

VORWORT

In dieser dritten Ausgabe des Newsletters Plattform Technologieevaluierung möchten wir Ihnen zwei Beiträge vorstellen:

Der erste Artikel bietet eine Zusammenfassung des Forschungsauftrages "Evaluierung von FTE-Programmen: Zwischen Best Practice Entwicklung und österreichischen Strukturen" im Auftrag des Bundesministerium für Wissenschaft und Verkehr. Besonderes Augenmerk ist auf die Empfehlungen für Österreich zu richten.

Als zweiten Beitrag drucken wir, mit freundlicher Genehmigung der Zeitschrift *Spektrum der Wissenschaft*, einen Aufsatz von Dr. Stefan Kuhlmann (ISI-Karlsruhe) ab, in dem Bewertungsgrundsätze für die wichtige Evaluierung technologiopolitischer Programme erläutert werden. Eine Präsentation des Aufsatzes fand im Februar dieses Jahres in Wien statt.

1997 werden weitere Ausgaben des Newsletters mit Schwerpunkt auf konkreten Evaluierungsvorhaben sowohl innerhalb als auch außerhalb des Innovations- und Technologiefonds erscheinen.

PREFACE

This third edition of the newsletter presents two papers: The first summarizes the results of a research project on the evaluation of RTD-programs, with a focus on the Austrian experience, commissioned by the Austrian Ministry for Science and Transport. The second paper is a reprint of an essay by Stefan Kuhlmann (ISI-Karlsruhe), which describes basic principles to be followed in the evaluation of RTD-programs. The upcoming 1997 editions of the newsletter will focus on individual evaluation projects.

Michael Stampfer, BMWV

Gernot Hutschenreiter, WIFO

Oliver Fritz, Joanneum Research

EVALUATION OF RTD-PROGRAMS: BEST PRACTICE CONSIDERATIONS AND THE AUSTRIAN EXPERIENCE

By Oliver Fritz, Gernot Hutschenreiter and Dorothea Sturn

INTRODUCTION

This paper summarizes a study¹ that was commissioned by the Austrian Ministry for Science and Transport to be carried out jointly by Joanneum Research and the Austrian Institute of Economic Research. Its goal was to discuss fundamental issues of program evaluations, to analyze the current evaluation practice with respect to RTD-programs in Austria and finally to give recommendations directed towards further improvement of this practice.

The first part of the paper discusses the contents of a program evaluation as well as some empirical methods that can be applied in this context. Furthermore, some of the specific characteristics of RTD-programs that should be considered in an evaluation of such programs are described. The Austrian practice of evaluating RTD-programs is discussed in the second part. Finally, based on this evidence and the conclusions from the first part, some recommendations are made as how to advance the Austrian evaluation practice.

CONTENTS OF A PROGRAM EVALUATION

Before turning to a discussion of the contents of a program evaluation, it should be pointed out that in every evaluation there is a tradeoff between adapting the evaluation design to the specific characteristics of the program and the need to apply standardized evaluation methodology. While standardization ensures that evaluations are comparable and of high quality, flexibility is needed so that the evaluation results meet the information needs of those in charge of the program. Since complete standardization is not achievable, a discussion of the contents of a program evaluation can only provide guidelines along which the evaluators select those issues they consider most appropriate to accomplish their specific tasks.

¹ Fritz O., Hutschenreiter G., Sturn D., Evaluierung von FTE-Programmen: Zwischen Best Practice Entwicklung und österreichischen Strukturen. Studie im Auftrag des Bundesministeriums für Wissenschaft, Verkehr und Kunst, Wien, 1997.

A program consists of four main elements: The basic concept underlying the program, the design of the program, its implementation, and finally the effects induced by the program. The evaluation of these elements will now be discussed.

Public programs are created to solve a problem society has to cope with. In the area of technology policy, the nature of such problems may comprise, for example, the insufficient diffusion of an important technology, low levels of R&D activities in the industrial sector or barriers for the creation of new products. Within the **evaluation of the program concept** the existence of the program has to be justified, based on the nature of the underlying problem and the type of intervention that is applied. The mere presence of a social problem is not sufficient reason for public action. Such action is only justified if the welfare improvement the intervention induces exceeds its costs.

