der österreichischen ESA-Beteiligung zum Gegenstand hatte (und daher weiter unten behandelt wird), gab es über längere Zeit keine weiteren Ex-post-Evaluierungen, sieht man von der Evaluierung der Pilotaktion Öko-Audit (O.Ö. Umweltakademie, o.J.) ab. Nun, nach zehnjährigem Bestehen des ITF, ist ein neuerlicher Aufschwung bei den Evaluierungsaktivitäten festzustellen. Derzeit befinden sich drei Ex-post-Evaluierungen in Arbeit:

- Eine Evaluierung des Schwerpunkts Energietechnik durch Zivilingenieur Bruck.
- Eine Evaluierung des Seed-financing-Programms durch das Institut für Soziologie der Universität Wien.
- Eine Ex-post-Evaluierung des Schwerpunkts Umwelttechnik. Dieser derzeit in Arbeit befindlichen Studie kommt strategische Bedeutung zu. Sie wird gemeinsam durch das in Evaluierungen von Technologieprogrammen erfahrene Fraunhofer-Institut für Systemtechnik (ISI, Karlsruhe) und das WIFO durchgeführt.

Weitere Ex-post-Evaluierungen von ITF-Schwerpunkten (Verkehrstechnik, FlexCIM) sind für die nächste Zukunft vorgesehen.

Zusammenfassend ergibt sich aus der Bestandsaufnahme über die Evaluierungen im Bereich des ITF folgendes Bild:

- Studien zur Entscheidungsvorbereitung - wenn auch nicht Ex-ante-Evaluierungen im strikten Sinn - wurden über die ITF-Schwerpunkte ziemlich flächendeckend durchgeführt.
- Die Auswahl der Schwerpunkte, die einer externen Ex-ante-Bewertung unterzogen wurden, war de facto vorgegeben. Gegen diese Art der Vorselektion bestehen unseres Erachtens auch keine grundsätzlichen Einwände. Die Frage nach der Transparenz und Effizienz der angewandten Verfahren der Schwerpunktidentifikation und -auswahl ist dann allerdings bereits in der vorgelagerten Entscheidungsphase zu lösen. Es sei in diesem Zusammenhang daran erinnert, daß anlässlich der Evaluierung des Projekts ATMOS im Jahr 1991 durch ein gemischtes Team von Professoren und Industrievertretern ein Mangel an Transparenz in Bezug auf die Auswahl der vorgegebenen Technologiefelder thematisiert wurde.
- Hervorzuheben ist auch der Umstand, daß das Spektrum der in den Ex-ante-Bewertungen vorgeschlagenen Maßnahmen ex-post auf solche Maßnahmen reduziert wurden, die im Rahmen des ITF umsetzbar waren. Dies ist selbstverständlich nicht dem ITF anzulasten, sondern verweist auf grundsätzliche Probleme der Koordination der Technologiepolitik mit anderen Politikbereichen, wie sie im Expertenentwurf zum Technologiepolitischen Konzept 1996 der Bundesregierung angesprochen wurden. Hinzu kommt das generell relativ geringe Volumen, das zur Finanzierung der ITF-Schwerpunkte jeweils zur Verfügung steht.
- Interim-Evaluierungen im engeren Sinn spielten eine geringe Rolle (Ausnahme: Seed-financing, FlexCIM). Über die Laufzeit des ITF hat sich allerdings eine erkennbare Tendenz

entwickelt, in den Schwerpunkten ein begleitendes Programm-Management unter Einbeziehung externer Expertise zu institutionalisieren.

- Derzeit befinden wir uns an einem Knotenpunkt in der Geschichte der Evaluierung im Bereich des ITF (und auch darüber hinaus). Nach zehnjährigem Bestehen des ITF befinden sich derzeit mehrere Ex-post-Evaluierungen von Schwerpunktprogrammen in Arbeit bzw. in Planung. Diese Kumulierung bietet eine Möglichkeit zur Setzung neuer, internationaler Standards in der österreichischen Evaluationspraxis. Die Evaluierung es Schwerpunkts Umwelttechnik könnte dabei Modellcharakter haben. Und letztlich bieten die anstehenden Evaluierungen auch eine Chance zur Erhöhung des Stellenwerts von Evaluierungen in technologiepolitischen Entscheidungsprozessen.
- Im Unterschied zu den Schwerpunktprogrammen wurde der ITF als Institution bisher keiner Evaluierung unterzogen.

4.4 Evaluierung von Instrumenten der indirekten F&E-Förderung

In Österreich wurde bislang keine Evaluierung im Bereich der indirekten F&E-Förderung durchgeführt. Dennoch wird das Thema hier aus den folgenden Gründen behandelt: Erstens liegt international - im Unterschied zu Österreich - eine Vielzahl von Studien vor, die sich mit den Effekten indirekter Förderungsinstrumente beschäftigen. Zweitens wurden in jüngster Zeit in Österreich mehrere Vorschläge unterbreitet, die auf eine Reform und einen Ausbau der indirekten F&E-Förderung abzielen. Sicherlich sollte bei der Realisierung größerer Reformschritte in diesem Bereich auf eine entsprechende Evaluierung ex-ante und ex-post Wert gelegt werden, wobei auf die vorliegenden internationalen Erfahrungen zurückgegriffen werden kann.

Als potentielle Vorteile der indirekten F&E-Förderung (siehe etwa Leyden - Link, 1992, Harhoff, 1994) werden die folgenden genannt:

- Indirekte F&E-Förderung ist durch ein hohes Maß an "Neutralität" gegenüber unternehmerischen Allokationsentscheidungen gekennzeichnet. Dies betrifft sowohl den Inhalt und Charakter der Projekte, als auch die Kombination von Inputs. Eine generelle steuerliche Förderung ist im Prinzip auch neutral gegenüber der sektoralen oder Branchenzugehörigkeit der Unternehmen.
- Steuerliche Förderungen weisen eine niedrige Zutrittschwelle auf, was insbesondere für kleine und mittelgroße Unternehmen von Bedeutung ist.
- Die Förderungsbedingungen sind vergleichsweise transparent.

- Indirekte Förderungen sind für die Unternehmen planbar. Sie unterscheiden sich diesbezüglich von der direkten (d.h. diskretionären) Projektförderung, es sei denn letztere ist nicht selektiv.
- Die Administrationskosten können vergleichsweise gering gehalten werden.

Diesen Vorteilen der indirekten F&E-Förderung werden häufig die folgenden potentiellen Nachteile gegenüber Instrumenten der direkten Förderung gegenübergestellt:

- Mitnahmeeffekte;
- ungleiche Verteilung des Nutzens ("inequity");
- der durch die indirekte Förderung induzierte Ausfall von Steuereinnahmen führt ceteris paribus zu höheren Steuersätzen;
- steuerliche Anreize unterminieren die Budgetkontrolle;
- die Anreize indirekter F&E-Förderung variieren über den Produktlebenszyklus.

Eine Abwägung der zitierten Argumente nehmen Leyden - Link (1992) vor. Häufig wird auch der Einwand vorgebracht, daß

- die tatsächlichen Anreizwirkungen indirekter F&E-Förderung unklar sind.

