Directorate for Science, Technology and Innovation www.oecd.org/sti

Enhancing Research Performance through Evaluation, Impact Assessment and Priority Setting

TABLE OF CONTENTS

Boxes

EXECUTIVE SUMMARY

Evaluation of publicly funded research has become a central concern of policy makers for two main reasons. First, there is growing demand for evidence-based policies and for evaluation of the results of public investments. More precisely, governments increasingly seek to determine how much they should invest in science and technology (S&T), research and development (R&D), and innovation. They wish to know where to invest and what society gets in return. Ideally, evaluation should help determine the economic effects of both public investment in R&D and innovation and the social impacts. Policy makers also increasingly want public investment to help meet global challenges, such as energy, security and climate change.

Second, the demand for evaluation has expanded because OECD countries have substantially increased public investment in R&D despite budget constraints. Governments not only finance R&D in various sectors of performance such as the business sector or the higher education sector, they also fund the performance of R&D on their own behalf. Government budget appropriations or outlays for R&D (GBAORD) measures the funds committed by federal/central governments for R&D. In aggregate, these have been climbing faster than GDP across the OECD, by 6.4% annually since 2001. In addition to direct support, governments also finance business R&D indirectly through the use of tax incentives. Although they may be significant, the costs of these tax credits, in terms of foregone revenue, do not usually appear as R&D support in government budgets.

The demand for effective evaluation tools to inform decisions on research funding and impacts will continue to increase in line with public investment in R&D and innovation as countries try to enhance competitiveness and improve innovation capacity. The need for evaluation is also likely to increase because of demands for greater accountability and effectiveness. Evaluation assists governments in their decisions to prioritise resources for R&D and innovation and can help them design research programmes. Moreover, it enhances public accountability, creates a better-informed society, and raises awareness of the contribution of research to a country's economic and social development. The demand for evaluation is also changing owing to increased interest in evaluating entire research systems and research portfolios. This greater emphasis on evaluation, however, has raised a number of important conceptual and methodological challenges.

The report has two parts. Part I (Chapters 1 to 4) first sets the stage by discussing key issues in the evaluation of public science and technology research performance. It presenting evaluation as a means of addressing both the demand for accountability on the part of various stakeholders and the need to consider the public value of such research. The report next discusses three specific evaluation contexts: *i)* expert reviews, *ii)* impact assessments, and *iii)* priority-setting activities. The first two deal more specifically with the evaluation of research programmes, projects and policies. The third demonstrates another possible area for the application of evaluation processes.

Chapter 1 presents key issues in the evaluation of S&T and sets the stage for addressing specific topics in more detail: the architecture of evaluation systems insofar as they exist, peer review as a tool for evaluation, and the interplay between evaluation and priority setting. It begins by recalling the emergence of evaluation of public research, from its initial focus on improving the quality of research by evaluating discrete outcomes of research (*e.g.* publications) to more complex attempts to evaluate the economic and non-economic social effects of research as well as environment/ecology/systemic effects. It discusses diverse approaches to evaluation, most of which use economic frameworks. While these have many advantages, they have limitations, particularly in relation to the impact of research on social change.

Evaluation is considered from the various perspectives of governments, civil society, research institutions, funding agencies, evaluators and social scientists. Finally, the chapter draws on recent history to assess the social impacts of research and describes new theoretical and empirical efforts (*e.g.* the science of science policy).

Chapter 2 assesses the role of "expert reviews" – a broader type of "peer review" – in shaping funding decisions (*e.g.* what, who and where to fund research) and in evaluating the output and impacts of public investment in R&D and innovation. Expert review is one of the methods most commonly used to evaluate public funding of S&T. Although it has many merits, including low cost and easy applicability, it currently faces its strongest challenges in several decades. Externally, there is some evidence that political decision makers question how well any type of peer review can address socioeconomic and political priorities. Internally, there is evidence of hollowing out as increasing pressures on researchers' time make it more difficult to find experts willing to undertake reviews.

Chapter 3 presents findings of the TIP Impact Assessment project that relate to the rationale, methods and tools for evaluating public R&D at country or economy level and at the institutional and programme levels. It discusses key methodological challenges and suggests ways to improve impact assessments based on country experiences.

Chapter 4 analyses priority setting and the role of evaluation in designing policy. Priority setting is the conscious selection of activities at the expense of others. It affects resource allocation. Here, priority setting differs from the type of priority setting that takes place in a self-organising system. Priority setting and evaluation are two distinct issues. The priority-setting process has several dimensions, the importance of which varies over time. The chapter examines, in particular, the links between *ex ante* and *ex post* evaluation in priority setting and decision making and seeks to identify best practices for improving both the quality of *ex ante* evaluations and their usefulness in the policy-making process, notably in priority setting. It also assesses the process of priority setting in S&T itself and identifies structural weaknesses as well as best practices.

Part II presents six case studies of systems for evaluating S&T programmes, projects and policies in China (Chapter 5), Finland (Chapter 6), Israel (Chapter 7), Norway (Chapter 8), Japan (Chapter 9) and Austria (Chapter 10). The case studies are organised so as to give the reader first the feel of a broader assessment of national evaluation systems before dealing with specific issues of evaluation methodology, processes, and scope and some current problems. The case studies are varied with regard to the comprehensiveness of the evaluation systems, the objectives of evaluations, and how they are conducted and used for policy making. Many of the issues discussed in previous chapters are addressed in the case studies.

The case studies show that in all countries, the rationale for evaluation is to improve R&D performance and to ensure efficient use of the natural and monetary resources invested. The roles of ministries differ, however, among the countries. While ministries play an active role in Japan and China, evaluation is delegated to research councils or other funding bodies in the other countries. In several countries, evaluation is also grounded in legal frameworks and regulations at the national government, funding agency or research council level.

The case studies illustrate differences in the comprehensiveness of evaluation systems, the objectives of evaluations, and the way they are conducted and used for policy making. In particular, they map institutional frameworks, actors, regulations and practices for designing and implementing evaluations (*ex ante* or *ex post*) in the area of S&T. Some case studies also offer suggestions or guidelines for improving the current system. For all six countries, the rationale for evaluation is to improve R&D performance and to ensure efficient use of the resources invested. The roles of ministries differ and the evaluation mechanisms and tools adopted also differ depending on the level of evaluation and country-specific factors. However, addressing social and economic impacts remains a challenge for most countries, mainly owing to the difficulty of attributing impact to a specific research programme, particularly after a lapse of several years. Scientific impact is easier to measure than social impact, and the impact of applied research is more easily measured than that of basic research. The outcomes of evaluations appear to have the greatest influence on improving the design, implementation and effectiveness of programmes, and a more limited impact on policy decisions for funding, with the exception of evaluations of basic research programmes.

The key messages of this report can be summarised as follows:

- ***There is growing demand for evaluation of S&T***. Demand for evaluation is increasing and changing from evaluating the quality of research (via peer review) to assessing the outcome, output and impact of public R&D. There is also increasing interest in evaluating entire research systems and research portfolios and in enhancing the role of evaluation as a tool for priority setting and decision making.
- ***The peer review process remains a fundamental mechanism for research planning, implementation and evaluation***. Peer review is under stress, but solutions exist (*e.g.* extended peer review processes involving non-scientific stakeholders; combing metrics and indicators). There is a need to facilitate and improve the internationalisation of peer review because of the increase in international research collaboration.
- ***Evaluation of social/economic impacts requires new metrics and approaches.*** In addition to new indicators and methodologies, assessing social/economic impacts requires stakeholder involvement as well as new communication channels (to decision makers, to agents, to stakeholders). Impact assessment must also balance the tension between (scientific) relevance and social/economic impacts.
- ***Evaluation capacity remains weak and fragmented.*** Evaluation is functionally fragmented in many countries, but elements of a system appear to be in place. These elements include a clear role for outside evaluators and stakeholders outside research establishments, the importance of setting standards in evaluation, and the role of self-evaluation. A system of evaluation also requires establishing follow-up processes on the consequences of evaluations.
- ***Evaluation can no longer be done solely in a national perspective.*** International comparisons are increasingly used in national policy analysis (*e.g.* FP7 participation rankings, University of Shanghai rankings). International benchmarks are necessary for public policy evaluations, especially in areas in which countries compete with others.
- ***Usefulness of evaluation for policy making varies but is often limited.*** Evaluations can be used for improving project management as well as for strategy. The use of feedback from evaluations can be constrained by a lack of data, negative findings or a lack of political buy-in by the leadership.
- ***Priority setting and evaluation interact but remain distinct dimensions of policy making.*** Priority setting is a more complex exercise, involving a broader range of actors (*e.g.* funding agencies) and relying on different approaches and methodologies. This increases the need for mechanisms to ensure greater coherence between priority setting and policy making.
- ***The conceptual underpinning of priority setting remains quite weak and expert opinion continues to predominate.*** Improving the process of priority setting through the use of *ex ante* evaluations requires political buy-in, commitment to invest in resources and develop skills, and the creation of indicators and data to monitor policy or programme effectiveness.

- ***Impact assessment is part and parcel of evaluation***. The rationale for impact assessment is becoming clearer as are policy needs for assessing the impacts of strategic research undertaken outside the context of mainstream public research in universities and government labs.

- ***Measuring impacts is neither straightforward nor easy especially as regards demonstrating causality.*** Many of the dimensions through which S&T affects society (*e.g.* societal, cultural or environmental impacts) are not easily captured by existing national statistical frameworks, as they are difficult to measure or evaluate. This makes it difficult to link social impacts to policy interventions. Nevertheless, methodological difficulties should not stand in the way of efforts to measure economic impacts.

- ***The choice of methodology is context-specific.*** Impact assessment methodologies are not universally applicable, but depend on the objective of the impact assessment exercise, its timing (*ex ante* and/or *ex post*); and the scope and nature of R&D.

- ***New and useful practices are being developed but methodological issues and (international) comparability remain challenges.*** The various methodologies are still evolving, but none of the available techniques has been able to capture the full range of the impacts of public R&D on society.

- ***The international community and the OECD in particular can play a role in diffusing good practices and stimulating policy learning.*** Evaluation is an increasing challenge, not only in technical and methodological terms, but also in terms of the political challenges. International co-operation can help improve countries' ability to effectively develop and foster the use of *ex ante* and *ex post* evaluation for research and innovation policy.

PART I

ISSUES IN SCIENCE AND TECHNOLOGY EVALUATION: PEER REVIEW, IMPACT ASSESSMENT AND PRIORITY SETTING

CHAPTER 1

EVALUATING THE SOCIAL VALUE OF RESEARCH[1]

> This chapter discusses the recent history of evaluation of public research and new theoretical and methodological efforts, including efforts to assess the social and economic impacts of science. It highlights current approaches to evaluating public value and the social impacts of research, including the process known as public value mapping.

Introduction

Despite three decades of progress in the ability to conceptualise, measure and evaluate research impacts, there remains a gaping hole in research evaluation methods and techniques: the ability to evaluate the *social and public value impacts* of research. Professional researchers have developed powerful tools to measure the economic impacts of research, sophisticated bibliometric tools to measure the impact of research outputs on scientific fields and the course of science and technology, and peer review techniques to assess projects, programmes and proposals. But there has been remarkably little progress in the ability to measure directly, systematically and validly the impacts of research on social change. Many scientists have spoken eloquently about the communal and cultural value of scientific knowledge. Possible cultural aesthetics notwithstanding**,** the value of science is inherently in its application. Without rejecting compelling arguments for the intrinsic value of research in intellectual, cultural and aesthetic terms (Fischer, 1997), most policy makers and citizens seem to agree that the chief purpose of public funding of research is to improve the quality of life (Johnson, 1965). Views of how to improve the quality of life through the use of technology, however, differ (*i.e.* weapons, automobiles, space exploration, etc.). Instead of arguing that the purpose of public funding and public programmes is to improve the quality of life, it may be better to state that funding and programmes are guided by visions of change (Pawson and Tilley, 1997). By formulating purpose in terms of change rather than quality of life, most people's goals for public research can be encompassed.