The identification of the main factors causing the problem is one of the keys for a successful intervention. Programs are often based on pure speculation about these factors, which are never scrutinized empirically. Any assumptions that are explicitly or implicitly contained in the program and have not been put to test should be assessed by the evaluators. Furthermore, the legitimacy and feasibility of the program goals have to be checked and potential tradeoffs with other policy goals, e.g. budgetary restrictions, have to be revealed. The evaluators should examine whether by implementing the program these goals are likely to be accomplished within a certain period of time.

The **evaluation of the program design** focuses on the specific layout of the program. This includes: All actions to be carried out as well as their expected direct and indirect effects; the organizational framework; and operational goals, which ought to be explicit, specific, measurable and to be accomplished by a given deadline. Such goals are often hard to set due to the nature of the subject which may restrict measurability, or due to the uncertainty about the effects of a program. While program planners and managers therefore tend to reject operational goals, which they often conceive as arbitrary, program evaluators require operational goals in order to determine the effectiveness of a program.

The actual **implementation** of a program should also be subjected to an evaluation. The evaluators must identify any deviations of the actual implementation from the program plan, as well as the causes for these deviations. Furthermore, it must be evaluated if the program is executed efficiently and effectively. The implementation is considered to be effective if all program management tasks are accomplished. It is efficient if these tasks are accomplished at the lowest possible cost.

Certainly the most challenging evaluation assignment concerns the **estimation of a program's effects**, which are key to its success or failure. Both the effects on the members of the target group and those that go beyond („spillovers“) should be measured. These spillovers are often the true motivation for carrying out the program. Some of the observed effects will be undesired and may be added to the program's cost. As many types of effects as possible - economic, technological, environmental, social or other - should be taken into account.

Short run effects have to be separated from medium and long run effects: Research and development activities may lead to one or more scientific articles in the short run, to patents in the medium run but only in the long run to new products or processes. This delay of some of the most important effects poses considerable difficulties to any evaluation effort. In order to measure these effects, the evaluation would have to be carried out long after the program expired. At that point, the evaluation results will have lost their relevance in the political decision-making process. Additionally, as time goes by, the evidence for a causal link between intervention and effects becomes more and more blurred.

The effectiveness and efficiency of a program can be determined based on the estimated effects. Effectiveness is defined as the extent to which the program accomplished its goals; efficiency, on the other hand, is based on the comparison of the benefits derived from the program (i.e. the positive effects) with its costs, including among others administrative cost, payments to the participants and any negative effects.

METHODS APPLIED IN A PROGRAM EVALUATION

Empirical methods used in social and other sciences can be classified in various ways. Based on Holzinger's (1993) classification, we distinguish two groups of methods:

Quantitative methods include standardized measuring tools like standardized questionnaires or statistical, econometric or mathematical models. They provide results that can be compared and generalized. **Qualitative methods**, on the other hand, provide detailed information on the subject of the analysis without following a predefined, standardized design. This group includes, for example, interviews that are not based on questionnaires, group discussions and observations.

Which methods are applied depends on the characteristics of the specific program. Before selecting a method and in order to interpret the results correctly the evaluator must understand the advantages and disadvantages the use of each of these methods implies.

Quantitative methods are regularly used to test hypotheses in an ex-ante evaluation or to estimate and compare program effects and determine causalities in an ex-post (or sometimes interim) evaluation. Mathematical models in particular provide an abstract description of reality, in which those variables considered crucial for the analysis are linked together. Such models help to understand complex processes by isolating their most significant elements and links; at the same time, however, they might lead to misinterpretations by omitting crucial variables or misspecifying the links between variables. Even though such models provide clear, standardized and comparable results, these results need to be carefully interpreted by the evaluator. All assumptions which affect the outcome significantly should be pointed out as well as any data deficiencies.

Qualitative methods are frequently applied in the evaluation of the program's implementation and the related organizational and managerial issues, and when target groups are small. They allow the evaluator to analyze the subject in all details, taking account of its complex nature. By paying too much attention to the details, however, the evaluator might easily miss the critical issues. Unlike quantitative methods, qualitative ones are able to deal with variables that cannot be easily measured. Additionally, the results are not based on any assumptions. The sometimes ambiguous nature of the qualitative results, however, leave much room for interpretation. In particular, these results cannot be generalized and are difficult to compare with the findings of other evaluations.