Zu dieser Frage, die letztlich nur empirisch zu beantworten ist, siehe das Beispiel zur Evaluierung indirekter F&E-Förderungsinstrumente in Abschnitt 3.2.

"R&D tax credits" (das sind Steuerabsetzbeträge für F&E-Ausgaben) sind international die verbreitetste Form der indirekten F&E-Förderung. In manchen Ländern gibt es eine "*Negativkomponente*" (Umwandlung des nicht beanspruchbaren Teils des Absetzbetrags in einen Zuschuß). Ein reiner Absetzbetrag ohne diese Negativkomponente benachteiligt in der Tendenz technologieorientierte Unternehmensneugründungen, die in der Startphase in der Regel keine Gewinne erzielen. Darüber hinaus wirkt er prozyklisch, was unter dem Gesichtspunkt der Verstetigung der F&E-Tätigkeit nicht wünschenswert erscheint. Alternativ zu einer Negativkomponente gibt es die Möglichkeit zu Verlustvor- bzw. -nachträgen (z.B. beim "R&D tax credit" in den USA). Dies wirkt der Zyklizität entgegen, Liquiditätsprobleme von technologieorientierten Unternehmen in der Start-up-Phase werden mit dieser Variante jedoch nicht gemildert.

Das derzeit bedeutendste Instrument der indirekten F&E-Förderung in Österreich ist der Forschungsfreibetrag. Ein Überblick über die ganze Palette der indirekten F&E-Förderung in Österreich bietet Taucher (1994).

Gemäß § 4 Abs. 4 EStG wird - zusätzlich zu den sofort als Betriebsausgaben geltend zu machenden - F&E-Aufwendungen ein Forschungsfreibetrag von 12% der Aufwendungen zur Entwicklung oder Verbesserung "volkswirtschaftlich wertvoller Erfindungen" anerkannt. Ein erhöhter Forschungsfreibetrag (18%) wird anerkannt, wenn die Forschungsergebnisse im wesentlichen eigenbetrieblich verwertet werden (die Einnahmen aus der Verwertung der Erfindungen nicht mehr als 25% der Forschungsaufwendungen des betreffenden Wirtschaftsjahres betragen).

Hinsichtlich der direkten budgetären Wirkungen liegen Schätzungen des Einnahmenausfalls vor. Im Förderungsbericht 1995 werden die geschätzten Einnahmenausfälle aufgrund § 4 Abs. 4 EStG (Forschungsfreibetrag, Spendenbegünstigung für Wissenschaft und Forschung) für die Jahre 1993, 1994 und 1995 mit jeweils 450 Mill. S per annum angesetzt. Am WIFO durchgeführte Schätzungen bewegen sich - ausgehend von den vorliegenden Daten bezüglich der Inanspruchnahme des Forschungsfreibetrags - in einer ähnlichen Größenordnung.

Eine Evaluierung indirekter Instrumente der Innovationsförderung steht in Österreich - wie eingangs erwähnt - noch aus. Allgemein dürfte die Meinung vorherrschen, daß dies vor allem auf den Umstand zurückzuführen ist, daß Österreich über kein ausgebautes Instrumentarium der indirekten Innovationsförderung verfügt. Dem kann entgegengehalten werden, daß das geschätzte Ausmaß des durch den Forschungsfreibetrag induzierten Einnahmenausfalls in Relation zu den anderen Instrumenten der Technologieförderung durchaus nicht vernachlässigbar gering ist. Es gilt daher wohl auch, daß "tax expenditures" weniger im Zentrum der öffentlichen Aufmerksamkeit stehen als direkte Förderungen, obwohl sie aus finanzwissenschaftlicher Sicht - grob gesprochen - äquivalent sind. Diese Einschätzung stimmt mit der Beobachtung überein, daß Fragen der Effektivität und Effizienz (etwa gestützt durch ökonometrische Analysen) von umfangmäßig noch viel bedeutenderen Instrumenten der indirekten Förderung, wie etwa des Investitionsfreibetrags, kaum in der öffentlichen wirtschaftspolitischen Diskussion stehen. Der Förderungsbericht 1995 beziffert den aus dem Investitionsfreibetrag resultierenden Einnahmenausfall in den Jahren 1993-1995 mit 17 Mrd. S (1993), 15 Mrd. S (1994) bzw. 8 Mrd. S (1995).

Im Expertenentwurf für das Technologiepolitische Konzept der Bundesregierung 1996 werden für den Bereich der indirekten F&E-Förderung unter anderem die folgenden Optionen vorgeschlagen:

- Reform des Forschungsfreibetrags.
- Steuerliche Förderung von (zusätzlichem) wissenschaftlichen und technischen Personal.

Darüber hinaus steht auch eine Erhöhung des Forschungsfreibetrags zur Diskussion. Zur Frage der Erhöhung des Forschungsfreibetrags werden in der wirtschaftspolitischen Diskussion

kontroversielle Meinungen vertreten. Wirtschaft und Industrie befürworten eine Erhöhung des Forschungsfreibetrags (auf 30%). Ablehnende Positionen hierzu werden primär mit der vermuteten geringen Effizienz des Forschungsfreibetrags ("Mitnahme") sowie mit budget- bzw. steuerpolitischen Notwendigkeiten begründet. Unverkennbar hat die Diskussion direkte versus indirekte Innovationsförderung auch ideologische Obertöne. Die Beurteilung der relativen Vor- und Nachteile korreliert mit der relativen Bewertung der Bedeutung von "Marktversagen" einerseits und "Staatsversagen" andererseits.

Die Motivation für eine F&E-Personalkostenförderung ist primär in folgendem zu sehen: Eine vordringliche Aufgabe der österreichischen Technologiepolitik besteht darin, den Kreis der regelmäßig F&E betreibenden Unternehmen zu erweitern und damit ihr Innovationspotential sowie die Fähigkeit zur Übernahme von extern produziertem Wissen zu erhöhen. Eine an den Arbeitskosten (für zusätzliches) F&E-Personal ansetzende Förderung ist transparent, hat hohe Signalwirkung und ist relativ leicht administrierbar. Ein weiterer Vorteil besteht in der hohen politischen Akzeptanz (Beirat für Wirtschafts- und Sozialfragen, 1994). Das Ziel besteht darin, erstmalig bzw. zusätzlich F&E-Personal in Unternehmen, vor allem KMUs, zu bringen bzw. deren F&E-Tätigkeit zu stabilisieren. Entlastende Effekte sind auch auf dem Arbeitsmarkt für Techniker zu erwarten.

Sowohl der geschätzte Umfang des Einnahmenausfalls als auch die Diskussion um Reform bzw. Ausbau des Forschungsfreibetrags legen die Durchführung einer Ex-post-Evaluierung des bestehenden Instrumentariums der indirekten F&E-Förderung nahe. Wie gezeigt liegen diesbezüglich umfangreiche internationale Erfahrungen vor, die dabei nutzbar gemacht werden können. Vor allem in den USA wurde eine Vielzahl, qualitativ z.T. sehr hochwertiger ökonomischer Forschungsarbeiten unternommen (siehe Abschnitt 3.2).