The critical problem for understanding the social impacts of science is the lack of satisfactory analytical tools for understanding the causal impact and the magnitude of the effects of research on social change. This is true whether the researcher defines social change broadly or narrowly. This is not surprising when one considers the relatively recent development of research evaluation. This chapter highlights current approaches to evaluating the public value and social impacts of research. It considers: *i)* why such approaches are needed and how they relate to current approaches to research evaluation; *ii)* special difficulties or challenges for developing such approaches; and *iii)* possible methods to use,

---

1. Barry Bozeman is Regents' Professor and Ander Crenshaw Professor of Public Policy, University of Georgia, and Distinguished Researcher and Professor of Public Affairs, Consortium for Science Policy and Outcomes and School of Public Affairs, Arizona State University. The author acknowledges the support of the US National Science Foundation's "Science of Science Policy" programme and is grateful for the assistance and ideas of the members of the Public Value Mapping project: Dan Sarewitz (principal investigator), David Guston, Catherine Slade, Ryan Meyer, Erik Fisher, Genevieve Maricle, Walter Valdivia and Stephanie Moulton. The views expressed are those of the author and do not necessarily reflect those of the US National Science Foundation, the OECD, Arizona State University or the University of Georgia.

including current "public value mapping of science outcomes".[2] Before turning to these questions, the recent history of analytical methods of research evaluation is briefly considered.

Historical roots of research evaluation

As recently as the 1980s, research evaluation was a field with few practitioners and mostly focused on the economic evaluation of industrial firms' internal rate of return.[3] In fact only the Canadian government[4] and some European countries[5] had begun a systematic evaluation of publicly funded research. Many other countries, such as the United States, did not consider evaluation of the impact of public research a field but instead an agglomeration of fragmented, largely isolated works, many of which were not published.

To understand the historical roots of research evaluation one should consider the state of the art as reported in one of the earliest reviews of the evaluation of publicly funded research. The intention of Salasin, Hattery and Ramsay's *The Evaluation of Federal Research Programs* (1980) was to "identify useful approaches for evaluating R&D programs conducted and sponsored by the federal government" (p. 1). In pursuit of that objective they interviewed more than 200 evaluation experts, most of them based in industry. They cited 49 papers, including one journal article (Rubenstein, 1976) and one book (Andrews, 1979) which focused explicitly on the systematic evaluation of government-sponsored R&D. The monograph identified four problems endemic to evaluating government R&D impacts: *i)* lack of a straightforward definition of effectiveness; *ii)* multiple and competing objectives; *iii)* problems in aggregating products and results, especially across programmes; and *iv)* reconciling political and scientific measures of success – a list that would be just as good today. The monograph concluded with a problem identified by many of the experts consulted: "It is not clear that it is possible to assess research quality based on the immediate outputs of a research project (*e.g.* reports or journal publications)" (Salasin, Hattery and Ramsay, 1980, p. 62). This means that to properly evaluate research requires a broader treatment of impacts. Today, there are numerous studies and methods of R&D evaluation, but most of the problems identified nearly three decades ago in that pioneering monograph remain, particularly those arising from a focus on discrete R&D outputs.

From the 1990s, documents on the practice of evaluating R&D were published. In 1993, Bozeman and Melkers (1993) edited *Evaluating R&D Impacts: Methods and Practice*, an R&D evaluation primer with contributions by leading authorities on such topics as case studies of R&D projects, rate of return on R&D investments, and operations research approaches. At about the same time, the Critical Technologies Institute of the RAND Corporation published a report prepared for the Office of Science and Technology Policy reviewing methods for evaluating fundamental science (Cozzens *et al.*, 1994). One of the earliest OECD research evaluation monographs was produced in the mid-1990s (OECD, 1995). In short, as of the mid-1990s interest in research evaluation reached an inflection point and the amount of literature, professionalisation, dedicated journal pages and policy applications has since increased substantially.

---

2. The method discussed in this chapter is the work of the author and his research colleagues as part of the US National Science Foundation programme on the science of science policy.

3. The most prominent approach to assessment of S&T policy and research evaluation has been peer review. While recognising that peer review is crucial, the focus here is on systematic and potentially quantitative or mixed-method approaches and does not discuss peer review. Similarly, this chapter does not deal with the many and increasingly useful bibliometic approaches to research evaluation. As the primary concern is social impacts, measures relating to scientific impacts and scientific changes are not addressed.

4. For a history of government-mandated research evaluation in Canada, including research evaluation, see Auditor General (1993). For a history of research evaluation activities in Canada see Barbarie (1993).

5. Several publications provide synoptic reviews of the history and methods of research evaluation in Europe, *e.g.* Luukkonen (2002); Callon, Laredo and Mustar (1997).

The economic basis of the evaluation of research impacts

Each of the three works identified just above provided diverse approaches to evaluation, but most of them adopted an economic framework for analysis. Economic assessments of research and technology generally fall into two related categories: social rate of return and aggregate-level production analysis. Social rate of return approaches can be used in a wide variety of contexts. With respect to research and technology, they attempt to estimate the social benefits that accrue from changes in technology and to relate the value of these benefits to the cost of the investments that produced the changes of interest. Among social rate of return approaches, cost-benefit analysis has been the most commonly used for evaluations at the project and programme level (*e.g.* Link, 1996a; 1996b; Ruegg, 1996; Audretsch *et al.* 2002; Saavedra and Bozeman, 2004). The second category, aggregate-level production analysis is influential for broad economic development policy making. It typically focuses on the contribution of technology to the national or regional economy (*e.g.* Solow, 1957).

Economic approaches to evaluation of research, especially cost-benefit approaches, have strong appeal because they focus on specific science and technology (S&T) outputs such as the number of articles or patents produced in research and development (R&D) projects, the number of jobs created by technology transfer programmes, and contributions of technology-based economic development programmes to regional economies. The utility of these approaches is obvious even to sceptics who question the accuracy of economic approaches to assessing larger social impacts of R&D. The benefits of such approaches to evaluation are explored in more detail elsewhere (Link, 1996b). For present purposes they have certain advantages: they typically yield numerical assessments which are useful in a public policy domain increasingly dominated by "metrics" (Kostoff, 2001); they draw on decades of development of strong economic theories relating to the firm, rational choice and economic growth; many have been quite creative in developing quasi-economic techniques based on preference functions and units that mimic rational economic choice (*e.g.* contingent value analysis; see Cummings and Taylor, 1999). Moreover, even if the chief focus, as here, is the social impact of research, it is demonstrably the case that economic change affects social change. If it is assumed (as is not the case here) that economic growth and development generally lead directly to desirable social outcomes, economic measures become satisfactory indices of social progress.

Despite their many advantages, economics-based approaches to evaluation of research have many limitations for gauging the social impact of research. In particular, cost-benefit or rate of return approaches provide only limited insights on the creation of research capacity or the transformational aspects of research. Most focus on the specific products of research projects, such as journal articles or marketable products. Such a focus works best when there are sharp contours (*e.g.* a single R&D project). However, most social objectives are over-determined, that is, several antecedent factors beyond S&T policy can lead to the social objective. Moreover, despite efforts to consider implications of future streams of benefits, economics-based evaluations tend to be static. They rarely take into consideration the mutability of the "products" evaluated, much less the changes in the persons and institutions producing them. They also tend to give short shrift to the generation of S&T capacity and to the ability to produce sustained knowledge and innovations. And, most importantly for present purposes, many of the social benefits and costs of science and technology are not well accounted for in monetary units.

The inadequacy of economics-based approaches for measuring and providing an understanding of the social impacts of research is a chief concern. There is currently, however, no satisfactory method (except, perhaps, case studies) for making a valid assessment of the impact of research on social change.

Social impact of research: Framing the problem

The need for new approaches to assessing the social impact of research may not be obvious. In a broad sense almost all research has social impact. Most research is socially embedded and the development and transmission of research results are social processes and thus have "social impact". For instance, societies are not concerned about economic development chiefly because it is inherently desirable but because of the social impact that is assumed to accompany it. If research leads to an extra percentage point gain in GDP and at the same time to a reduction in public health, environmental quality or personal security, the benefits are not obvious. With some conspicuous exceptions (*e.g.* funding for astronomy) public investments in research are invariably rationalised by expected improvements in the social well-being of the citizens who provide the revenue to support the research. Countries fund work on the human genome because they expect that the funding will ultimately lead to improvements in their citizens' health. Similarly, support for developing technologies is justified because these technologies are expected to make people better off. To be sure, the logic is somewhat indirect. One might also argue that support for research leads to new technology, which leads to economic growth which, in turn, provides more disposable income to citizens, which they can use to improve their conditions as they deem fit. This linear model does, at least in some circumstances, lead to desirable, even optimal social outcomes, especially if the distribution of the economic outcomes of research are equitable.

An evaluator's concern is with more directly observable social and public impacts and with monitoring the direct effects of research on such impacts. Rather than a linear model, this is more a "churn model" (Bozeman and Rogers, 2002) in which research leads through circuitous routes to dead ends, to positive social outcomes, to negative outcomes, and often to both positive and negative outcomes. The question, then, is "to what extent does research contribute, either positively or negatively, to desired social change and to public value?"

It is easy to identify the difficulties for making a valid assessment of the ultimate social impacts of research but impossible to resolve them completely. Research is often only one factor in determining social outcomes and rarely the most important one. For example, some research helps to reduce disease, but in most cases factors such as life style, economic opportunities and environmental conditions also play an important role. The social benefits of research in terms of disease reduction are often defined by a "best case scenario". However, it is difficult to link many impacts precisely to prior research. Desirable social outcomes such as poverty alleviation, public education, improvements in housing and protection of public safety are generally highly over-determined. Research is one of a great many social, economic and natural determinants. In such circumstances it is virtually impossible to define the contribution of research. Whether one employs standard economics-based approaches such as cost-benefit analysis, social indicators monitoring, social accounting or even in-depth case studies, the attribution of causality for complex social impacts is always fraught with difficulty.

A related problem pertains to the "dependent variables". Determining causation is difficult enough, but the effects are often interwoven in ways that are not obvious and difficult to unravel. Social outcomes occur in clusters. Some obvious examples: new birth control technology reduces unwanted pregnancy *and* gives rise to promiscuity and socially transmitted disease; smoking cessation programmes reduce tobacco-related cancer *and* lead to increased obesity rates; technological innovations lead to increased wealth *and* to greater inequities. In short, when modelling social outcomes from research it is difficult not only to trace cause to effect but also to set boundaries on effects.

The list of obstacles to assessing the social impacts of research is unfortunately far from exhaustive (for a more detailed discussion, see Bozeman, 2007). While working towards finding means of assessing them is daunting, it is important to do so because policy makers will continue to make choices about research funding on the basis of a causal logic. They make assumptions about the effects of those investments on such social outcomes as public health, transport systems, education and wealth creation, often on the basis of limited information. The evidence that can be brought to bear, even when imperfect, is likely an improvement over intuition, habit, rough-hewn ideology and political self-interest, among others.