SPECIAL ASPECTS TO BE CONSIDERED IN THE EVALUATION OF RTD-PROGRAMS

Programs that focus on R&D and technological issues differ from other programs in various aspects relevant for the evaluation of these programs. In particular, their effects on the participants of the program are only a small part of their total effects, so that, due to spillovers, their social returns are considerably larger than their private returns. If, for example, the competitiveness of a specific sector increases due to a program promoting R&D activities, all its supplying sectors will benefit from the program as well. The evaluation of the effects of RTD-programs should therefore not be restricted to the participants but has to go beyond. In order to capture the economic effects, for instance, multisectoral or macroeconomic models are needed.

The estimation of the social returns needs to take into account the following factors:

- **Unintended effects**, e.g. a reduction in the level of competition due to a program promoting firm cooperation with respect to R&D activities.
- **Network externalities**, which imply that the effects of a program depend on the number of participants.
- **Learning phenomena**; the adaptation of one technology by a firm, for example, makes the successful adaptation of other technologies by that firm more likely.
- **Delayed effects**: Many of the crucial effects of RTD-programs (e.g. the development of new products as a result of increasing R&D activities) occur much after the program expired and are therefore out of reach for most evaluations.
- **Uncertainties**, implying that social returns cannot be foreseen at the start of the program.

EVALUATION OF RTD-PROGRAMS IN AUSTRIA

Until now, the Austrian public RTD support system is only incompletely covered by evaluations. Some of the basic elements of this system have not been subjected to systematic evaluations or public debates. In fact, most of the evaluation efforts carried out so far in Austria were concentrated on targeted technology programs. However, targeted technology programs constitute just one part of the total RTD support systems, other major elements being the Industrial Research Promotion Funds (FFF), technology-related support activities of the ERP Fund and - sometimes forgotten - tax incentives for R&D.

Targeted technology programs do not have a long tradition in Austria. The pioneering activity in this field was the implementation of the Federal Government's technology programs (1985-87) targeted at information technology, biotechnology and CAD/CAM. These programs were the subject of the first comprehensive RTD program evaluation study in Austria (Hutschenreiter et al., 1991). In 1987, the Innovation and Technology Funds (ITF) was established. Since then, more than a dozen - mostly relatively small - technology programs were financed through the ITF. Studies commissioned in preparation of these programs (ex ante) together with the ex-post evaluations of the technology programs 1985-87 and of the Austrian participation in European Space Agency programs (Leo, 1991) constituted the first generation of RTD program evaluation studies in Austria.

Now, after ten years of operation of the ITF, several ex-post evaluations of specific programs are in preparation or being carried out, among them the current "mixed-team" evaluation of the environment technology program conducted by ISI and WIFO. Thus, the establishment of the ITF continues to exert substantial impact on evaluation and related activities in Austria. In contrast to its specific programs, the ITF has not been evaluated as an *institution* so far. The same applies to the other major RTD support institutions in Austria. The assessment of their activities as a whole should be the focus of coming Austrian evaluation efforts.

As noted above, there are some "blind spots" in the Austrian RTD support system as far as public scrutiny is concerned. One of the remarkable facts is the complete absence of attempts to evaluate the impact of the indirect R&D support through tax benefits in Austria. This fact cannot be explained by the low scale of the schemes in operation. Indeed, according to official estimates the tax expenditure associated with tax incentives to R&D amounts to ATS 450 million annually. This brings it to a range comparable to the total annual funds of the ITF. Recently, a number of proposals have been made to extend and/or reform the indirect support for R&D. In this context, an ex-post evaluation of the current practice and an ex-ante evaluation of new instruments or schemes is strongly recommended.

The Austrian participation in international RTD programs has an impact both on evaluators and on users of evaluation studies such as technology policy makers. In particular, the evaluation of European RTD programs is of strategic importance for the development of evaluation practices in Austria:

- Firstly, it can be expected that the participation in European evaluation efforts will lead to strengthening the "institutionalization" of evaluations in Austria.
- Secondly, due to the European Commission's sustained support for the improvement of evaluation procedures and the establishment of international networks of evaluators (e.g. in the framework of MONITOR/SPEAR), the Austrian involvement in European evaluation practices will lead to a transfer of know-how to Austria.