Zur Vorbereitung größerer Reformschritte im Bereich der indirekten Innovationsförderung erscheint es darüber hinaus erforderlich, eine Ex-ante-Evaluierung vorzunehmen. Nicht zuletzt zeigen internationale Forschungsarbeiten, daß das Design indirekter Förderungsinstrumente gravierende Auswirkungen auf die davon ausgehenden Anreizwirkungen haben kann. Welche Art von Fragen im Rahmen einer Ex-ante-Evaluierung zu untersuchen ist, soll hier beispielhaft anhand der vorgeschlagenen Einführung einer indirekten Förderung von F&E-Personalkosten illustriert werden:

- Sollen die gesamten F&E-Personalkosten gefördert werden oder nur die zusätzlichen ("inkrementellen") Ausgaben? Eine inkrementelle Förderung ist grundsätzlich die effizientere Form der Förderung. Sie führt in der Regel auch zu einer relativ weniger ausgeprägten Begünstigung von Großunternehmen. Eine Bestimmung der "zusätzlichen" F&E-Personalkosten bedarf natürlich der Festlegung eines Referenzwerts ("baseline"). Die

zusätzlichen Ausgaben sind als Differenz zwischen den aktuellen und den Referenzausgaben definiert. Die Definition der "baseline" ist allerdings tückisch, eine Fehlspezifikation kann die Wirksamkeit eines indirekten Förderungsinstruments drastisch vermindern. Im Fall des "R&D tax credit" in den USA wurde die "baseline" als Durchschnitt der F&E-Ausgaben der letzten drei Jahre definiert. Dies hatte zur Folge, daß die Erhöhung der F&E-Ausgaben in einem Jahr die "baseline" in den folgenden beiden Jahren erhöhte. Diese Erhöhung vermindert unter sonst gleichen Umständen den Umfang der *zusätzlichen* F&E-Personalausgaben in den Folgejahren und damit die Anreizwirkung der Förderung. In Verbindung mit einer Obergrenze für die zusätzlich anrechenbaren jährlichen F&E-Ausgaben konnte der Modus des "R&D tax credit" rechnerisch auch zu durchaus paradoxen Ergebnissen führen.

- Die betriebsinternen F&E-Ausgaben der österreichischen Industrie bestehen zu 62% (1993) aus Personalkosten. Ist die Schaffung eines eigenen Instruments (neben dem Forschungsfreibetrag) gerechtfertigt oder ist eine spezielle F&E-Personalkostenförderung in das allgemeine Schema (Forschungsfrei/absetzbetrag) integrierbar?
- Soll es - wie in anderen Ländern - eine Beschränkung der Förderung auf KMUs oder (wie spezifizierte?) Förderungsobergrenzen geben? Soll es eine Begünstigung z.B. der "ersten" F&E-Mitarbeiter eines Unternehmens oder für junge F&E-Mitarbeiter geben?

4.5 Evaluierung der österreichischen Teilnahme an internationalen Forschungs- und Technologieprogrammen

Von der Teilnahme an internationalen Forschungs- und Technologieprogrammen gehen bedeutende Impulse aus, die sowohl die Evaluatoren als auch die Nutzer der Evaluierungen, insbesondere die technologiepolitischen Entscheidungsträger, betreffen. Aufgrund der hohen Evaluierungsstandards der EU ist von dieser Form der Internationalisierung eine Weiterentwicklung der Evaluierungskultur im Land sowie die Einführung internationaler Standards der Evaluierung zu erwarten.

Die erste Studie, die die Teilnahme an einem internationalen Technologieprogramm zum Gegenstand hatte, liegt bereits einige Zeit zurück. Etwa zur selben Zeit, in der die Evaluierung der Technologieförderungsprogramme 1985/1987 durchgeführt wurde, untersuchte Leo (1991) die Effekte der Beteiligung Österreichs an den Programmen der European Space Agency (ESA). Die Studie von Leo enthält nicht nur einen Überblick über den "State-of-the-art" der quantitativen Methoden der FTE-Programmevaluierung, sondern stellt einen eigenständigen Versuch dar, fortgeschrittene Methoden der Programmevaluierung einzusetzen. Im Rahmen der Studie wurde eine Unternehmensbefragung zur bisherigen Entwicklung der Unternehmen der

Weltraumindustrie und zu den bereits eingetretenen und erwarteten Nutzen des Raumfahrtengagements durchgeführt. In einem weiteren Schritt wurde versucht, die Ausstrahlwirkungen der Raumfahrt zu quantifizieren. Die Methode, die der quantitativen Abschätzung der Ausstrahlwirkungen zugrundeliegt, beruht auf jener von Mansfield et al. (1977). Darüber hinaus wurde ein Entwicklungsszenario für die österreichische Weltraumindustrie entwickelt. Eines der wichtigsten Ergebnisse der Studie bestand darin, daß die häufig reklamierten "Spin-offs" der Weltraumindustrie in Form kommerzialisierbarer Produkte in der Praxis nicht nachweisbar waren. Was die Wirkung dieser ambitionierten Evaluierungsstudie betrifft so kann heute - mit einiger zeitlicher Distanz - festgestellt werden, daß diese zwar einen Einfluß auf die Rhetorik der "Raumfahrer" ausübte, die sich jedoch nur teilweise in den Investitionsentscheidungen niedergeschlagen hat.

Ein wesentlicher Vorschlag der Evaluierungsstudie zielte darauf ab, Schwerpunktbereiche (Erdbeobachtung und Telekommunikation) zu schaffen. Diese Empfehlung wurde insbesondere durch den Einstieg in die bemannte Raumfahrt (Austro-Mir) konterkariert - mit der Konsequenz, daß die Zersplitterung der österreichischen Weltraumaktivitäten verstärkt wurde. Die Struktur der Programmbeteiligung an den fakultativen Programmen der ESA entspricht hingegen eher den vorgeschlagenen Schwerpunkten der Evaluierungsstudie. Die Rückflußstatistik (per 31.12.1996) der ESA-Programme zeigt, daß rund die Hälfte der fakultativen Programmbeteiligungen auf den Bereich Telekommunikation, ein Viertel auf die Erdbeobachtung und ein weiteres Viertel auf die Raumtransportsysteme entfallen. Diese Programmbeteiligung resultiert auch aus der europaweiten Ernüchterung über die Kosten-Nutzen-Relationen der bemannten Raumfahrt.

Vor kurzem wurde auch eine erste Evaluierung der österreichischen Teilnahme an den FTE-Programmen der EU vorgelegt (Ohler et al., 1997). Diese folgt - aus Gründen der Vergleichbarkeit - im Kern den in anderen EU-Mitgliedsländern durchgeführten Evaluierungen. Der Inhalt der Unternehmensbefragung wurde jedoch an österreichische Bedingungen angepaßt. Die österreichische Evaluierungsstudie stellt einen eigenständigen Beitrag zur Praxis der FTE-Programmevaluierung in der EU dar. Eine erste Einschätzung gibt Ohler (1997). Eine Abschätzung des "Impact" dieser Studie ist derzeit noch nicht möglich.

Der Evaluierung von FTE-Programme im europäischen Rahmen kommt in mehrerer Hinsicht "strategische" Bedeutung für die Etablierung einer entwickelten "Evaluierungskultur" zu.

- Im europäischen Kontext gehören Evaluierungen zum Standard. Die Einbeziehung Österreichs in diese Praxis wird zweifellos zu einer stärkeren Verankerung der

Evaluierungspraxis in Österreich beitragen, und zwar über den Anlaßfall, die Teilnahme an europäischen Programmen hinaus.