Assessing the social and public impact of research: the "public value mapping of science" project

Researchers at the Consortium for Science, Policy and Outcomes (CSPO) initially developed the idea of "public value mapping of science". Current work is undertaken by a team of researchers at CSPO, now at Arizona State University.[6] The project is funded by the US National Science Foundation's "Science of Science Policy" programme.

The primary rationales for the public value mapping of science (PVM) are that: *i)* the focus of science policy should be on end-state social goals and public values; and *ii)* current research evaluation and science policy analysis methods and techniques, while useful in many important respects, are not sufficient for analysing the impact of research on public value. Box 1.1 provides further detail on PVM assumptions. It is important to bear in mind that PVM is not, and does not aspire to be, a unified method; it is an approach, or set of approaches.

Box 1.1. Core assumptions of public value mapping

1. PVM is either prospective (analysing planned or projected research activities), "formative" (analysing such activities as they occur), or "summative" (evaluating activities and their impacts after they have occurred).
2. It seeks to take into account the highest order impacts of activities (*i.e.* broad social aggregates) and thus focuses on social indices and social indicators.
3. It is multilevel in its analysis, seeking to show linkages among particular programme activities of an agency or institution, activities of other agencies or institutions, relationships – intended or not – among various institutional actors and their activities.
4. PVM is concerned with understanding the environmental context of research and related programmatic activities, with locating the activities and their institutional actors in terms of other actors in the environment, and with the constraints, opportunities and resources present in the environment.
5. Research in any field by any method is embedded in a social context; in PVM analysis of the social context of research (*i.e.* characteristics of research performers, their attributes and social relations) is part of the analysis.
6. PVM is guided by a "public value model of science outcomes", rather than a market-based or market failure model. PVM explicitly rejects evaluation and assessment based on commodification of research values and outcomes. Market prices are viewed as weak partial indicators of the social value of research and research outcomes. Even as a partial indicator, market value is considered in terms not only of magnitude but also of distribution and equity.
7. Since market value is eschewed in PVM and since social values are not interpersonally transmissible, PVM anchors its outcomes values in a wide range of criteria derived from diverse sources including: official, legitimated statements of policy goals; goals implicit in poorly articulated policy statements; government agencies' goal statements in strategic plans; aggregated statements of value represented in opinion polls; official policy statements by government actors; and official policy statements by relevant non-governmental organisations (NGOs).
8. PVM analyses (maps) the causal logic relating goals (any of the above) to measured and hypothesised impacts and outcomes of science and research activities. When possible, the analysis begins from the causal logic articulated by responsible officials. The causal logic, explicit or implicit, is then considered in relation to various plausible alternative hypotheses and alternative causal logics invented by the analyst.
9. PVM is not an analytical technique or even a set of analytical techniques, but a model that includes a

[6] The project, "Public Value Mapping: Developing a Non-Economic Model of the Social Value of Science and Innovation Policy", began in 2007 and will end in 2010.

guiding theoretical framework (public value theory) and a set of assumptions and procedures. Research techniques employed in PVM depend upon the needs and possibilities afforded by the context of application. The only technical approach used in all applications of PVM is the case study method.

10. After gathering data to test hypotheses about causal logics and outcomes, the hypotheses are tested using appropriate analytical techniques and the impacts and outcomes are measured. The results of the analysis focus on relations among the causal logic, the environmental context, and the measured impacts and outcomes.
11. PVM links impact and outcome measures back to aggregate social indicators or other broad-based, trans-institutional, or trans-research programme measures of social well-being.
12. PVM concludes the analysis with recommendations for possible changes (in research or programme activity, causal logic, implementation) that seem likely to lead to better social outcomes.

Source: Bozeman, B. (2003), "Public Value Mapping of Science Outcomes: Theory and Method", in D. Sarewitz *et al.*, *Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development*, 2(1).

As part of the original project, two "beta tests" were developed as initial PVM applications: one focused on breast cancer research (Gaughan and Bozeman, 2002) and the other on genetically modified crops (Gupta, 2003). In both cases, the analysis involved developing public policy statements about research and innovation goals (as surrogate indicators of public value) and developing indicators to determine the degree to which public value and social outcomes meet those goals. Subsequently, most of the work in developing PVM has aimed at theory building and, more recently (*i.e.* since the outset of the project funded by the National Science Foundation), development of multi-method analytical tools. Two aspects of PVM theory building are particularly relevant. First, a theory of innovation has been developed that matches the aims of PVM, the "churn theory of innovation" (Bozeman and Rogers, 2002). Second, efforts have been made to develop a theory of assessing public value (Bozeman, 2007; Jorgensen and Bozeman, 2007; Bozeman and Moulton, 2008) and to apply the PVM theory in case contexts, including influenza vaccine research, development and commercialisation of S&T, genetic suppression technology for seeds, and climate change technology (Bozeman and Sarewitz, 2005; Feeney and Bozeman, 2008; Bozeman, 2007).

The "building blocks" of PVM theory

New ways of thinking about the social value of knowledge

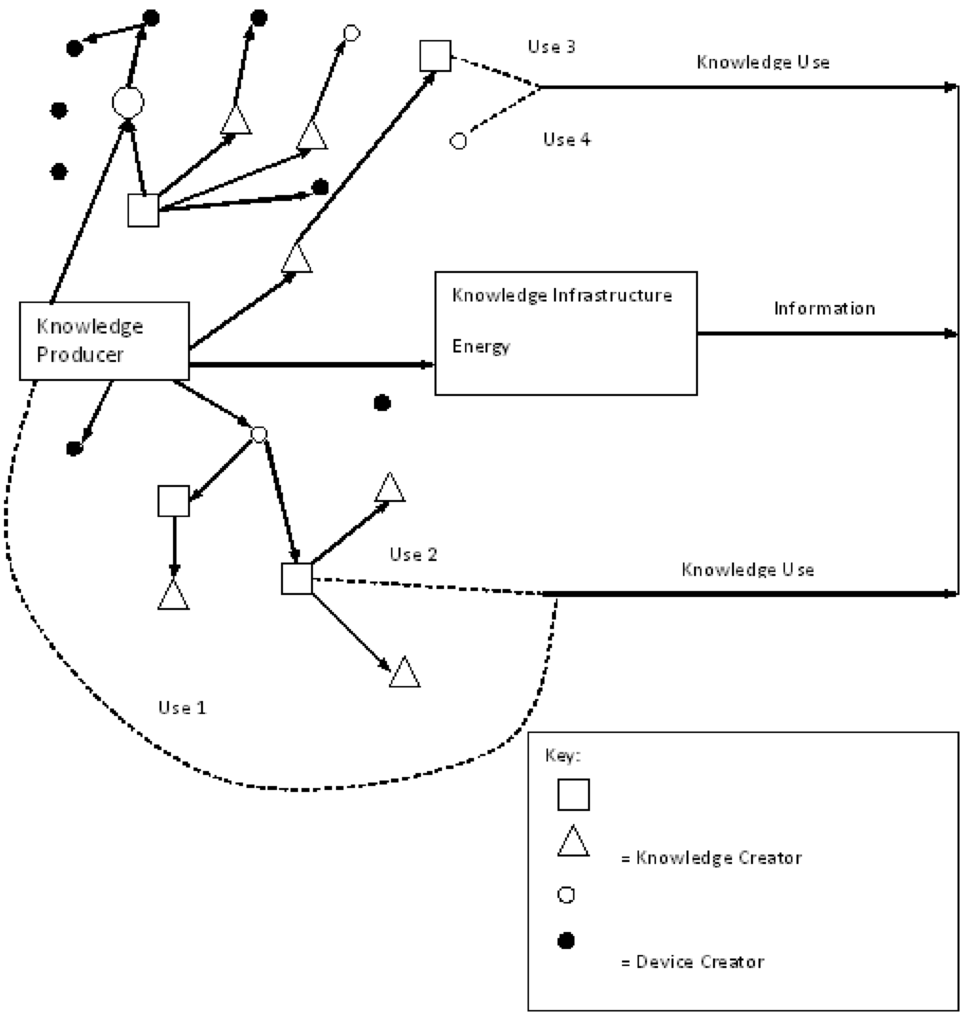
PVM offers a new way to think theoretically about the value of knowledge and its assessment. Philosopher Elizabeth Anderson (1993) presents an especially interesting analysis of economic value and value theory as it pertains to economics. Anderson's position is that economic values are inherently monistic and that this undermines richer and generally more useful pluralistic analyses of value. That is, an analysis that values exchanges, commodities and services on the basis of market standards pre-empts simultaneous reference to other standards such as social impact (Marmolo, 1999; Anderson, 1993).

The monistic nature of economic analysis also fails to view programmes as open systems (Pawson and Tilley, 1997). Pawson and Tilley's "theory-based evaluation" considers that programmes cannot be isolated from the many social factors surrounding them: an S&T programme cannot be separated from its social impacts but a monistic evaluation fails to fully grasp the impacts of the programme because it fails to treat the programme as an open system.

The assertions of Anderson and of Pawson and Tilley have implications for models of innovation and the impact of scientific and technical knowledge. Economists have never made much headway in valuing scientific knowledge (Machlup, 1962) and PVM provides an excellent means of understanding the relations among intrinsic value, economic value and public value.

Figure 1.1 provides a simple depiction of the churn model. It shows that information is created from knowledge and that information can result, via use, in new knowledge or can lie fallow, depending on whether or not it is used. The figure also indicates the possibility of information being put to multiple uses, in each instance creating value as knowledge.

Figure 1.1. Churn model of knowledge use and transformation



New ways of thinking about public values

Certainly there is no single agreed definition of public value, but one definition is closely related to PVM:

> *"Public values: A society's "public values" are those providing normative consensus about i) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; ii) the obligations of citizens to society, the state and one another; and iii) the principles on which governments and policies should be based."* (Bozeman, 2007)

Public value may be viewed as a criterion by which to judge institutional arrangements for goods and services, but should not be confused with them. Thus, public value neither supports government action nor abjures markets. Criteria based on market failure or economic valuation often miss this fundamental point, which is critical to assessing research and innovation.

An important challenge for any analytical approach to assessing the public value of research or other social goods is the identification of *particular* public values. To say that public values are held in common does not mean that they are universally embraced or that people agree on their exact nature or content. Where does one look for public values? A nation's fundamental laws and, if there is one, its constitution provide good starting points for identifying public values, although public law is best viewed as reflecting, rather than establishing, public values. In some legal regimes, statutory and case law may reflect public values as some want them to be, rather than as how the population as a whole sees them.[7] Public values can be found in a country's fundamental myths. Myths such as "the land of opportunity" often contain several broad public values. Public values can also be found in the authoritative statements of duly authorised and legitimate policy makers. Civil societies are permeated by the public values that provide much of their structure. The problem is not to find public values but to understand them in some analytically useful form.

It is less vital to agree on a precise approach to identifying public values than to agree that it is useful to evaluate research impacts from the standpoint of public values (as opposed to conventional analysis of economic impacts, intermediate goals or inputs). So long as the analyst is clear about the definition of public values used, evaluation of research can proceed and, if transparent, promote debate about the social impact of research. If the aim is to gauge the extent to which some public value has or has not been obtained, rather than whether or not the value is a public value to begin with, there may be little objection to positing it as a public value. Thus, "improvements in public health and longevity" would seem to entail only minimal controversy, as would "decreased infant deaths" or "cleaner air". Posited as a public value, even the last example is unlikely to stir much controversy.