In this context, an important trade-off arises: Austria is a small country with corporatist traditions and a rather strong persistence of institutions characterized by long-lasting and close institutional and personal relations. This is also true for the small-sized Austrian RTD sector with its limited number of actors. As a consequence, Austrian evaluators have inside knowledge of institutional and political contexts which is of considerable advantage

in the implementation of evaluations. On the other hand, the same factors may lead to undesirable cooperative behavior.

Under these conditions it appears appropriate to involve international evaluators. At the same time, Austria should utilize the institutional knowledge of its experts and not forgo the creation of domestic capabilities in an important field of research and expertise. A feasible solution to this dilemma is the involvement of mixed teams of evaluators consisting of international and Austrian experts. This approach was chosen in the first evaluation of the Austrian participation in the EU Framework Programs presented recently (Ohler et. al., 1997).

RECOMMENDATIONS

1. Evaluation of the entire RTD system

In Austria, RTD program evaluation is focused on ITF technology programs. This reflects the fact that the implementation of targeted RTD programs is in practice confined to the ITF. It is therefore recommended

- to emphasize the evaluation of the *institutions* concerned (i.e. institution for technology promotion)
- to extend evaluations to also cover *indirect support for R&D* so that the entire system of RTD support is assessed.

2. Make use of the current cumulation of evaluation activities

Currently, several ITF programs are being evaluated or will be evaluated soon and an evaluation of the FFF is in planning. This cumulation of evaluation activities should be used to advance the Austrian evaluation practice by discussing results and methodological foundation of these evaluations among experts, policy makers and program planers.

3. Learn from the EU evaluation practice

All EU-programs are evaluated on a regular basis. With respect to RTD-programs, the EU carries out monitoring and ex-post evaluations, impact studies and regular evaluations of the R&D framework programs and supports evaluation research. Austrian RTD-policy makers and evaluation experts should learn from these activities.

A rather sophisticated evaluation framework has been developed for EU structural programs, which are subjected to regular and mandatory appraisals, monitoring and ex-post evaluations. In order to take advantage of such a system, the following methodological requirements have to be met:

- the application of a *useful mix of qualitative and quantitative methods*;
- setting *arbitrary operational goals*, that were not derived from an analysis of the problem and whose only purpose is to meet required evaluation standards, *should be avoided*.
- the *data from the monitoring system* should be used to develop quantitative models like simulation models or microeconomic models.
- this data should also be used to *develop indicators* measuring the effects of a program.

4. Improve procedures and make evaluations a requirement

High quality evaluations require both inside knowledge of the political, economic and social system in which a program is embedded and a distant look at that system. Therefore, „*mixed teams*“ of evaluators which consist of local and foreign experts are recommended.

Programs as well as institutions must be evaluated on a *regular and mandatory* basis. However, regular evaluations, should not result in „evaluation routines“ on part of the programs and institutions to be evaluated. Such routines are characterized by a conduct merely oriented towards meeting the evaluation criteria.

Evaluations have a control as well as a learning function. In order to fulfill both functions, *program planning and implementation, monitoring and evaluation have to be coordinated*. This requires that during the program planning general and operational goals and the expected effects of the program must be defined. Furthermore, data from the monitoring should be made available to the evaluators. It is also recommended that both program participants and intermediary institutions should be required to make any relevant information available.

5. Improve methods in order to increase quality

Evaluation methods should be further advanced with respect to:

- *indicators*: the Austrian RTD-data base has to be improved and new indicators developed.

- *qualitative and quantitative empirical methods*: robust, precise and partially standardized quantitative methods are needed to be able to measure the success or failure of programs based on their total effects.
- *evaluation criteria based on total program effects* should have priority over managerial or organizational criteria or results from participant surveys.
- *the use of the latest findings in economic research* for evaluation purposes.

References:

Holzinger, E., "Programm Evaluation. Theoretische Grundlagen und Anwendungsmöglichkeiten in Raumordnung und Regionalpolitik." Schriften zur Regionalpolitik und Raumplanung, Heft 23, Wien, 1993.

Hutschenreiter, G. (Project Leader) et al., Evaluierung der Technologieförderungsprogramme der Bundesregierung 1985/87 (Evaluation of the Technology Support Programs of the Federal Government), Austrian Institute of Economic Research (WIFO), Vienna, 1991.

Leo, H., Die österreichische Weltraumindustrie. Ausmaß und Perspektiven (The Austrian Space Industry. Scale and Perspectives), Austrian Institute of Economic Research (WIFO), Vienna, 1991.