- Die EU hat in nicht unbeträchtlichem Umfang in die Weiterentwicklung von Evaluierungsverfahren und den Aufbau von internationalen Evaluatorennetzwerken investiert (etwa im Rahmen von MONITOR/SPEAR). Die Involvierung in die Evaluierungspraxis auf europäischer Ebene wird daher einen Know-how-Transfer nach Österreich mit sich bringen.

In diesem Zusammenhang ist ein Trade-off anzusprechen: In Österreich als kleinem Land mit korporatistischen Traditionen und vergleichsweise hoher Persistenz von Institutionen sind über lange Zeiträume aufgebaute enge institutionelle und persönliche Kooperations- oder auch Konkurrenz-Beziehungen. Dies trifft auch auf den FTE-Bereich zu, zumal es sich dabei um einen überschaubaren Bereich mit einer relativ geringen Anzahl von Akteuren handelt. Dies hat den Vorteil, daß österreichische Evaluatoren über intime Kenntnis der institutionellen und politischen Zusammenhänge verfügen, der im Zusammenhang mit Evaluierungen große Bedeutung zukommt.

Allerdings hat diese Beziehungsstruktur auch Nachteile. Wie die mathematische Spieltheorie festgestellt hat, gibt es in "wiederholten Spielen" eine Tendenz zur Kooperation. Siehe etwa Gibbons (1992) oder die experimentelle Anwendung bei Axelrod (1984). Die Herausbildung kooperativen Verhaltens mag in einer Vielzahl von Zusammenhängen positiv sein, ist im Kontext von Evaluierungen aber zumindest teilweise abträglich. Weniger problematisch sind Interim-Evaluierungen durch "Quasi-Insider", bei denen es u.a. häufig darum geht, in Kooperation mit den Programm-Management zu einer Verbesserung der Abläufe zu finden. Vor allem (aber nicht ausschließlich) bei Ex-post-Evaluierungen, bei denen die Kontrollfunktion eine prominente Rolle spielt, ist ein Mangel an Distanz offensichtlich problematisch.

Unter diesen Bedingungen erscheint die Heranziehung internationaler Evaluatoren erforderlich. Andererseits sollten die institutionellen Kenntnisse österreichischer Experten genutzt werden und, was noch wichtiger ist, Österreich sollte nicht auf den Aufbau eigener Kapazitäten im Evaluierungsbereich verzichten. Dafür gibt es eine Reihe von Gründen, darunter die folgenden:

- Wie auch in anderen Bereichen muß Österreich auch im Bereich Evaluierung Kompetenzen aufbauen, um auf europäischer Ebene - durch Evaluierungsexperten - angemessen vertreten zu sein.
- Der europäische Markt für Evaluationsexpertise ist expansiv. Aus forschungspolitischer Sicht ist es wünschenswert, daß österreichische Forschungseinrichtungen an diesem dynamischen Nachfragesegment teilhaben.

- Die Mitarbeit an internationalen Evaluationen vermittelt wertvolle Einsichten, von denen sowohl die Technologiepolitik als auch potentielle Programmteilnehmer im Land profitieren können

Als gangbare Antwort auf dieses Dilemma erscheinen gemischte Evaluatoren-Teams aus internationalen und österreichischen Experten. Die Evaluierung der österreichischen Teilnahme an den europäischen FTE-Programmen folgte diesem Ansatz. Auch bei der derzeit laufenden Evaluierung des ITF-Schwerpunkts Umwelttechnik wurde eine solche Konstruktion gewählt.

5 Empfehlungen

In Hinblick auf die Entwicklung einer österreichischen FTE-Evaluierungskultur lassen sich aus den bisher gewonnenen Erkenntnissen folgende Empfehlungen formulieren bzw. Schlußfolgerungen ableiten.

Empfehlung 1: Evaluierung des gesamten FTE-Systems

Jener Teil der Studie, der sich mit den österreichischen FTE-Evaluierungserfahrungen auseinandersetzt, konzentriert sich im wesentlichen auf die Programme im Rahmen des ITF. Das ergibt sich daraus, daß praktisch nur innerhalb des ITF strategische FTE-Programme formuliert und auch evaluiert werden. Der Umfang der finanziellen Mittel der ITF-Programme ist aber im Vergleich zu anderen Technologiefördermaßnahmen relativ gering. Daraus lassen sich folgende Schlußfolgerungen ableiten:

- Die Entwicklung einer österreichischen Evaluierungskultur im gesamten Bereich der Technologieförderung bedingt eine **stärkere Fokussierung der Forschung und Erfahrungsbildung auf die Evaluierung der relevanten Institutionen**, also der Technologiefördereinrichtungen.
- Um zu Aussagen über das Gesamtsystem der FTE-Förderung kommen zu können, ist auch eine **Evaluierung der indirekten F&E-Förderung** erforderlich.

Der Blick auf das gesamte FTE-System kann helfen, Komplementaritäten zwischen Programmen, Institutionen und Projekten zu identifizieren und auch eine stärkere Integration von Forschungs- und Technologiepolitik unterstützen, wie sie etwa in den achtziger Jahren bei den Technologieschwerpunktprogrammen der Bundesregierung versucht wurde.

Klare Zielvorgaben der nationalen Technologiepolitik - etwa im Rahmen eines Technologiepolitischen Konzepts - erleichtern die Evaluierung des FTE-Systems und seiner Elemente.

Empfehlung 2: Aktuelle Kumulierung von Evaluierungsaktivitäten nutzen

Derzeit stehen im Rahmen des ITF mehrere Schwerpunktprogramme unmittelbar vor ihrer Ex-post-Evaluierung bzw. sind solche Evaluierungen schon im Gange. Auch im Bereich des FFF ist eine Evaluierung geplant. Diese **zeitliche Kumulierung von Evaluierungsaktivitäten** bietet

die Chance, die dabei von den Evaluierungsteams gemachten Erfahrungen zur Erhöhung des Standards der österreichischen Evaluierungspraxis auszunützen und so die Evaluierungskultur in kurzer Zeit wesentlich weiterzuentwickeln. Voraussetzung dafür ist die systematische Beobachtung und Auswertung dieser Evaluierungen in Bezug auf die in ihnen angewandten Methoden und Ansätze. Eine eingehende Diskussion der Ergebnisse im Kreis von Evaluierungsexperten, Programmanministratoren und Entscheidungsträgern sowie die Veröffentlichung der Evaluierungsergebnisse ist ebenfalls erforderlich. Dadurch kann auch der Stellenwert von Evaluierungen in technologiepolitischen Entscheidungsprozessen erhöht werden.

Empfehlung 3: Aus der Evaluierungspraxis in der europäischen FTE-Politik lernen

Programme und größere Aktivitäten aller gemeinschaftlicher Politikbereiche der EU unterliegen einer verbindlichen Begleitung und Bewertung.