A public value mapping criterion model

Disagreement about the measurement of public value is less troubling when there are public value criteria. The criteria used in the PVM model are a set of diagnostics which can be applied to science policy and the evaluation of research (Bozeman and Sarewitz, 2005). To some extent, like the market failure model and related concepts, the PVM approach seeks to identify the failure to achieve public value. This occurs when neither the market nor the public sector provides the goods and services required to achieve public value. PVM criteria change the discussion of public policy and management by assuming that both government and market organisations need to be more than a means of ensuring market successes and technical efficiency in pricing structures. A fundamental assumption of the PVM model is that market failure says little about whether government should intervene. With PVM, the key policy question becomes: Even if the market is efficient, does the investment fail to provide adequate public value? The PVM model is not a decision-making tool (like cost-benefit analysis), but a framework to promote discussion of public value (and its relation to economic value and public values). Its primary use is for policy deliberation and promoting public dialogue. Table 1.1 presents the PVM model.

---

7. There are several instances of decisions of high courts in common law countries that do not necessarily reflect agreed public values. Courts' interpretations of the Constitution or the law may lead to reading "public values" into certain texts. For instance, in the United States, *Roe v. Wade* (1973) was decided by the Supreme Court and established a woman's right to choose to have an abortion. The Court asserted this right under the substantive due process portion of the 14th Amendment to the US Constitution. At the time of the decision, many laws stated that there was no right to choice. The court asserted that a woman's right to choice is a public value (or a fundamental right). The Court's holding still has the weight of law today, even though many Americans have different views.

Table 1.1. Public failure and public policy: a general diagnostic model

Public failure criterion	Failure definition	Science policy example
Mechanisms for articulation and aggregation of values	Political processes and social cohesion insufficient to ensure effective communication and processing of public values	Peer review, the favoured means of decision making for individual projects, is appropriated for decisions on huge scientific programmes, resulting in displacement of social goals for more easily defined technical goals
Imperfect monopolies	Private provision of goods and services permitted even though government monopoly deemed in the public interest	When public authorities abrogate responsibility for overseeing public safety in clinical trials for medical research, there is potential for violation of public trust and public value
Scarcity of providers	Despite the recognition of public value and agreement on the public provision of goods and services, they are not provided because providers are not available	The premature privatisation of the Landsat programme shows that a scarcity of providers can create a public failure that may be remedied by government action
Short time horizon	Adopting a short time horizon when a longer-term view shows that certain actions lessen public value	Policy for energy R&D, by considering the short term, fails to fully capture the costs of global climate change on future generations
Substitutability* vs. *conservation of resources	Policies focus either on substitutability or indemnification even in cases when there is no satisfactory substitute	No-net-loss policies, ranging from wetlands protection to prohibiting the sale of human organs on the open market. fail to take into account the non-substitutability of many natural organisms
Benefit hoarding	Public commodities and services are captured by individuals or groups, limiting broad distribution to the population	A prime technical success of genetic engineering, the terminator gene, is an excellent means of enhancing the efficiency of agricultural markets, potentially to the detriment of millions of subsistence farmers throughout the world

Source: Adapted from Bozeman (2007) and Bozeman and Sarewitz (2005).

Developing and applying the PVM model

While considerable conceptual work has already been undertaken to provide building blocks for PVM, these have not yet been integrated to create a viable model that can generate practical analytical tools. That is the purpose of the work currently under way. All PVM approaches begin as case study analyses. Current PVM cases use the following analytical approaches:

- A search for pertinent "public values": Several approaches to identifying public values have been discussed, including *i)* surrogate public values (government mission statements, strategic plans, and broad policies, statutes); *ii)* distillation of public values from relevant academic literature; *iii)* public values as expressed in public opinion polls and public statements. The public value failure criteria will be used to guide the analysis of possible public value deficits or public value failures.

- The application of the public values grid: After developing information about putative public values and gathering data about the social impacts of SIPs and consequent STEM (Science, Technology, Engineering, Mathematics) research and innovation, it will be possible to map historical and current cases on an improved version of the prototype public values grid. A key aspect of the proposed study will be to improve and further specify both the public value failure criteria model and the public values grid, including further extending their direct relevance to SIPs.

- Developing value analysis chains: Among the many reasons why public value analysis of SIPs has made little headway is that values analysis is remarkably underdeveloped. One of the difficulties of values analysis (Gaus, 1990) is that analysts sometimes fail to consider relations among values, including features such as value hierarchies, conditional relations among values, logical structures of multiple and related values, and ends-means relations (Braybrooke and Lindblom, 1963). One of the key objectives of the research is to develop the capacity to clarify relations among values.

The analytical lenses for the case studies can be thought of as master hypotheses about possible determinants of the public value outcomes. The case studies use four important contextual factors that affect the social impacts of research and S&T policy.

- *Characteristics of the knowledge produced by the research*. In some instances knowledge creation processes, innovation and, ultimately, social impacts are largely governed by the inherent characteristics of the science or technology (*e.g.* "technology push" in instances when companies push innovation through R&D processes without a defined need).

- *Institutional arrangements and management affecting knowledge production and use.* Institutional arrangements pertain to the configuration of producers and users of scientific and technical knowledge, the ways in which they interact, their internal and network management.

- *Policy and political domains of knowledge production and use.* These are the political, legal, public policy and normative factors that determine research choices, utilisation and impact (*e.g.* characteristics of intellectual property policy or structures of budgets for research).

- *Market settings for knowledge production and use.* Public value may be achieved (or thwarted) by markets, quasi-markets or government entities. In some instances, much can be understood about public value by considering such market features as the relative scarcity of resources, market actors controlling resources, market segmentation, extent and nature of competition.

The case studies currently under way aim to apply and to further develop PVM approaches. They focus not only on research developments and the evaluation of research but also on the use and social impact of knowledge produced within the framework of the case studies. Case studies at various stages of development include nanotechnology-based water filtration, alternative fuels, climate change, inequities in cancer research and treatment, and public value impacts of technology transfer.

These case studies will not only perform the traditional role of "thick description" case studies but will also: *i)* provide a context for the application of a variety of analytical approaches including logic models and value chain analysis; *ii)* help determine the extent to which it is possible to distil public values satisfactorily; *iii)* extend the theories upon which PVM is based; and *iv)* point the way for further development of analytical tools.

Conclusion

The assessment of the impacts of research on ultimate social and public value is at a quite primitive level. Few tools have been developed and many of the “borrowed” tools have important limitations. By simultaneously working to develop theory, concepts, cases and early analytical approaches it is perhaps possible to develop approaches that supplement the rational choice, cost-benefit, bibliometric and other approaches to the evaluation of research that have been developed chiefly for purposes other than assessing public value. This chapter has reviewed some of the motivations for a public value-oriented approach to assessment, some of the problems for developing such an approach and a few of the early steps taken to fulfil the perceived need for approaches that are better able to determine the social impacts of research.

REFERENCES

Andrews, F. (1979), *Scientific Productivity, the Effectiveness of Research Groups in Six Countries*, University of Michigan Press, Ann Arbor, Michigan.

Auditor General (1993), *Program Evaluation in the Federal Government. The Case for Program Evaluation*. Treasury Board of Canada.

Audretsch, D. B., B. Bozeman, K. L. Combs, M. Feldman, A.N. Link, D.S. Siegel, P. Stephan, G. Tassey and C. Wessner (2002), "The economics of science and technology", *Journal of Technology Transfer* 27 (2), pp. 155-203.

Barbarie, A. (1993), "Evaluating federal R&D in Canada", pp. 155-162 in B. Bozeman and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, Norwell, Massachusetts.

Bozeman, B. (2003), "Public value mapping of science outcomes: theory and method", in D. Sarewitz, *et al*. (eds.), *Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development*. 2(1).

Bozeman, B. (2007), *Public Values and Public Interest: Counter-balancing Economic Individualism*, Georgetown University Press, Washington, DC.

Bozeman, B. and J. Melkers (eds.) (1993), *Evaluating R&D Impacts: Methods and Practice*. Kluwer Academic Publishers, Boston, Massachusetts.

Bozeman, B. and J.R. Rogers (2002), "A churn model of scientific knowledge value: Internet researchers as a knowledge value collective", *Research Policy*, 31, 5, 769-794.

Bozeman, B. and D. Sarewitz (2005), "Public values and public failure in U.S. science policy", *Science and Public Policy*, 32 (2), 119-136.

Callon, M., P. Laredo and P. Mustar (1997), *The Strategic Management of Research and Technology: Evaluation of Programmes*, Economica International, Paris.

Cozzens, S., S. Popper, J. Bonomo, K. Koizumi and A. Flanagan (1994), *Methods for Evaluating Fundamental Science*, Report prepared for the Office of Science and Technology Policy, RAND Corporation, Critical Technologies Institute Washington, DC.

Cummings, R. and L. Taylor (1999), "Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method", *American Economic Review*, 89 (3), pp. 649-665.

Fischer, E.P. (1997), *Beauty and the Beast: The Aesthetic Moment in Science*, trans. Elizabeth Oehlkers, Plenum Trade, New York.

Johnson, H. (1965), "Federal support of basic research: some economic issues", *Minerva*, 3 (4), pp. 500-514.

Jones, C.I. and J.C. Williams (1998), "Measuring the social return to R&D", *Quarterly Journal of Economics*, 113 (4), pp. 1119-1135.

Kostoff, R. (2001), "The metrics of science and technology", *Scientometrics*, 50, 2, pp. 353-361.

Link, Albert N. (1996a), "Economic performance measures for evaluating government sponsored research", *Scientometrics*, 36(3), pp. 325-342.

Link, A. (1996b), *Evaluating Public Sector Research & Development*, Greenwood, New York.

Luukkonen T. (2002), "Research evaluation in Europe: state of the art", *Research Evaluation*, 11 (2), pp. 81-84.

OECD (1995), "Cost/benefit analysis of large S&T projects: some methodological issues", *www.oecd.org/dsti/sti/s_t/ms/prod/e_95-57.pdf.*

Pawson, R. and N. Tilley (1997), *Realistic Evaluation*, Sage Publications Ltd, London.

Rubenstein, A. (1976), "Effectiveness of federal civilian-oriented R&D programs", *Policy Studies Journal*, 5 (2), pp. 217-227.

Ruegg, R. (1996), "Guidelines for economic evaluation of the Advanced Technology Program", NIST Internal Report 5896.

Saavedra, P. and B. Bozeman (2004), "The 'gradient effect' in federal laboratory-industry technology transfer partnerships", *Policy Studies Journal*, 32 (2), pp. 235-252.

Salasin, J., L. Hattery, and T. Ramsay (1980), "The evaluation of federal research programs", MITRE Technical Report MTR-80W123, June.

Sen, A. (1982), *Choice, Welfare and Measurement*, MIT Press Cambridge, Massachusetts.

Solow, R.M. (1957), "Technical change and the aggregate production function", *Review of Economics and Statistics* 39(3), pp. 312-320.

CHAPTER 2

ASSESSING THE ROLE OF PEER REVIEW

This chapter explores the role of expert review and the need to include the social impacts discussed in the previous chapter in the evaluation process. It reviews the role of expert review as a key tool for evaluation of public R&D. It explores current challenges faced by the peer review process and expert review as a tool for *ex ante* and *ex post* evaluation of research policies, programmes and public research organisations (PROs). It aims to provide a comprehensive assessment of "expert review" at the programme and policy level and to summarise methodological issues and good practices that have emerged from OECD member countries' experiences. Although it offers principles and suggestions, it does not aim to be a guide or handbook for programme managers in evaluation agencies.