Ohler, F. (Project Leader) et al., Evaluation of the Austrian Participation in Community RTD Programs, Seibersdorf, 1997.

Dr. Oliver Fritz, Institute of Technology and Regional Policy, Joanneum Research, Wiedner Hauptstraße 76, A-1040 Vienna. Phone ++43-1-5817520-15; fax -20; e-mail: oliver.fritz@joanneum.ac.at

Mag. Gernot Hutschenreiter, Austrian Institute of Economic Research (WIFO), P.O. Box 91, A-1103 Vienna, Austria. Phone ++43-1-7982601-238; fax ++ 43-1--7989386; e-mail: hutsch@wsr.ac.at

Dr. Dorothea Sturn, Institute of Technology and Regional Policy, Joanneum Research, Wiedner Hauptstraße 76, A-1040 Vienna. Phone ++43-1-5817520-18; fax -20; e-mail: dorothea.sturn@joanneum.ac.at

WIE BEWERTET MAN FORSCHUNGS- UND TECHNOLOGIEPOLITISCHE PROGRAMME²?

Eine Analyse bisher durchgeführter Evaluationen zeigt die Kriterien auf, mit denen sich Nutzen und Wirkungen staatlicher Forschungsförderung umfassend bewerten lassen.

Von Stefan Kuhlmann

Hat je eine wissenschaftliche Studie unabhängiger Experten ein Programm der Forschungsförderung umsteuern können? Tatsächlich, auf verschiedene Weise. Im Jahre 1989 wurde sogar einmal eine beabsichtigte Fördermaßnahme gestoppt: Ein Gutachten des Deutschen Instituts für Wirtschaftsforschung empfahl damals, eine vom Bundesministerium für Forschung und Technologie erwogene finanzielle Förderung der Verbreitung von Faserverbundwerkstoffen in der mittelständischen Industrie zurückzustellen - weder die Technologieentwicklung noch die Industrie wurden seinerzeit als dafür reif genug bewertet. Das Gutachten hat verhindert, daß viele Millionen Mark quasi in den Sand gesetzt wurden.

Sind solche unabhängigen Bewertungen demnach nützlich? Die Forschungs- und Technologiepolitik wird immer unübersichtlicher, so viele Akteure sind hier mittlerweile tätig: die Bundesministerien, die relativ eigenständigen großen Forschungs- und Förderungsorganisationen, die Landesregierungen und immer weitreichender auch die Europäische Union. Sie alle bieten Fördermaßnahmen für Forschung und Technologieentwicklung in Wissenschaft und Industrie. Seit Anfang der neunziger Jahre allerdings stagnieren die staatlichen Mittel. Die Öffentlichkeit und auch Experten haben den ökonomischen oder gesellschaftlichen Nutzen so mancher Fördermaßnahme in Zweifel gezogen. Andererseits fordern Industrie und Wissenschaft immer lauter ein wieder deutlich verstärktes staatliches Engagement bei der Förderung von Forschung und Technologie.

Vor diesem Hintergrund steigt in jüngster Zeit das Interesse an Bewertungen - den Evaluationen - des Nutzens und der Wirkungen staatlicher Forschungsförderung. Das heutige Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) ließ 1992/93 vom Fraunhofer-Institut Systemtechnik und Innovationsforschung (ISI) in Karlsruhe mehr als 50 Evaluationen seiner Förderprogramme im Rahmen einer sogenannten Metaevaluation kritisch analysieren (siehe Stefan Kuhlmann / Doris Holland:

² Nachdruck, mit freundlicher Genehmigung, aus: Spektrum der Wissenschaft, Januar 1997, 110-112

"Evaluation von Technologiepolitik in Deutschland. Konzepte, Anwendung, Perspektiven", Physica-Verlag, Heidelberg 1995).

Arten der Evaluation

Evaluationsstudien im hier verstandenen Sinne sollen die Wirkungen von forschungs- und technologiepolitischen Maßnahmen feststellen, um Planungs-, Entscheidungs- und Bewertungsprozesse auf politischer Ebene unterstützen zu können.