So sieht die **FTE-Politik** der Europäischen Union Monitoring und Ex-post-Evaluierungen der einzelnen Programme, "Impact Studies" aller Mitgliedsländer sowie regelmäßige Bewertungen der Rahmenprogramme für Forschung und Technologische Entwicklung selbst vor. Daneben werden methodische Weiterentwicklungen unterstützt und Diskussionen initiiert. Wenngleich auch hier methodische und inhaltliche Probleme (z.B. das Problem der Vergleichbarkeit verschiedener Interventionsformen, die Messung der Effizienz und Wirksamkeit etc.) auftreten, so kann die österreichische FTE-Politik von dieser Praxis sowohl in Bezug auf Abwicklung und Verfahren als auch in der inhaltlichen und methodischen Diskussion lernen.

Daneben demonstriert die vergleichsweise hochentwickelte Evaluierungspraxis im Bereich der gemeinschaftlichen Strukturpolitik das Zusammenwirken von **Ex-ante-Evaluierung (Appraisal)**, **begleitender Evaluierung (Monitoring)** und **Ex-post-Evaluierung** in Kombination mit einer genauen Festlegung von Regelmäßigkeit, Ablauf und institutioneller Verankerung.

Um die Vorzüge eines solchen Systems optimal nutzen zu können, bedarf es folgender methodischer Voraussetzungen:

- Beim Vergleich verschiedener Interventionsformen sollte ein **Mix aussagekräftiger qualitativer und quantitativer Methoden** zur Anwendung kommen, die jedoch nicht zu starren Pseudo-Quantifizierungen führen sollen.

Eine Pseudo-Quantifizierung liegt beispielsweise vor, wenn quantitative Zielgrößen (z.B. Schaffung von 20.000 zusätzlichen Arbeitsplätzen) ohne methodisch fundierte Analyse der Sachlage formuliert werden, sondern nur in Erfüllung der Vorgabe, quantifizierte Ziele in der Programmplanung zu inkludieren. Dies scheint in vielen Programmplanungen im Rahmen der Strukturfondsförderungen der Fall zu sein und sollte vermieden werden.

Ein **sinnvoller methodischer Mix** könnte etwa darin bestehen, qualitative Methoden für die Evaluierung des Programmanagements bzw. der Programmimplementierung einzusetzen, bei der die persönlichen Erfahrungen der im Programm involvierten Personen von Bedeutung sind, quantitative Methoden jedoch für die Schätzung indirekter Wirkungen, etwa die Effekte auf andere Sektoren, Arbeitsplatzwirkungen, Konsequenzen für staatliche Budgets etc.

- Zusätzlich zu der gängigen, rein deskriptiven Interpretation der Indikatoren sollten die Daten des Monitoring-Systems auch zur **Entwicklung quantitativer Modelle** (z.B. Simulationsmodelle oder theoriegeleitete mikroökonomische Modelle) genutzt werden.

Mikroökonomische Modelle versuchen unter anderem, unternehmerisches Handeln auf der Grundlage ökonomischer Rationalität zu erklären und können somit helfen, Ansatzpunkte für das Eingreifen der öffentlichen Hand zur Beeinflussung unternehmerischer Entscheidungen aufzuzeigen. Mikroökonomische Modelle auf Basis von Unternehmensdaten eignen sich zur quantitativen Analyse von Zusammenhängen, die für den Erfolg eines bestimmten FTE-Programme von besonderer Bedeutung sind. So zum Beispiel ist für die Wirksamkeit eines F&E-Förderungsprogramms die Reagibilität der F&E-Investitionen auf die Senkung der (privaten) Kosten von F&E zentral und auch ökonomisch schätzbar.

- Auch sollte das Monitoring-System die Grundlage für die Entwicklung von **Wirkungsindikatoren** sein.

Daten über durchgeführte bzw. geplante kooperative Forschungsprojekte können etwa die Grundlage zur Entwicklung von Indikatoren sein, die den Erfolg eines Programms zur Förderung kooperativer betrieblicher F&E-Aktivitäten anzeigen.

Die Weiterentwicklung quantitativer Modelle sollte in einem solchen Rahmen vor allem dahin gehen, **indirekte Effekte** abschätzen zu können und auch dort Aussagen zu treffen, wo die Interventionssummen im Vergleich zu den Makro-Größen relativ klein sind. Indirekte Effekte umfassen Auswirkungen auf Branchen, die keine Programmförderungen erhalten haben, Beschäftigungseffekte, Produktivitätseffekte etc.

Empfehlung 4: Verfahren verbessern und Verbindlichkeiten herstellen

Bei jeder Evaluierung gibt es ein Spannungsverhältnis zwischen der erforderlichen Objektivität der Evaluatoren und ihrer für die Evaluierung ebenfalls vorteilhaften Vertrautheit mit dem Evaluierungsgegenstand bzw. der institutionellen und politischen Landschaft, in die ein Programm eingebettet ist. Um "den Blick von außen" und "den inneren Blick" zu vereinbaren, ist die Zusammenstellung von Evaluatorenteams vorteilhaft, in denen sowohl ausländische als auch inländische Experten vertreten sind. Solche **gemischte Teams** bieten zudem die Gelegenheit eines Erfahrungsaustauschs zwischen den Experten, der den Aufbau einer österreichischen Evaluierungskultur fördert, die nicht von internationalen Standards abgekoppelt ist.

Voraussetzung für die Entwicklung einer heimischen Evaluierungskultur ist auch, Evaluierungen

- **verbindlich** an die Durchführung von Programmen zu knüpfen, und sie
- **regelmäßig** auszuführen.

Bei der Initiierung eines Programms sollte ein verbindlicher Evaluierungsplan erstellt werden. Auch Forschungs- bzw. Förderinstitutionen müssen regelmäßig evaluiert werden. Diese Regelmäßigkeit und Verbindlichkeit *sollte aber nicht zu einer Evaluierungsroutine im negativen Sinne führen, die durch ein automatisiertes, an den spezifischen Evaluierungskriterien orientiertes Verhalten seitens der evaluierten Programmverantwortlichen bzw. Institutionen gekennzeichnet ist.* Dieser Gefahr muß bei der Planung des Evaluierungsdesigns Rechnung getragen werden, etwa mit unangekündigten Evaluierungen, periodischem Wechsel der Evaluierungsteams, Variationen des Evaluierungsdesigns oder Kontrolle der Empfehlungen, die sich aus der Evaluierung ergeben haben.

Damit Evaluierungen sowohl ihre Kontroll- als auch Lernfunktion erfüllen können, bedarf es einer **engen Abstimmung zwischen Programmerstellung, Umsetzung, Begleitung und Bewertung.** Dies hat verschiedene Facetten:

- Erstens zeigt sich, daß qualitativ hochwertige, richtungsweisende Evaluierungen umso eher realisiert werden können, wenn bereits im Stadium der Programmerstellung Ziele, Intentionen und erwünschte Wirkungen der Programme klar und möglichst widerspruchsfrei spezifiziert werden. Gute Evaluierungen benötigen eine **avancierte, informierte Nachfrage** und umgekehrt.