Introduction

Aims and scope

This chapter addresses definitions, procedures, the underlying issues and suggestions for research policy and programme expert review. It first summarises the definitions, the purposes and applications, and the merits and limits of expert review. It then describes in detail the expert review process protocols found in member countries' good practice guidelines before addressing methodological issues regarding expert review and suggests solutions based on countries' research experiences. Finally, many key principles and suggestions are summarised.

The problems in "expert review"

The expert review is one of the main methods used to evaluate science and technology (S&T). It plays a significant role in many of the key stages of research. It is the main mechanism for deciding who and what gets funded, who publishes in the scientific literature, and who is selected and promoted by research institutions (Scott, 2006). It is also the core tool used in various research and development (R&D) programmes and innovation policies.

Expert review has many merits. It is a relatively quick, low-cost, well-known, widely accepted and versatile tool which can be used to answer a variety of questions throughout the project performance cycle as well as in other applications. It also provides an opportunity for mutual learning. Expert review may well be the best of all known methods of assessing R&D programmes and policies so long as it is properly managed.

There are, however, concerns that the expert review system is under pressure and that users are losing confidence in it. The lack of confidence stems from the fact that the system depends on the professional but subjective decisions of individuals. Moreover, the process is increasingly time-consuming and resource-intensive. It is not an exaggeration to say that expert review currently faces some of its strongest challenges in several decades. They are both external and internal. Externally, there is some evidence of dissatisfaction among political decision makers about the capacity of expert reviews to reflect socioeconomic and political priorities. Internally, there is a hollowing out as increasing pressures on researchers' time make it more difficult to find experts willing to undertake reviews. From the perspective of evaluation methods, it is an appropriate time to assess the status of expert review and to identify possible challenges and solutions.[8] Among the key issues that arise are:

- How to reflect socioeconomic and political priorities effectively and link these priorities to decision making in expert review processes.
- How to combine expert review with other quantitative and qualitative methods to improve evidence-based policy.
- How to enhance cost efficiencies at the various stages of the peer review process.
- How to develop an effective international frame of reference for expert review.
- How to manage conflicts of interest in the expert review process.
- What opportunities the Internet offers for improving expert review.
- What type of expert review is appropriate for the evaluation of policy, programmes or public research organisations (PROs).
- What the key principles are for ensuring a high-quality programme/policy expert review.

Definitions and applications

To fully understand the scope of expert review and its relative strengths and weaknesses, it is necessary to understand what separates expert review from traditional peer review. This calls for a comprehensive definition, as well full understanding of its purpose and application. This section considers definitions of expert review, describes its purpose and application, and discusses the merits and challenges of the process.

Definitions

There are several definitions of peer review. Hartmann and Neidhardt (1990) define peer review as various processes to evaluate the quality of research by peer scientists. Chapman and Farina (1983) define peer review as "a process of assessment of research proposals by peer scientists". Kruytbosch (1989) also provides a simple definition of peer review in science as "advice about proposed actions solicited by decision makers from recognised experts in relevant technical areas". Chubin and Hackett (1990) say that peer review is an organised method for evaluating scientific research in order to enhance the exactitude of the research process, evaluate the authenticity of results, and allocate scarce resources.[9] An OECD document provided a comprehensive definition of peer review (Gibbons and Georghiou, 1987):

---

8. For these reasons, the 2005 OECD-BMBF Conference on Evaluation and subsequent meetings have highlighted a number of issues in the area of peer review of research. See *www.internationales-buero.de/de/2193.php* and *www.pragueforscience.cz/Scientific-Programme.php*.

9. Some use the term "peer advice", "peer evaluation", "peer judgement", "quality control", "peer censorship", "merit review" and "refereeing" as an equivalent.

Peer review is the name given to the judgement of scientific merit by other scientists working in, or close to, the field in question. Peer review is premised upon the assumption that a judgement about certain aspects of science, for example its quality, is an expert decision capable of being made only by those who are sufficiently knowledgeable about the cognitive development of the field, its research agenda, and the practitioners within it.

Peer review in this form is intrinsic to the practice of science and is used for publication, career and resource allocation decisions. It is widely used by industry, government and academia. It is increasingly used as an instrument for *ex post* evaluation. The peer review model has also been extended to encompass additional criteria, notably socioeconomic criteria, the potential to contribute to innovation, and other considerations of merit beyond scientific quality. Given this trend, EERE's *Peer Review Guide* (2004) defines in-progress peer review as:

A rigorous, formal, and documented evaluation process using objective criteria and qualified and independent reviewers to make a judgement of the technical, scientific, and business merit, the actual or anticipated results, and the productivity and management effectiveness of programmes and/or projects.

Clearly, no single definition of peer review is used in the evaluation literature. However, all definitions of peer review adhere to the fundamental concept of a review of scientific or technical merit and socioeconomic impacts by individuals with professional competence and no unresolved conflicts of interest (GAO, 1999; Guston, 2001).

From peer to expert review

Expert review is a broader concept than peer review.[10] The classical definition of a peer is "a person who has equal standing with another". Therefore, one could define peer review as "a review of a person or persons by others of equal standing". The crucial issue is how "equal standing" is defined. For example, although scientists who participate in an evaluation may be identified as the "peers" of the applicants when evaluating research proposals, experts in other fields, in addition to peer scientists, should be included in a programme evaluation. The term "expert review" is therefore more appropriate than peer review for the evaluation of a programme. The term "expert review" could be defined as follows (Ruegg and Jordan, 2007):

Qualitative review, opinion and judgement from individuals with professional competence on the subject being evaluated, based on objective criteria.

The best-known form of expert review is actually peer review, on the premise that a scientist's or engineer's peers have the essential knowledge and perspective to judge research quality. Peer review is commonly used to make many kinds of judgements, such as those about the careers of individual researchers, about the value of their publications, about the standing of research institutions, and about the allocation of funds to individuals and to fields of research (COSEPUP, 1982). Therefore, some people use the term "peer review" instead of expert review.

10. According to COSEPUP (1999), "expert review" can be classified into three types: *i) peer review*, which is commonly used to make judgements about the careers of individual staff members, the value of publications, the standing of institutions, and the allocation of funds to individuals, organisations and fields of inquiry; *ii) relevance review*, which is used to judge whether an agency's programmes are relevant to its mission; and *iii) benchmarking*, which is used to evaluate the standing of an organisation, programme or facility relative to another.

In sum, expert/peer review has distinguishing characteristics: a *qualitative* method, judgement by qualified *individuals*, and *objective criteria*. Whichever definition one uses, the following issues may be important for high-quality expert review or peer review:

- Who should be the evaluator?
- How can the credibility of subjective opinions and judgements of individuals be enhanced?
- How can materials and criteria be developed and provided to evaluators for objective evaluation?

Purposes and applications

Expert review is but one of many methods of evaluation policy makers use to reach their decisions. To evaluate policy or programmes involves assessing one or more of five domains (Rossi, Lipsey and Freeman, 2004): *i)* the need for the policy/programme, *ii)* the policy/programme's design, *iii)* its implementation and service delivery, *iv)* its impact or outcomes, and *v)* its efficiency. The general goals of the evaluation relate mainly to programme improvement, enhancement of accountability, or knowledge generation (Chelimsky, 1997).

According to the literature (Kostoff, 2004; Alassaf, 1996; Armstrong, 1997; Cram, 1992; Levine, 1988; Palli, 1993; Rainville, 1991; Ramsay, 1989; Stull, 1989; Wakefield, 1995; Wicks, 1992) expert reviews of projects and programmes serve a broad range of purposes:

- They serve as a quality filter to conserve scarce resources.
- Papers published in peer-reviewed journals are assumed to be above a threshold of minimal quality, so that readers can assume that the documents contained in journals are of high quality.
- Projects and programmes selected by expert review for initiation or continuation are assumed to be above a minimal threshold of quality.
- Precious labour and hardware can be focused on selected high-quality tasks.
- Expert review has the potential to add value to, and improve the quality of, the manuscript or programme under review.
- Expert review can help provide legitimacy and competency and increase a programme's visibility and support.
- Expert reviews range from being an efficient resource allocation mechanism to being a credible predictor of research impact.
- A properly conducted expert review of a research programme can provide its sponsors with a credible indication of the programme's quality, relevance, management and appropriateness of direction.

Policy makers and programme managers want to learn from an evaluation whether the research is done correctly (*e.g.* has high quality and efficiency); whether a programme's R&D efforts are focused on the right areas; whether programme-created knowledge can find varied applications that generate additional benefits to the nation; how collaboration and other activities stimulated by the programme have affected the nation's R&D capabilities; and ways in which past efforts or new planned initiatives are worthwhile. A good expert review should be able to provide programme managers and policy makers with answers to these questions. Ruegg and Jordan (2007) provide a good summary of uses of programme expert review:

- to conduct in-progress reviews of scientific quality and productivity;
- to help answer questions about the relevance, timeliness, riskiness and management of existing programmes' research activities and sufficiency of resource for new programmes;
- to score and rate projects to aid decisions to continue, discontinue or modify existing or planned project, programmes or initiatives;
- to help assess the appropriateness of programme mechanisms, processes and activities and how they might be strengthened;
- to integrate multiple evaluation results and make judgements about the overall success of a programme or programme initiative;
- to provide information to help programme managers make decisions to design or revise their programme, redirect existing R&D funds, or allocate new funds.

Merits and limitations

Like other methods of evaluation, expert review has its strengths and limitations. This section summarises the merits and limitations of expert review. Its merits are several:

- *It is relatively fast and convenient*. Assuming that the most appropriate experts are selected, an expert review can be very efficient in terms of the time required.
- *It may be carried out in various situations*. There are few types of projects or programmes that would not benefit from expert review.
- *It can easily persuade stakeholders to accept.* It is relatively easy to persuade both the party to be evaluated and the stakeholder to undertake an expert review.
- *It is relatively cheap*. The need for further analysis is lessened by reliance on the existing knowledge of the experts.[11]
- *It provides those involved with opportunities for mutual learning*. The expert review process involves much discussion and exchange of ideas. This can lead to intended and/or unintended benefits.

Despite the foregoing merits, expert review has certain limitations:

- *It is difficult to ensure the accuracy and quality of the resulting evaluations.* Expert review has limited usefulness as a method to guarantee reliability and consistency (or repeatability).

---

11. There are, however, considerable hidden indirect costs. For programme evaluation, the actual cost may increase significantly owing to the additional resources needed to analyse the programme.

- *The quality of a review can be affected by the reviewers' biases and conflicts of interests*. Although various measures can help to reduce biases and conflicts of interest, they can never be completely eliminated.
- *Expert reviews tend to perpetuate orthodox and conservative paradigms* and to reject new paradigms that threaten the *status quo*.

The second and third drawbacks relate to the reliability of or confidence in expert review. These risks generally appear in the review of grant applications or scientific papers (*i.e.* project level expert review) and have most often been examined in the context of "peer review". While reviewers should be as objective as possible, in practice, peer judgement is affected by factors (*e.g.* bias, favouritism, conservatism, discrimination), which have nothing to do with the subject of the evaluation. The possible lack of objectivity can lead to a lack of confidence.