Man kann grundsätzlich zurückschauende (ex post-) Evaluationen, ein Förderprogramm begleitende Studien sowie vorausschauende (ex ante-) Analysen unterscheiden. Ex post-Evaluationen sollen erzielte Wirkungen feststellen und bewerten, sollen helfen zu lernen, sollen aber auch die staatlichen Fördermaßnahmen nachträglich legitimieren. Begleitende Studien sollen das Management laufender Programme unterstützen und gegebenenfalls Kurskorrekturen ermöglichen. Strategische ex ante-Analysen sollen vorausschauend die möglichen Wirkungen unterschiedlicher technologiepolitischer Eingriffe bewerten; operationale ex ante-Studien prüfen die Erreichbarkeit bereits festgelegter Ziele bei variierenden Programmgestaltungen.

Alle diese Studien sind als wissenschaftlich begründete empirische Untersuchungen angelegt - und dürfen darum nicht mit den im Forschungsbetrieb üblichen Begutachtungen der wissenschaftlichen Qualität durch Fachkollegen (peer review) verwechselt werden.

Die Mehrheit der Förderbereiche weist eine geringe Evaluationspraxis auf. Relativ häufig wurden bisher die Bereiche marktnahe industrielle Forschung und Entwicklung, Technologie-Transfer und Technik-Diffusion, also die breiten- und öffentlichkeitswirksamen Maßnahmen, evaluiert; seit dem Ende der achtziger Jahre ließ der Bund verstärkt aber auch Programme analysieren, die Forschung und Entwicklung ausgewählter Technologielinien (wie Lasertechnologie und Mikrosystemtechnik) stärken wollen; in jüngster Zeit begann man auch mit Bewertungen im Bereich der Vorsorgeforschung (etwa der Gesundheitsforschung). Demgegenüber wurden die Förderung von Grundlagenforschung und die institutionelle Förderung von Forschungseinrichtungen bisher kaum durch unabhängige Studien bewertet.

Leitfragen

Die Beurteilung von Fördermaßnahmen ist ein schwieriges und heikles Unterfangen: In Deutschland verfolgt man mit forschungs- und technologiepolitische Programmen häufig

strategische Ziele, etwa die Steigerung des technologischen Niveaus von Forschungseinrichtungen oder Unternehmen; die politischen Absichten sind also eher vage formuliert. Dies unterscheidet sie von Maßnahmen, die einem bestimmten Zweck dienen, etwa einen Menschen auf den Mond zu bringen oder die Häufigkeit von Krebserkrankungen zu senken. Programmevaluationen müssen deshalb sozusagen mehrdimensional angelegt werden: Ein aus der Sicht von Evaluationsexperten akzeptables Konzept umfaßt fünf Leitfragen, die befriedigend zu beantworten sind, wenn eine Fördermaßnahme positiv bewertet werden soll:

- Ist das Programm geeignet? Sind die zugrundeliegenden Annahmen richtig?
- Wird die Zielgruppe erreicht?
- Welche direkten und indirekten Wirkungen gibt es oder sind absehbar?
- Wurden Ziele erreicht?
- Sind Implementation und Verwaltung effizient?

Diese Fragen werden nicht einfach nur den Nutznießern und Verwaltern eines Förderprogramms gestellt, sondern - im Idealfall - durch ein ausgefeiltes sozial- und wirtschaftswissenschaftliches Instrumentarium untermauert: Methodisch gehören dazu Vorher-/Nachher-Vergleiche oder der Vergleich von geförderten und nicht geförderten Forschungsprojekten. Analysiert werden solche Daten im Rahmen von Fallstudien, durch Indikatorenbildung, Kosten-/Nutzen-Analysen und ökonometrische Modellbildung. Die dafür benötigten Informationen und Daten gewinnt man in Befragungen und aus statistischen Quellen (etwa aus Forschungsaufwand, Patentstatistiken und Expertendaten).

Metaevaluation

Das ISI hat in seiner Metaevaluation 50 Evaluationsstudien externer Experten daraufhin geprüft, ob und wie diese Leitfragen beantwortet wurden. Untersucht wurden unter anderem

- Evaluationen der Förderung mittelfristiger technologischer Entwicklungslinien (wie Lasertechnologie, Materialforschung und Mikrosystemtechnik);
- Evaluationen der Förderung technisch-organisatorischen Wandels (insbesondere in der Industrie, beispielsweise Humanisierung des Arbeitslebens/Technik und Arbeit);
- Evaluationen der Förderung industrieller Innovation (insbesondere bei kleinen und

- mittleren Unternehmen, beispielsweise Auftragsforschung und -entwicklung);
- Evaluationen experimenteller Programme (beispielsweise Modellversuche technologieorientierter Unternehmensgründungen).