- Die Formulierung **operationaler Ziele** ist ein weiteres Beispiel für ein "evaluierungsfreundliches" Programmdesign. Operationale Ziele sind eine unabdingbare Voraussetzung für eine sinnvolle Bewertung der Zielerreichung. Da viele FTE-Projekte jedoch mit großer Unsicherheit behaftet sind, können ihre Resultate ex-ante nur schwer abgeschätzt werden, woraus sich große Unterschiede zwischen Ex-ante-Einschätzungen und Ex-post-Ergebnissen ergeben können.
- Weiters soll die Programmbegleitung - im Idealfall in Form eines Monitorings - die Bewertung unterstützen. Dies betrifft unter anderem die **Weitergabe von Daten** und Informationen an die Evaluatoren. In diesem Zusammenhang empfiehlt es sich, die Verpflichtung der geförderten Institutionen und intermediären Einrichtungen zur Informationsweitergabe für Zwecke der Evaluierung in den Förderrichtlinien festzuschreiben.

Empfehlung 5: Methoden weiterentwickeln, um Qualität zu erhöhen

Um die Qualität von Evaluierungen sicherzustellen, müssen auch bei den Evaluierungsmethoden Fortschritte gemacht werden, und zwar mit Bezug auf:

- Indikatoren
- qualitative und quantitative empirische Methoden
- Anwendung neuer ökonomischer Erkenntnisse
- erfolgsorientiertes Evaluieren.

Um die Meßbarkeit der Programmwirkungen und damit auch die Feststellung der Zielerreichung zu erleichtern, ist der Einsatz technologierelevanter Indikatoren notwendig. Es gilt deshalb, **die in Österreich unbefriedigende Situation in bezug auf die Verfügbarkeit technologierelevanter Indikatoren zu verbessern**. Selbst zum internationalen Standard zählende Indikatoren fehlen entweder gänzlich oder werden nur mit großer zeitlicher Verzögerung veröffentlicht. Zur Behebung dieses Informationsdefizits ist einerseits die österreichische FTE-Statistik deutlich zu verbessern und andererseits - in Abstimmung mit internationalen Aktivitäten (z.B. OECD) - auch die Entwicklung neuer, aussagekräftigerer FTE-Indikatoren zu forcieren. Positiv zu vermerken ist, daß es in Österreich auch im internationalen Vergleich beachtliche Erfahrungen bei der Konstruktion und Analyse verschiedener Wettbewerbsindikatoren auf Basis von Außenhandelsdaten (z.B. "unit values") gibt. Dieses Potential ist bisher im Zusammenhang mit FTE-Programmevaluierungen allerdings noch nicht genützt worden.

Gerade im Bereich des technologischen Wandels sind die derzeit vorhandenen empirischen Methoden noch bei weitem nicht ausgereift. Um die Sinnhaftigkeit von FTE-Programmen beurteilen zu können, sind robuste, präzise und zumindest teilweise standardisierte quantitative Methoden erforderlich. Die Eigenschaften einzelner Modelle hängen dabei erstens von Eigenschaften der Methoden ab, auf denen sie basieren: So zum Beispiel gilt die Input-Output Analyse als robuster als ein allgemeines Gleichgewichtsmodell, welches wiederum den Wirtschaftskreislauf vollständiger beschreibt. Zweitens sind die modellspezifischen Charakteristika von Bedeutung. Im Sinne einer laufenden Verbesserung der Evaluierungspraxis, mit der ein gezielterer Einsatz öffentlicher Mittel möglich wird, ist die ständige Weiterentwicklung **quantitativer wie qualitativer Methoden** gefragt, die einerseits zur prognostizierenden Beurteilung geplanter, andererseits zur Ex-post Bewertung abgeschlossener Programme verwendet werden können.

Es kann festgestellt werden, daß eine Reihe neuerer Entwicklungen der ökonomischen Theorie bisher kaum in die Evaluierungspraxis Eingang gefunden haben. Beispiele sind etwa die umfangreichen Arbeiten zu Forschungsk Kooperationen in der neueren Industrieökonomie oder die Analyse von unvollständiger/asymmetrischer Information oder von ökonomischen Anreizen in der Informationsökonomie. Selbst wichtige quantitative Untersuchungen von FTE-Programmen entstanden quasi außerhalb des Evaluierungs-"Mainstream". Das Defizit im Bereich der **Verknüpfung neuer ökonomischer Erkenntnisse** (z.B. aus der Informations- und Industrieökonomie) **mit der Evaluierungspraxis** sollte daher im Rahmen der Evaluierungsforschung verringert werden. Nur aus einem tiefen Verständnis theoretischer Zusammenhänge zwischen staatlichen Eingriffen und ökonomischen und technischen Entwicklungen heraus kann es gelingen, die Erfolgsquote von FTE-Programmen durch Anpassung ihres Designs zu verbessern.

Nach Analyse der FTE-Programmevaluierungspraxis kann auch festgestellt werden, daß die Evaluierung von Programmdesigns viel weiter entwickelt ist und öfter praktiziert wird als die Evaluierung der Wirkungen eines Programms. Der Grund dafür liegt meist in den methodischen Schwierigkeiten, die mit der Durchführung einer Wirkungsanalyse verknüpft sind. Geht man davon aus, daß auch FTE-Programme dem allgemeinen Legitimationszwang staatlichen Handelns unterliegen, müssen **Evaluierungen solcher Programme in Zukunft stärker erfolgsorientiert angelegt werden**. Die Messung der direkten und indirekten Programmwirkungen ist ein unverzichtbares Element einer solchen Erfolgskontrolle, wobei der Schwerpunkt auf der Messung ökonomischer Effekte liegen sollte.

Literatur

Aiginger, K., "Creating a Dynamically Competitive Economy", in Devine, P., Katsoulacos, Y., Sugden, R. (Eds.), *Competitiveness, Subsidiarity and Industrial Policy*, Routledge, London - New York, 1996, 121-146.

Aiginger, K., Bayer, K., "Die TOP-Aktion - Eine neue Form der Investitionsförderung", *WIFO-Monatsberichte*, 1982, 55(10), 594-605

Aiginger, K., Hutschenreiter, G., "Förderungsberichte als Instrumente der Effizienzkontrolle", in: Gantner, M. (Hrsg.), *Handbuch des öffentlichen Haushaltswesens*, Manz, Wien, 1991, 183-201.

ATMOS, *Entscheidungsgrundlagen für die Schwerpunktspolitik des Innovations- und Technologiefonds für die beginnenden 90er Jahre*, Seibersdorf, 1990.

Axelrod, R., *The Evolution of Co-operation*, Basic Books, New York, 1984.

Becher, G. und Kuhlmann, S. (Hsg.), *Evaluation of Technology Policy Programmes in Germany*, Kluwer Academic Publishers, Boston, Dordrecht, London, 1995.

Beirat für Wirtschafts- und Sozialfragen, *Wirtschaftsstandort Österreich*, Wien, 1994.

Berk, R. A., *Thinking about Program Evaluation*, Sage Publications, Newbury Park, 1990.

Capron, H., (Hsg.), *Proceedings of the Workshop on Quantitative Evaluation of the Impact of R&D Programmes*, EC MONITOR-SPEAR Programme, Luxemburg, 1993.

Capron, H., *Economic Quantitative Methods for the Evaluation of R&D Programmes: A State of Art*, Luxemburg, 1992.

Chen H. T., *Theory-driven Evaluations*, Sage Publications, Newbury Park, 1990.

Cohen, L., "When Can Governments Subsidize Research Joint Ventures? Politics, Economics, and Limits to Technology Policy", *American Economic Review, Papers and Proceedings*, 1994, 84 (May), 159-163.