The first bias is known as the "Matthew effect", *i.e.* the allocation of research funds may be skewed towards more famous and influential researchers. Moreover, the effect also indicates that researchers who received funding in the past are more likely to receive further funding (Merton, 1973). Gustafson (1975) showed that 46% of all research funds in the United States were awarded to the top ten research organisations funded by the National Institutes of Health, and the top one-third of total funds went to the top 20 organisations funded by the National Science Foundation. The "Matthew effect" can be a severe problem, especially when research funds are scarce. Those who point out the problem are usually unsuccessful applicants (Pouris, 1988).

Peer review is not immune to the risk of cronyism. Personal connections play an important role especially in the evaluation of a major project that may have a strong impact on a researcher's reputation. The selection of panel members and their evaluation processes may also be influenced by favouritism and discrimination. For example, if a person holds a key post on the evaluation committee for a long time and then appoints his/her successor, the evaluation committee may represent the interests of a certain group rather than the entire science community. This can lead to discrimination against certain groups, such women, young researchers or researchers who work for less renowned institutes and universities (Gustafson, 1975). Objective and fair evaluations are important to avoid social replication or the so-called "old boys' network".

Many critics also call attention to the conservatism of peer review and a tendency to support an orthodox paradigm rather than look for, promote and fund more innovative research. In this case, peer review is likely to reinforce the views of mainstream scholars (this is known as "Pied Piper effect") (Kostoff, 1996). This makes the choice of members of the review panel important; a weak point of peer review is the fact that panels are generally composed of specialists in specific fields rather than experts with a broader view. A review panel composed only of mainstream researchers will be more concerned with questions such as "is this research likely to be successful?" rather than more fundamental questions such as "is this research really needed?" Certain authors also note that these scholars may tend not to acknowledge the scientific achievements of other fields (Bozeman and Melkers, 1993). Established fields may also have more ready access to the mass media and government (Pouris, 1988).

Ethical issues also arise in the peer review process: fraud, plagiarism, fabrication, image manipulation, leakage of commercial confidentiality, etc. (Campbell, 2006). It is relatively easy for reviewers to appropriate or use a grant applicant's ideas by delaying the evaluation process if they conduct research on similar topics. Alternatively, a leading scholar may not want to see a rival who might challenge his/her authority succeed and so may criticise the researcher's project (Pouris, 1988). Scientific misconduct like this has an enormous impact but is often hard to document. In fact, much misconduct in science and technology originates in the peer review process, even though the academic world expends a great deal of effort to prevent it, since one purpose of peer review is to protect the science and technology community's ethical values (Goodstein, 1995).

Key processes of expert review

This chapter highlights challenges to expert review and presents some emerging solutions. It is useful first to describe the key processes of expert review of programmes/policies because they differ from the evaluation of research projects in terms of their use, stakeholders and complexity. Good examples of the expert review process at programme level are provided in various national or institutional guidelines (EERE, 2004; Kostoff, 2003; Kostoff, 2004; Rigby, 2002; British Academy, 2007; EPA, 2000).[12] Expert review is generally understood to have three main phases: pre-review, implementing review and post-review. This section describes these phases, highlighting some important steps in the process.

Pre-review

The pre-review phase is a preparation and planning stage and includes the following three activities: establishing the foundations of the review, selecting reviewers and preparing tools and materials.

Establishing the foundations of the review

Initiation of the review: Assigning responsibilities. A successful expert review of an R&D programme requires the full participation of the unit being reviewed. The necessary motivation and participation derive from the actions of an organisation's senior management when the process is initiated. It is essential for a senior manager (in the evaluation agency) to send out an initial letter to all participants setting out: the purpose of the review and its importance; the goals, objectives, and scope of the review; the identity and responsibilities of the review manager(s), the general responsibilities of the reviewers; and the responsibilities of the reviewers through all phases of the review process (Kostoff, 2004).

Identifying the purpose and scope of the review. Once the senior manager has assigned responsibilities, he/she must establish the principles that govern the review. The first step is to determine its purpose and scope within the context of other review and management activities. Clearly identifying the objectives of the review and the boundaries of the programme to be reviewed provides a framework. If the purpose is unclear and the scope is too broad, the result is confusion and lack of precision. However, if the scope is too narrow, it is difficult to gain an overall perspective and draw conclusions on how to redistribute resources and make changes. Evaluations of R&D programmes may include in-depth technological reviews of the S&T projects within the programme. Generally speaking, at the project level, the review focuses on whether the "projects are being done right". At the programme level, the focus is on whether the "right things are being done".

12. Opinions on the phases of expert review differ. The EERE guide and Kostoff (2004) provide information and examples useful for planning, conducting and using expert reviews based on best practices in the United States. Kostoff suggests the following five phases: *i)* initiation of the review; *ii)* establishing the foundations for the review; *iii)* preparing for the review; *iv)* conducting the review; *v)* enacting post-review actions. EERE's guide describes four phases: *i)* preparations; *ii)* pre-review; *iii)* conduct of the review; *iv)* post-review activities. The United States' EPA's *Peer Review Handbook* (2000) describes three stages: *i)* planning a peer review, *ii)* conducting a peer review, *iii)* completing a peer review. Rigby (2002) suggests twelve key steps: *i*) setting the terms of reference; *ii)* overall time available; *iii)* appointment of panel chair; *iv)* appointment of panel members; *v)* appointment of the panel secretary or scribe; *vi)* operating procedure; *vii)* schedule of work of the panel; *viii)* links from panel to programme/client and other sub-contractors; *ix)* identifying the requirement for external support; *x)* interim reporting; *xi)* final reporting; *xii)* dissemination.

Usually a review unit is established to define the scope of the review and assess how well the programme fits the policy objectives and consider its relation with other programmes, the relevance of the project portfolio and the programme's relation to the external environment. The review unit therefore needs to be selected on the basis of the objectives and uses of evaluation.

Identifying the evaluation criteria and the review questions. Expert review requires establishing evaluation criteria in advance. The project/programme's mission and the objectives and nature of the review should help to identify and select evaluation criteria. The criteria and standards for judging aspects of the programme should reflect the programme's definition of success and the characteristics of the programme or projects. They should focus on the aspects that most need to be discussed by an objective expert group. The criteria and the relevant questions need to be stated as clearly and succinctly as possible to guide reviewers towards the appropriate goals (EERE, 2004).

The fundamental criteria for evaluating an R&D programme are research quality, research relevance and overall programme quality. They are sometimes subdivided into research merit, research approach/plan/focus/co-ordination, match between resources and objectives, quality of research performers, probability of achieving research objectives, programme productivity, potential impact on mission needs (research/technology/operations), probability of achieving that impact, potential for transition or utility, and overall programme evaluation (Kostoff, 1997; Kostoff, 2004).

For example, in the United States, a few criteria are recommended and used by the Department of Energy (DOE), the Office of Management and Budget (OMB), the National Academies of Science (NAS) and other agencies. The most prominent are usually accomplishments, relevance and method.[13] Because specific questions make it easier for the reviewer to do the job requested, evaluation criteria are often presented to evaluators as questions tailored to the particularities of the evaluated project/programme.

Identifying the information needed and data collection/analysis processes. After determining the purpose, scope, criteria and questions, attention should turn to the review process itself. During this phase several questions arise: What type of review should take place? How should the necessary data be collected, analysed and transmitted to the evaluators? How should the evaluation results from the evaluators be presented? This will depend on the particularities of the programme/policy and its objectives and uses. For example, if the main objective of the evaluation is to determine how the programme is performing, the data reflect the programme's performance and the analysis focuses on the programme's output, input and impact. On the other hand, if the objective is to modify a programme or to decide on its continuation, the analysis would address the relevance of the programme and examine its past portfolio.

The data collected must be sufficient for reviewers to judge past and ongoing activities against the criteria and the specific questions. The data usually include material that is provided prior to and during the review. A balance must be struck between too much and too little data. To the extent possible, materials that are already developed or planned for other purposes should be used to minimise the burden on researchers. Depending on the type of programme, data can include the following (EERE, 2004): information on the programme/project mission, goals, targets and other milestones. Additional data are sometimes necessary, such as: data on how funding is allocated across key activity areas; summary project reports, plans and budgets; presentations by principal investigator or project manager; lists of publications

---

13 Although programmes may choose to define additional criteria, all EERE programmes are expected to use, at a minimum, the following three criteria (referred to as "core criteria"): *i)* quality, productivity and accomplishment; *ii)* relevance; *iii)* management. In addition, reviewers may be asked to provide an overall assessment. The OMB R&D Scorecard provides another example of criteria (US DOE FY 2002 R&D Scorecard): *i)* accomplishments, *ii)* relevance, relevance of future research; *iii)* approach to performing technology transfer/collaboration.

or patent applications and the results of citation analysis; customer surveys; available impact studies; reports prepared by other external groups; and/or other data and information reviewers may request.

Identifying the type of review group and the audience. In programme/policy review, the competence of the review group may be more important than individual reviewers' technical competence., The selection of type of review group is therefore an important issue and should be addressed at the start of the review process. Various types of groups are possible, depending on the aim of the review: for example, there may be an independent panel, *i.e.* a group of experts who are independent of the agency and typically funded under a contract or an external group consisting of experts individually contracted to the agency. Generally, an independent panel is used when the purpose of the evaluation is to assess the performance or the accountability of a programme.

A programme review can provide an excellent means of disseminating programme information and results to a wide audience. It needs to be determined early in planning whether or not the public will be invited to be present or participate in the review sessions. Care should be taken to ensure that a review audience includes: actual and potential customers, stakeholders and other oversight groups, co-sponsors, users, and other agency representatives (Kostoff, 2003, 2004).

Establishing a timeline and determining the logistics of the review. The primary goal of a programme review is to provide information that assists programme staff and managers in their efforts to improve programme performance. Because evaluation has a practical purpose, timing is important. Consequently, after setting a date for presenting evaluation results, major deadlines should be established. Although in theory, resources (time, money, people, etc.) for the review depend on programme size, objectives of evaluation, etc., in practice these resources are scarce. These limitations need to be considered when determining the logistics in order to concentrate on key issues and fundamental processes.

Selecting and inviting the reviewers

Identifying criteria for selecting reviewers. When seeking nominations for reviewers, it is important to state the selection criteria clearly. The review manager, working with staff, the external steering group and others, establishes the qualifying criteria which can include: *i)* in-depth knowledge of the subject area for which the reviewer is selected; *ii)* lack of real or perceived conflicts of interest.

Developing a list of possible reviewers and nominate. Reviewers are appointed after the overall technical description of the programme has been established and descriptions of the technical sub-areas have been presented. Potential sources of candidate reviewers include: recommendations of programme managers, membership lists of prestigious organisations, agency review boards, agency consultant pools, and contributors to technical databases (such as authors of journal articles or technical reports). The review manager, working with the external steering group and/or others, develops an initial list of candidate chairpersons and reviewers by: *i)* arranging for several independent, external objective groups familiar with the programme to nominate candidates; *ii)* identifying candidate chairpersons and reviewers among experts identified in a bibliometric search of the published literature on the topic or on the basis of their roles in research or management institutions or professional societies; *iii)* employing a co-nomination approach, *i.e.* selecting reviewers among persons nominated by more than one external expert in the relevant field.

Gathering background information and developing an initial selection list. The review manager gathers information on the candidate chairpersons and reviewers, for example by:

- Reviewing the performance of reviewers in past reviews.

- Contacting candidates to determine their interest and availability; sending them project summary descriptions to identify their interests and possible conflicts; and requesting and reviewing self-assessment forms.

- Obtaining staff and/or public input to identify candidates who may have known biases or other issues. Considerable care is needed to prevent the gathering of materials or other inputs that could unfairly or inappropriately characterise an individual, while ensuring that privacy or other concerns are not raised.