Zusammenfassend lassen sich die seit 1985 durchgeführten Evaluationsstudien in folgender Weise charakterisieren:

Evaluationslaufzeiten: Die Mehrheit der Studien dauerte etwa zwei Jahre. Das erscheint zunächst viel, ist aber angesichts der weitreichenden Fragestellungen akzeptabel. Einige sehr kurzläufige Studien gingen nur sehr eingeschränkten Fragestellungen nach und können kaum als Evaluation bezeichnet werden.

Evaluationstyp: Reine ex post-, begleitende und ex ante-Analysen gibt es kaum. Oft kommen die verschiedenen Typen gemischt vor. Dies weist auf kombinierte Verwendungszwecke hin.

Zielgruppe einer Fördermaßnahme: Fast alle Studien haben relativ ausführlich die Situation beziehungsweise das Verhalten der potentiellen oder tatsächlichen Zielgruppen forschungs- und technologiepolitischer Fördermaßnahmen analysiert.

Erreichter Stand von Forschung und Technologie im Bereich einer Fördermaßnahme, möglichst im internationalen Vergleich: Diese Analyse kann Aufschluß geben, ob eine weitere staatliche Förderung von Forschung und Technologie überhaupt erforderlich ist. Nur etwa die Hälfte der Studien, in denen eine solche Fragestellung sinnvoll war, enthält entsprechende Untersuchungen.

Technologieverbreitung: In knapp der Hälfte der Fälle, in denen Aufschluß über die erreichte oder zu erwartende industrielle Verbreitung neuer Technologien erforderlich gewesen wäre, wurden ausführliche Diffusionsanalysen durchgeführt.

Daten: Nur wenige Studien gründen ihre Analysen auf eine Kombination von statistischen Daten (wie Forschungs-, Patent- oder Handelsstatistiken) und empirischen Erhebungen (etwa Ergebnissen eigener Umfragen und Daten über Antragsteller). Die Mehrheit der Studien basiert ausschließlich oder überwiegend auf Ergebnissen eigener Erhebungen.

Veröffentlichung: Die Mehrzahl der Studien ist für die breite Öffentlichkeit nur schwer zugänglich; die Abschlußberichte liegen als sogenannte graue Literatur oder sogar nur als regierungsinternes Dokument vor. Nur knapp ein Viertel aller Studien wurde bisher

als Buch veröffentlicht (und erscheint damit in den Katalogen von Bibliotheken und Buchhändlern).

Die Durchsicht der 50 Evaluationsstudien hinterließ bei uns ein widersprüchliches Bild: Ein großer Teil dieser Studien wurde systematisch und gründlich durchgeführt und liefert bemerkenswerte Einblicke in die Verwendung und die Wirkung staatlicher Forschungsförderung, andere hätten aber durchaus konsequenter und kritischer betrieben werden können.

Folgerungen

Als Ergebnis dieser Durchsicht und auf der Grundlage vertiefender Fallstudien gewannen wir viele Hinweise für eine künftige, effektivere Evaluationspraxis:

Seit langem fordert die Evaluationsforschung die systematische Prüfung der Annahmen über wirtschaftliche oder technologische Probleme, die eine staatliche Maßnahme begründen. In der Praxis erweist sich diese Annahmenprüfung immer wieder als äußerst schwierig und wird häufig nur sehr unzureichend realisiert. Zum Teil ist das eine Konsequenz ungelöster methodischer Probleme, zum Teil drückt sich hier aber auch eine Vermeidungshaltung der Programmgestalter, teils auch der Evaluatoren aus: Die kompromißlose Frage nach dem Sinn einer Fördermaßnahme könnte unangenehme, radikale Antworten bewirken. Solange Programmverantwortliche zugleich Auftraggeber von Evaluationsstudien sind, besteht hier ein systematischer Interessenkonflikt.