Cordes, J.J., "Tax Incentives and R&D Spending: A Review of the Evidence", *Research Policy*, 1989, (18), 119-133.

d'Aspremont, C., Jaquemin, A., "Cooperative and Noncooperative R&D in Duopoly with Spillovers", *American Economic Review*, 1988, 78(5), 1133-1137.

De Bandt, J., "Research and innovation: Evaluation problems and procedures at different levels", *International Journal of Technology Management*, 1995, 10 (4-6), 365-377.

Fritsch, M., "Förderprogramme", *Discussion Paper der Bergakademie Freiberg*, 1993, Technical University, Faculty of Economics and Business, Freiberg.

Geroski, P., "Markets for Technology: Knowledge, Innovation and Appropriability", in Stoneman, P. (Ed.), *Handbook of the Economics of Innovation and Technical Change*, Blackwell, Oxford, 1995, 90-131.

- Gibbons, R., *A Primer in Game Theory*, Harvester Wheatsheaf, New York, 1992.
- Grimmer, K. (Hsg.), *Politische Techniksteuerung*, Leske & Budrich, Opladen, 1992.
- Gual, J., "The Three Common Policies: An Economic Analysis", in Buiges, P., Jacquemin, A., Sapir, A. (Eds.), Buiges, P., Jacquemin, A., Sapir, A. (Eds.), *European Policies on Competition, Trade and Industry. Conflict and Complementarities*, Edward Elgar, Aldershot, 1995, 3-48.
- Hahn, F., *Die Fördertätigkeit des ERP-Fonds*, Österreichisches Institut für Wirtschaftsforschung, Wien, 1991.
- Hall, B. H. (1992A), "Investment and Research and Development at the Firm Level: Does the Source of Financing Matter?", NBER Working Paper, 1992, (4096).
- Hall, B. H. (1992B), "R&D Tax Policy During the Eighties: Success or Failure?", NBER Working Paper, 1992, (4240).
- Harhoff, D., *Zur steuerlichen Behandlung von Forschungs- und Entwicklungsaufwendungen. Eine internationale Bestandsaufnahme*, ZEW Dokumentation, 1994, (94-02).
- Hellstern, G.M. (Hsg.), *Handbuch zur Evaluierungsforschung*, Schriften des Zentralinstituts für Sozialwissenschaftliche Forschung der Freien Universität Berlin.
- Holzinger, E., *Programm Evaluation: Patentrezept für Effizienz?*, Forschungsbericht, Österreichisches Institut für Raumplanung, Wien 1994.
- Hornschild, K. (Projektleiter), Meyer-Krahmer, F., Becher, G., Gielow, G., Kuntze, U., *Wirkungsanalyse der Forschungspersonal-Zuwachsförderung*, Deutsches Institut für Wirtschaftsforschung, Beiträge zur Strukturforschung, 1990(115).
- Hutschenreiter, G. (1993A), *Die "Regionale Innovationsprämie". Eine begleitende Evaluierung*, Österreichisches Institut für Wirtschaftsforschung, Wien, 1993.
- Hutschenreiter, G. (1993B), *Neue Ansätze zu einer Industriepolitik der EG*, WIFO-Monatsberichte, 1993, 66(5).
- Hutschenreiter, G., "Intersektorale und Internationale 'F&E-Spillovers'", WIFO-Monatsberichte, 1995, 68(6), 419-427.
- Hutschenreiter, G. (Projektleiter) et al., *Evaluierung der Technologieförderungsprogramme der Bundesregierung 1985/1987*, Österreichisches Institut für Wirtschaftsforschung, Wien, 1991.
- Irwin, D.A., Klenow, P.J., "High Tech R&D Subsidies: Estimating the Effects of Sematech", *Journal of International Economics*, 1996, 40, 323-344.
- Kastrinos, N., "The Impact of EC R&D Programmes on Corporate R&D Practice in Europe: A Knowledge-base Approach", *R&D Management*, 1995, 25(3), 269-279.
- Katz, M.L., Ordover, J., "R&D Cooperation and Competition", *Brookings Papers on Economic Activity. Microeconomics*, 1990, 137-203.
- Kealy, T., *The Economic Laws of Scientific Research*, MacMillan, London, 1996.

- Knoll, N., Orthofer, R., Polt, W., Programmanagement ITF-Schwerpunkt "Umwelttechnik". Stufe 1: Themebewertung, Forschungszentrum Seibersdorf, 1995.
- Kreps, D.M., A Course in Microeconomic Theory, Harvester Wheatsheaf, New York, 1990.
- Krugman, P., Obstfeld, M., International Economics. Theory and Practice, 2nd ed., Harper Collins, New York, 1991.
- Krugman, P., Pop Internationalism, MIT Press, Cambridge, Mass., 1996.
- Kuhlmann, S., Holland, D., Evaluation von Technologiepolitik in Deutschland, Physica-Verlag, Heidelberg, 1995
- Leo, H., Die österreichische Weltraumindustrie. Ausmaß und Perspektiven, Österreichisches Institut für Wirtschaftsforschung, Wien, 1991.
- Leo, H., Knoll, N., Ohler, F., Peneder, M., Telekommunikation im Umbruch. Innovation - Regulierung - Wettbewerb, tip, Wien, 1994.
- Leyden, D.P., Link, A.N., Government's Role in Innovation, Kluwer Academic Publishers, Dordrecht, 1992. OECD, Impacts of National Technology Programmes, Paris, 1995.
- Mansfield, E., et al., "Social and Private Returns from Industrial Innovation", Quarterly Journal of Economics, 1977.
- Mas-Colell, A., Whinston, M.D., Green, J.R., Microeconomic Theory, Oxford University Press, New York - Oxford, 1995.
- Meyer-Krahmer, F., "Technology policy evaluation in Germany", International Journal of Technology Management, 1995, 10(4-6), 601-621.
- Milgrom, P., Roberts, J., Economics, Organization and Management, Prentice Hall, Englewood Cliffs, 1992.
- Mowery, D. "The Practice of Technology Policy", in Stoneman, P. (Ed.), Handbook of the Economics of Innovation and Technical Change, Blackwell, Oxford, 1995, 513-557.
- O.Ö. Umweltakademie - Institut für angewandte Umweltforschung, Öko-Audit. Evaluierung der ITF-Pilotförderung, Bundesministerium für Wissenschaft, Verkehr und Kunst, o.J.
- OECD (1995A), Impacts of National Technology Programmes, Paris, 1995.
- OECD (1995B), National Systems for Financing Innovations, Paris, 1995.
- OECD (1995C), Impacts of National Technology Programmes, Paris, 1995.
- OECD, (1995D), Main Science and Technology Indicators, Paris, 1995.
- OECD, Evaluation of Research: A Selection of Current Practices, Paris, 1987.
- OECD, Transparency for Positive Adjustment. Identifying and Evaluating Government Intervention, Paris, 1983.

Ohler, F., "Lessons on Methodology from the Austrian Impact Study", Plattform Technologie - Evaluierung, 1997, (2).

Ohler, F., Hutschenreiter, G., Mühlberger, M., Entscheidungsgrundlagen für die Bildung eines Schwerpunktes Energietechnik im Rahmen des Innovations- und Technologiefonds (ITF), Seibersdorf, 1992.