Selecting the chairperson and reviewers from the list of nominees. The review manager selects the chairperson and reviewers from the list of nominees by working with the external steering group, the chairperson (after selection) and/or others, for example by:

- arranging for independent, external, unbiased and objective university, professional society or other groups familiar with the programme, as identified above, to select the chairperson and/or the reviewers from the nominees;
- having the review chairperson select the rest of the reviewers from the nominees;
- identifying the chairperson and the reviewers based on a co-nomination process among the candidates;
- using an independent, unbiased, objective contractor to select the nominees either directly or in collaboration with the steering group and independent, external, unbiased universities, professional societies, or others.

To ensure transparency the selection process should be carefully and fully documented and included in the final report.

Preparing tools and materials

Developing guidelines and tools for the review. Both the review panel and those being reviewed should clearly understand the review's objectives and guidelines as well as the specific evaluation criteria. The review leader and chairperson should determine how the project/programme is to be rated and distribute this information to both parties, generally in the form of a written description of the evaluation method. These guidelines should describe the purpose and scope of the review, the evaluation criteria and questions, the data to be presented, and how the data are to be collected, analysed and reported.

Rating or scoring systems are often used to improve the effectiveness of the evaluation. To ensure comparability of ratings across reviewers and review groups, reviewers need to use the rating scale in the same way. The scale must be well defined so that a given rating (adjective or number) will represent the same appraisal made by different reviewers.

Developing presentations. During the review process, it is easy for project leaders to prepare presentations. It is more difficult to present a programme because of the need to account for various socioeconomic factors as well as the numerous components of the programme itself. Evaluation managers should provide appropriate guidelines on presentation to relevant managers.

Providing evaluation material. Before embarking upon an evaluation, the evaluators, as well as those being evaluated (*e.g.* presenter, programme manager) need clear instructions on the materials needed for the evaluation. This allows those being evaluated to prepare for the evaluation effectively.

A variety of background materials should be supplied to reviewers (and to the invited audience) before the review. When those being evaluated submit background material and results based on the guidelines provided by the evaluation manager, these materials should be distributed in a timely manner to reviewers along with guidelines on evaluation criteria, processes and indicators. To ensure the quality of the evaluation it is important to provide sufficient time for reviewing these materials. Evaluation managers can provide documents describing programme accomplishments at this time, although it is better to distribute them in advance so that reviewers can, if necessary, request additional materials.

Creating an expert review record. The expert review record is established at the beginning of the review process and maintained throughout. It should contain all the key review documents. It is important for transparency and will help the evaluation manager to improve the expert review process.

Conducting the review

Providing on-site instructions to reviewers. It is recommended that the review leader or chairperson orally reinforce the previously distributed guidelines at the opening of the review in order to clarify its purpose. This gives reviewers an opportunity to address outstanding concerns or questions before the review begins. The leader or chairperson should also remind reviewers to keep all evaluations strictly confidential during and after the review process.

The specifics of the on-site instruction depend on choices made by the review leader, review chairperson and/or group. However, in general, reviewers may be instructed to: *i)* read and understand the evaluation criteria and peer review procedures; *ii)* evaluate each programme element; *iii)* prepare preliminary comments on the merits of the project/programme in accordance with the peer review evaluation criteria; *iv)* be prepared to discuss each project/programme at the meeting or assign a rating or ratings that reflect the reviewer's opinion of the merit of the project/programme in accordance with the specific evaluation criteria, and; *v)* complete the post-review evaluation form.

Programme presentation and questions and answers (Q&A). Presentation is a crucial step in the review process because expert reviews generate new ideas through discussions between evaluators and those evaluated and provide opportunities for mutual learning. Concerned parties from various levels – organisation unit head, programme manager, technical unit head – can make presentations. For example, technical presentations are initiated by the head of the organisational unit in which the programme resides, and include the following information: the mission and objectives of the organisational unit, a list of all of its programmes, a description of the objectives of each programme, the funds and people associated with each programme and with the programme to be reviewed, an overview of the stages and accomplishments of programmes not being reviewed and their relation to those of the organisational unit's mission, potential national impact, etc. The programme manager's presentation can provide a more detailed overview of the programme under review, including its objectives, the requirements to be satisfied and derived target capabilities for the S&T initiative.

Discussion and judgement. Reviewer-to-reviewer interaction, for example during a special closed session, to discuss their preliminary ratings and then finalise them can improve the quality of the review findings. This discussion can help clear up misconceptions or introduce new information. Such interactions may be particularly important for a higher-level programme review as they can help to better understand the full range of issues. The review chairperson needs to ensure that no single reviewer dominates the ratings discussion and to make clear that consensus is not expected.

After reviewers have discussed their ratings, they establish their overall judgement. Sometimes reviewers' individual opinions are compiled; sometimes, consensus is reached on the basis of the individual review results. In programme evaluations the latter option is often preferred. It is important for the choice of final judgement method to be determined in the preparatory stage of the review.

Synthesising reviewers' evaluation results. After discussion and judgement, the evaluation results are confirmed and synthesised for the final report. In the case of evaluations that determine priorities among different programmes, a rating or scoring system is often used. In this case, the type of rating or scoring system should be determined in advance.

Developing the review documents and report. The review report provides managers with an independent assessment of the programme's productivity, relevance and management. Where applicable, the report should include the following: programme/project identification, description and budget; a narrative summarising the salient features of the comments of the individual reviewers and their reasons for their judgements; conclusions supported by specific observations; a summary of reviewers' rating or assessment of each individual criterion as well as the overall assessment; recommendations aimed at improving programme performance, including areas where further study is desirable; comments on the status of recommendations made at prior reviews, if applicable; and annexes with the full text of reviewer input.

The review chairperson concurs with and signs off on the report, which is often also sent to reviewers to review their own responses. With this report, the "conducting review" phase is concluded and the report is distributed to stakeholders such as the programme manager.

Post-review process

Integrating additional comments. Before the report is distributed publicly, those evaluated and the programme manager can respond to the reviewers' comments and recommendations. At this time, additional comments about the review from the reviewers, the external audience or senior management should be considered for integration into the review report.

Drafting a final report. In general, there is a long version and a short version of the final report. The long version includes all the written material generated during the course of the review. It provides an archival record of what was done. The short version summarises the details of the process and focuses on reviewers' comments and other significant inputs, conclusions and recommendations. The final report should include the viewpoints of all reviewers, with appropriate weighting for the judgements and expertise of specific contributors.

Making the report available to the public. When the final report is presented to policy-level decision makers or higher-level committees and is recognised as an official evaluation, the report should be available to related parties as well as the general public through publications and the Internet.

Assigning action items and evaluating responses to action items. If internal management accepts the conclusions and recommendations of the report, action items should be assigned to the appropriate personnel to respond to problems identified in the report. Possible types of response include corrective action or rebuttal of the conclusions and recommendations. The response should be evaluated and appropriate follow-up action taken. These action items, responses and follow-up actions should be presented at the introduction of the subsequent review.

Evaluating the expert review process itself, including the lessons learned. This step is considered as a type of meta-evaluation. Expert review is used as a valuable resource for improving future expert reviews by providing information on problems faced during the process.

Table 2.1. Phases and key actions for the expert review

Phases	Key actions
Pre-review	**Establishing the foundations of the review** • Initiating the review: Assigning responsibilities (K) • Identifying the purpose and scope of the review • Identifying information needed and data collection/analysis processes • Identifying the evaluation criteria and review questions to be used • Identifying the types of review group and the audience (K) • Establishing timeline and determining logistics for the review **Selecting and inviting the reviewers** • Identifying criteria for selecting reviewers • Developing a list of possible reviewers and nominating • Gathering background information and developing initial selection list • Selecting the chairperson and reviewers from list of nominees **Preparing tools and materials** • Developing guidelines and tools for the review • Developing the presentations • Providing evaluation materials • Creating the expert review record
Conducting review	• Providing final instructions to the reviewers • Presenting the programme and Q&A • Discussing and judgement • Synthesising evaluation results from reviewers • Developing review documents and report
Post-review	• Integrating additional comments • Writing a final report • Making the report available to the public • Assigning action items and evaluating response to action items • Evaluating the expert review process itself, including lessons learned

Source: Adapted with changes from EERE (2004), *EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review*, August; Kostoff, Ronald N. (2003), *Science and Technology Peer Review: GPRA*, Office of Naval Research; Kostoff, Ronald N. (2004), *Research Program Peer Review: Purposes, Principles, Practices, Protocols*, Office of Naval Research; Rigby, John (2002), “Expert Panels and Peer Review”, in Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke, and Klaus Zinocker (eds.), *RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies*. IPTS Technical Report Series, EUR 20382 EN.

Issues and suggested solutions

The preceding section draws attention to the main steps in an expert review. The following discussion further defines the issues raised and discusses possible solutions.

The changing context

A number of changes affect how expert review operates. These changes offer new challenges and opportunities for expert review.

Emphasis on performance. Since the 1990s, the "new public administration" of Australia, New Zealand, the United Kingdom and the United States has emphasised both evaluation of public policies and utilisation of the results. For example, to enhance the accountability of government programmes, the US GPRA (Government Performance and Results Act) requires performance-based management and performance-based budgeting.

Progress on international benchmarking and the internationalisation of evaluation. Many OECD countries have increased the international benchmarking of their S&T policies (OECD, 2007d). This may be seen as a continuing effort to promote the quality and objectivity of evaluation.

Development of methodologies. More efforts have been made to evaluate programmes/policies using quantitative indicators. This has led to the development of new indicators. Various methods have also been developed to measure certain programmes' socioeconomic impacts. This indicates a growing interest in developing ways to effectively complement expert review with other evaluation methods.

Need for greater transparency. The limited resources available for R&D create more competition when setting priorities. This calls for greater transparency in the priority-setting process. The elimination of biases and conflicts of interest in the evaluation process also remains a challenge.

Development of information and communication technologies. The development of communication tools, such as phone-conference, videoconference and the Internet, allows for greater flexibility in expert review. The Internet and online databases also make possible real-time entry and review of evaluation data/information because there are no temporal or spatial limits on accessing and exchanging information. These developments have helped to improve the effectiveness and quality of evaluations and allowed for network-centred expert review. Electronic communication means that expert review is more easily an international process, potentially widening the range and number of reviewers.

Methodological issues and solutions based on country experience[14]

Over the years peer review has received much attention in the evaluation literature. Studies have suggested a number of challenges, solutions and issues. Most relate to project-level evaluations of grant applications, publication of papers and *ex post* project evaluation. Wood and Wessley (2007) cover issues mainly related to peer review for grants. They consider various issues: Is peer review of grant application fair? Are peer reviewers really peers? Is there institutional bias? Do reviewers help their friends? Does gender or age bias exist in peer review? How reliable is grant peer review? Does peer review of grant applications serve the best interests of science? Is peer review of grant application cost-effective? Can peer review of grant applications be improved? Should peer review of grant application be replaced?

Kostoff (2004) also describes the strengths and weaknesses of major peer review components and issues: objectives and purposes of peer review; quality of peer review; impact of peer review manager on quality; selection of peer reviewers; selection of evaluation criteria; secrecy (reviewer and performer anonymity); objectivity/bias/fairness of peer review; normalisation of peer review panel; repeatability/reliability of peer review; effectiveness/predictability of peer review; global data awareness; cost of performing a peer review; ethical issues in peer review; alternatives to peer review; and recommendations for further research in peer review.