Ein verstärkter Einsatz von querschnittartigen Evaluationskonzepten bietet sich vor allem für die übergreifende Planung der mittelfristigen Forschungsförderung an. Horizontale Studien können den Interessen und Aktivitäten verschiedenartiger forschungs- und technologiepolitischer Akteure, Adressaten und Betroffener Rechnung tragen und die zunehmende Überlassung, Verflechtung und Wechselwirkungen von Forschungs- und Technologiegebieten thematisieren. Bisher gibt es erst wenige Beispiele für solche Studien.

Die Verknüpfung von Wirkungsanalysen laufender Fördermaßnahmen mit vorausschauenden Analysen des technologischen, ökonomischen oder sozialen Potentials neuer Forschungs- und Technologiegebiete wie der 1993 durchgeführten Delphi-Studie zur künftigen Entwicklung von Wissenschaft und Technik könnte hilfreich für eine verbesserte Ausrichtung forschungs- und technologiepolitischen Handelns sein.

Kombinierte Evaluationen von institutioneller Forschungsförderung und Förderprogrammen werden künftig an Bedeutung gewinnen, nicht zuletzt weil knappe staatliche Budgets gezielt eingesetzt werden müssen. Methodisch ist hier das Problem angemessener Bewertung der Leistungen von Forschungsinstitutionen zu bewältigen, eine Aufgabe, die bisher von den großen Forschungsorganisationen wie Wissenschaftsrat, Max-Planck-Gesellschaft und Fraunhofer-Gesellschaft in Eigenregie wahrgenommen wurde. Künftig sind konkurrierende Evaluationsbemühungen verschiedener Akteure im Forschungs- und Technologie-System durchaus denkbar.

Die Bedeutung von Gutachtergremien und von Bewertungsverfahren nach dem Modell des peer review für die Interessenformulierung in der Förderung darf nicht unterschätzt werden; sie werden - vor allem im Bereich der Grundlagenforschung - immer ein gestaltender Faktor bleiben. Die Schnittstelle solcher Bewertungsverfahren zu den professionellen Evaluationsstudien ist bisher aber nicht befriedigend gestaltet.

Viel zu tun gibt es bei der Verbesserung der Nutzerfreundlichkeit von Evaluationen. Im Prinzip könnten unterschiedliche Interessen an einer Bewertung (etwa aus Parlament, Ministerien und organisierten Interessen aus Wissenschaft, Wirtschaft und Gesellschaft) bewußt in eine Evaluationsstudie aufgenommen und damit mögliche Widersprüche produktiv genutzt werden. Dabei sollte gelten: Evaluationsstudien werden durchgeführt, um forschungs- und technologiepolitische Optionen und ihre eingetretenen oder potentiellen Wirkungen freizulegen. Deshalb sollen sie auch politische Handlungsempfehlungen enthalten; eine echte Unterstützung im Diskussions-, Planungs- und Entscheidungsprozeß leisten bevorzugt solche Empfehlungen, die gleichzeitig mehrere alternative Pfade präsentieren. Von elementarer Bedeutung ist hierbei: Wenn die Ergebnisse der Analysen andere politische Schlußfolgerungen nahelegen, als der Finanzier einer Studie erwartet oder wünscht, muß gewährleistet sein, daß solche unbequemen Voten dennoch nachdrücklich vorgetragen und auch publiziert werden können. Bringen unsere forschungspolitischen Institutionen die dazu erforderliche Stärke auf? Dies ist eine Frage der politischen Kultur.

*Dr. Stefan Kuhlmann, Fraunhofer-Institute Systems and Innovation Research (ISI),
Breslauer Straße 48, D-76139 Karlsruhe, Phone ++49-721-6809-170/2; fax -260; E-mail:
sk@isi.fhg.de*

Für den Inhalt verantwortlich:

Bundesministerium für Wissenschaft und
Verkehr - Dr. Michael Stampfer
Renngasse 5, A-1010 Wien
E-mail: michael.stampfer@bmwv.ada.at

weitere Kontaktadressen:

Mag. Gernot Hutschenreiter
WIFO
Arsenal Objekt 20, PF 91
A-1031 Wien
E-mail: hutsch@wsr.ac.at

Dr. Oliver Fritz
Joanneum Research
Wiedner Hauptstraße 76
A-1040 Wien
E-mail: oliver.fritz@joanneum.ac.at

"Plattform Technologieevaluierung" ist ein unregelmäßig erscheinendes offenes Forum zur Diskussion methodischer und inhaltlicher Evaluierungsfragen in der Technologiepolitik.