Ohler, F., Jörg, L., Polt, W., Guy, K., Hutschenreiter, G., Husz, M., Sieber, A., Gluske, H., Patsios, S., Evaluation of the Austrian Participation in Community RTD Programmes, Seibersdorf, 1997.

Ohler, F., Katzenberger, W.E., Ortner, G., Polt, W., Stepan, A., Empfehlungen zur Schwerpunktsetzung Softwaretechnik, Seibersdorf, 1992.

Ohler, F., Leo, H., Schweiger, W., Stepan, A., Entscheidungsgrundlagen für die Bildung eines Schwerpunktes Industrial Design im Rahmen des Innovations- und Technologiefonds (ITF), Seibersdorf, 1992.

Patton, M.Q., How to Use Qualitative Methods in Evaluation, Sage Publications, Newbury Park, 1988.

Pavitt, K., "The Road to Ruin", New Scientist, 3 August 1996, 32-35.

Phlips, L., Competition Policy: A Game-Theoretic Perspective, Cambridge University Press, Cambridge, 1996.

Polt, W., Buchinger, E., Jörg, L., Kopcsa, A., Leo, H., Mustonen, L., Ohler, F., Patsios, S., Evaluierung des ITF-Förderschwerpunktes "Flexible Computerintegrierte Fertigung" (FlexCIM), Österreichisches Forschungszentrum Seibersdorf und Österreichisches Institut für Wirtschaftsforschung, Seibersdorf, 1994.

Romer, P.M., "Implementing a National Technology Strategy with Self-Organizing Investment Boards", Brookings Papers on Economic Activity. Microeconomics, 1993, (2), 345-397.

Rossi, P.H., Freeman H.E., Evaluation: A Systematic Approach, Sage Publications, Newbury Park, 1994.

Scriven, M., Evaluation Thesaurus, Sage Publications, Newbury Park, 1991.

Shadish, W. R., et. al., Foundations of Program Evaluation: Theories of Practice", Sage Publications, Newbury Park, 1991.

Steiner, M., Sturn, D., "Elements of Evaluation of Science and Technology Policy in Austria", Research Evaluation, 1995, 5(1), 98-108.

Stiglitz, J., Whither Socialism?, MIT Press, Cambridge, Mass., 1994.

Szopo, P., Aiginger, K., Lehner, G., Investitionsförderung in Österreich, Österreichisches Institut für Wirtschaftsforschung, Wien, 1985.

Szopo, P., Direkte Wirtschaftsförderung in Österreich: Reformimpulse durch Budgetkonsolidierung und EG-Integration, Österreichisches Institut für Wirtschaftsforschung, Wien, 1990.

Szopo, P., Subventionen in Österreich, Österreichische Strukturberichterstattung - Spezialbericht 1985, Österreichisches Institut für Wirtschaftsforschung, Wien, 1986.

Taucher, "Prüfung des Einsatzes des Steuerrechts für die indirekte Förderung von F&E nach dem EU-Beitritt Österreichs", Finanz-Journal, 1994, (12), 274-280.

Tirole, J., The Theory of Industrial Organization, MIT Press, Cambridge, Mass., 1988.

Triconsult, Seed-financing. Experten und Unternehmerbefragung, Wien, 1991.

Anhang

Selektion von Förderungsschwerpunkten

Verschiedene theoretische und praktische Modelle versuchen, bei der Selektion von Förderungsschwerpunkten Konkurrenzmechanismen oder kompetitive Marktmechanismen wirksam werden zu lassen. Ein Beispiel dafür ist Paul Romers (1993) Vorschlag zu "self-organizing industry investment boards". Grundsätzlich geht es Romer darum, den Konflikt zwischen koordinierten Aktionen zur Beschleunigung des Innovations- und Wachstumsprozesses einerseits und den "kollektivem Handeln" innewohnenden Risiken der Ineffizienz andererseits zu entschärfen. Diese Risiken ergeben sich aus dem politischen Prozeß (wie etwa die "positive Theorie der Regulierung" der Chicago-Schule betont) und aus dem Phänomen der asymmetrischen Information.

Durch ein geeignetes institutionelles Arrangement soll für die beteiligten Akteure Anreize geschaffen werden, sich in einer Weise zu verhalten, daß die Wohlfahrtseffekte maximiert werden.

Paul Romers Vorschlag zu "Self-organizing Industry Investment Boards"

Der Vorschlag von Paul Romer (1993), eines der renommiertesten US-Ökonomen der jüngeren Generation, wird hier nur in den Grundzügen dargestellt. Romer selbst hat weitere Details ausgearbeitet und auch potentielle Probleme der Umsetzung seines Vorschlags angeschnitten.

Das von Romer vorgeschlagene Verfahren sieht vor, daß Unternehmen einer Branche, die die Notwendigkeit zur Bereitstellung industriespezifischer öffentlicher Güter sehen, die Einhebung einer zweckgebundenen Steuer petitionieren können. Erst nach Abhaltung einer "Urabstimmung" unter den Mitgliedsfirmen der Branche wird eine Steuer, etwa auf die Umsätze aller Unternehmen der entsprechenden Branche eingehoben. Die Verpflichtung aller Unternehmen - und damit der Einsatz hoheitlicher Befugnisse - ist erforderlich, um Trittbrettfahren zu vermeiden. Die entsprechenden Steuereinnahmen fließen jedoch nicht in die allgemeine Staatskasse, sondern sind ausschließlich für die Finanzierung von Industry Investment Boards (IIBs) reserviert. Diese IIBs konzentrieren sich auf die Bereitstellung bestimmter industriespezifischer öffentlicher Güter (z.B. durch die Finanzierung universitärer Forschung in ausgewählten Bereichen, die Entwicklung im Bereich von Zulieferfirmen etc.). Die IIBs operieren als private, nicht auf Gewinn ausgerichtete Stiftungen, deren Organe von den beitragenden Firmen kontrolliert werden.

Die Entscheidungsbefugnis über die Verwendung ihrer Steuerbeiträge, d.h. über die Allokation der Mittel auf verschiedene IIBs verbleibt ausschließlich bei den Unternehmen. Diese entscheiden in regelmäßigen Abständen, welches der bestehenden oder neu in den "Markt" eintretendes IIBs sie mit ihren Beiträgen unterstützen. Neue IIBs treten nach Absolvierung eines entsprechenden Zulassungsverfahrens mit den eingesessenen IIBs in Konkurrenz um die verfügbaren Beiträge der Unternehmen.

Nicht nur die von den IIBs zur Bereitstellung öffentlicher Güter beauftragten Institutionen (Unternehmen, Universitäten, Forschungseinrichtungen etc.) stehen daher im Wettbewerb um die bereitstehenden Mittel, sondern die Verwendungszwecke (repräsentiert durch die IIBs) selbst stehen in Konkurrenz zueinander. Mangelnde Attraktivität von IIBs wird sich in sinkenden Beiträgen niederschlagen und IIBs können letztlich auch gänzlich aus dem Markt ausscheiden. Umgekehrt besteht für die beteiligten Unternehmen ein Anreiz, IIBs mit dem größten Ertrag zu fördern.