The complexity of the issues raised in the literature makes it impossible to cover all of them. The focus here is therefore on a few issues for the evaluation of policy, programmes and institutions. Although

14. This section centres on issues discussed at an OECD workshop held in Paris in October 2007.

targeted towards research policy/programme expert review, most of these issues apply to many kinds of expert review including project selection review.

Issue 1. *Socioeconomic factors: How to effectively reflect socioeconomic and political priorities and link these to decision making in the expert review processes.*

This issue is arguably the most important issue in policy making and evaluation. Some decision makers doubt whether expert review is able to reflect socioeconomic and political priorities. Expert review is, in fact, likely to ignore wider social and economic effects owing to its strongly scientific orientation. Expert review panels depend on sound, detailed information on a programme's progress or impact, and poor or insufficient information will affect the value of the results. As the type of data needed for retrospective impact assessment does not lend itself to an expert review panel format, expert review tends not to be appropriate for evaluating the impact of programmes (Ruegg and Jordan, 2007).

A couple of solutions can be suggested. To begin with, reviewers could receive advance analysis of socioeconomic needs and priorities. For example, the Korean government has provided evaluators of the National Master Plan of Science and Technology with the results of technology foresight, expenditure priorities at national level, the status of public R&D expenditure, and the analysis of programme portfolios and performances during the R&D programme evaluation process (Oh and Kim, 2006).

The most common solution is diversification of the fields of experts. Although it is reasonable to compose a review panel of peers from the relevant field to assess the excellence of the research proposal and to judge whether or not to award a grant, similar expert panels may be inappropriate when evaluating a programme or a policy which aims to address general social and economic problems. To put it simply, it is important to seek balance when selecting the review panel (see Box 2.1).

Evaluators of R&D should have not only have technical expertise but also a perspective on broader issues (for example, the impact of the research, mandate of the programme, economic utility, political and economic effects, etc.) (Klahr, 1985; Marshall, 1996). Although it would be ideal for an evaluator to have both, evaluators can complement one another to provide the expertise and the necessary broad perspective. Some of the experts should have a non-S&T background and have expertise in economics, business, accounting, public relations and policy, industrial policy, and other areas. Even when projects relate to a specialised technological area such as biotechnology or nanotechnology, the social and economic impact, needs, relevance and value of those projects for society as a whole should be considered as important as the scientific merits of the technological advances.

Box 2.1. Balancing expertise on the peer review panel

The most important aspect of "balance" is to include a range of intellectual perspectives that are respected in the scientific and technical community. Considerations in developing the balance also include:

- balance between technical specialists and multidisciplinary types, while ensuring adequate coverage of critical technical disciplines for each project and the overall programme;
- balance between academic, industrial, national laboratory, governmental and non-governmental organisation perspectives, as well as that of customers;
- balance between "old hands" and "young bloods";
- gender balance;
- geographic balance, possibly including international expertise and perspectives;
- balance across time, maintaining some continuity with prior peer reviews;
- in some cases, if appropriate, balance across interest groups, including representation from environmental, labour and other organisations, particularly for higher, programme-level reviews.

Source: EERE (2004), *EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review*, August 2004.

The question of how science can be made more relevant to the needs of society is increasingly central not only to the science policy debate but also to project selection (Scott, 2006; Nightingale and Scott, 2007). In Canada, political and socioeconomic priorities are increasingly considered. At NSERC, expert review is mostly used to evaluate applications for research grants. Generally, programme managers guide the work of panels regarding goals, criteria and applicable polices, but they are not involved in the review itself. In some programmes, officers make recommendations based on the review input and analysis of merit relative to the selection criteria. Most panels have members from the industry, government and university sectors. The panels also include a mix of national and international geographic representation, stage of career, gender, language and size of institution. This diversity works well in "problem/priority areas" (OECD, 2007d).

Another solution is to establish a dual review committee, in which one review group focuses on the scientific and technological excellence of the subject of the evaluation, while the other focuses on the relevance and socioeconomic priorities of research. For example, the US National Institutes of Health (NIH) has such a system for grant applications (Scarpa, 2006):

- The first level of review by a scientific review group (SRG) provides initial scientific merit, reviews grant applications, rates applications and makes recommendations for the appropriate level of support and duration of the award.
- The second level of review by council makes a recommendation to institute staff on funding, evaluates programme priorities and relevance, and advises on policy.

This dual review system makes it possible to carry out a proper evaluation of both the scientific/technical quality of the research and the socioeconomic value and utility.

"Bicameral review"‒ a slightly modified expert review – may also be used to evaluate programmes. In this case, research grants are assessed according to two different, independent criteria. One is the past accomplishments of the researcher and the other is the proposed research project. The former is assessed through peer review; the latter is assessed internally, based on the budget (Forsdyke, 1991; Forsdyke, 1993). These methods could be applied to evaluations in order to set priorities or allocate resources on the basis of past achievements and consideration of a country's strategic priorities and budgetary concerns. Focusing on past achievements, however, may bring new challenges, as this may favour better-known researchers and lead to the above-mentioned "Matthew effect". It is important to find ways to use past achievements while maintaining openness to new researchers and fields.

It is also possible to use the Delphi method, which is frequently used in technology foresight. The two-stage Delphi procedure is used to preclude bias and misunderstandings. In the first stage, each grant applicant receives feedback consisting of questions, criticism and advantages/disadvantages from four to six evaluators. In the second stage, on the basis of the grant applicant's responses, the evaluators assess the applicant's abilities, objectives, methods and the general level of research (Pouris, 1988). As an example, the Netherlands assumes that "when evaluating scientific projects, the best standards come from outside the field". Thus, when evaluating a grant in the field of physics, evaluators are selected from fields such as physics, chemistry, mathematics and astronomy in order to include "relevance" in the set of criteria. This improves the quality of the evaluation, as evaluators are made aware of opinions from other fields. Most importantly, it promotes relevance as a criterion in national research development policy. Given this, the Delphi method could be very useful for evaluating programme/policy.

Issue 2. *Interface of expert review with other means of judgement: How to use objectives indicators or ranking tables effectively in order to enhance the objectivity of evaluation result. How to combine expert reviews with other quantitative and qualitative methods for evidence-based policy.*

With the rise of indicator-driven judgement and ranking tables, the interface of these with expert review is an important issue. Policy makers and R&D programme managers have attempted to adopt more quantitative, indicator-based evaluations to complement or replace expert review. The problem is to identify performance indicators that are closely linked to the desired outcomes and to expert review procedures. Programme theory (or a logic model) is frequently used to develop the most suitable performance indicators for programme or policy evaluation.

It is important to raise the quality of expert evaluation by employing both qualitative and quantitative methodology properly. Each of these methods has its advantages, and weak points in the current evaluation system can be overcome by taking advantage of the strong points of each. Currently, there are few examples of combining expert review with other tools to improve the evaluation system. For instance, surveys, case studies, sociometric/social network analysis, bibliometrics and historical tracing have been used along with expert judgement in evaluations of the US Advanced Technology Program (Ruegg and Feller, 2003). The technology development programmes of the US Department of Energy (DOE) utilise expert review to evaluate R&D activities at the project and programme levels. In addition to expert review, R&D programme managers at DOE are encouraged to use other evaluation methods to obtain information on programme effectiveness and the benefits generated that cannot be provided using the peer review method (Ruegg and Jordan, 2007). In terms of quantitative indicators, expert reviewers need to be provided with condensed, systematic, verified, objective information on research performance, and the grounds for their judgement, or the assumptions underlying it, should be made more explicit, thus making the process more transparent (Moed, 2007).

Issue 3. *Cost efficiency of expert review: How to enhance the cost effectiveness of the various phases of the expert review process.*

Given that the aim of evaluation is to create new value, its benefits should outweigh the costs. These costs are easily underestimated because they are usually an implicit opportunity cost rather than an explicit payment. For instance, evaluators have to sacrifice working time and performance to carry out an evaluation. Research indicates that the indirect costs of the time spent by evaluators, presenters, staff and others are ten times or more than the direct costs, such as travel expenses (Kostoff, 1996). Evaluators should therefore do their best to achieve cost efficiency.

Efforts should also be made to reduce direct expenses. Expenditures vary depending on the number of projects reviewed, the number of reviewers, whether the meeting is open to the public and the length of the review. Typically, meeting logistics are a major cost of an expert review. Scheduling the event using public facilities, meal planning and audiovisual requirements should be organised well in advance of the actual meeting. Ways of controlling the cost of the review meeting include structuring the agenda carefully so that time is used efficiently and making maximum advance use of teleconferences, videoconferences and other electronic media to prepare the review panel. This is particularly helpful when international reviewers are involved (EERE, 2004).

Building a database of evaluators can help reduce the cost of selecting evaluators. Given the present internationalisation of science, there is much value in promoting international co-operation in building evaluator databases.

Ways to reduce the cost of research project evaluation have also been suggested. For example, Klahr (1985) points out that National Science Foundation (NSF) was able to reduce by one-third the number of final proposals to be evaluated by using a screening method to compare the results of “mail review” at the first stage of the evaluation with those of the “panel review” at the second stage. In addition, the NIH runs the Center for Scientific Review (CSR) to maximise the efficiency of the evaluation process. The CSR also operates streamlined review procedures (SRP) in order to concentrate on quality proposals (Lee, Om and Ko, 2000).

Efficiency can also be increased through various tools supported by the Internet, which can deliver real-time news and information and facilitate networking among those concerned. For instance, NSF operates the NSF Fast Lane System for more effective, convenient and faster administration through applications for preparing, submitting and revising research proposals (*www.fastlane.nsf.gov*). NIH has also announced assessment of research proposals through SRP via the Internet so that researchers are made aware of the most recent evaluation criteria and policies (*www.drg.nih.gov/refrev.htm*).

Issue 4. *International frame of reference: How to develop an effective international frame of reference for expert review.*

A review panel can be national or international. A national panel is composed of local experts in the field. This approach is useful in large countries with well-developed S&T systems and a large pool from which to choose experts and where there is less possibility of subjective evaluation. An international panel is mainly composed of foreign and internationally recognised experts. Each approach has certain weaknesses. The first may be unable to cope effectively with lobbying by interested groups within the scientific community, while the second may suffer from external evaluators’ lack of familiarity with the country’s particularities and the expertise might be affected by the fact that the evaluators come from different scientific environments (OECD, 1998).

The internationalisation of expert panels is needed more in countries that have a small S&T community. In Finland and Portugal, for example, proposals are submitted in English to increase the pool of international reviewers. Moreover, studies have shown that evaluation results of foreign experts are better accepted (Pouris, 1988; OECD, 2007). The Korean government is also aware that the Korean S&T society and expert pool is very limited. The government thinks the internationalisation of evaluation would help enhance objectivity and reliability and therefore tries to enlarge the expert pool to include foreign experts. At the same time, however, Korean policy makers know that it is difficult for foreign experts to evaluate Korean R&D programmes because they are unfamiliar with the Korean scientific community, the context of programmes and related policy, and national strategies (OECD, 2007).

Evaluations require criteria, standards and benchmarks to assess the quality and achievements of policy, programmes, projects or institutes. Given the internationalisation of research, expert reviews and evaluations at every level need to rely on international reference points to measure outcomes. This reflects in part growing concerns about competitiveness in science, but caution is needed when comparing different cultures and contexts.

Issue 5*. Managing conflicts of interest: How to manage conflicts of interest in the expert review process.*

One of the basic hypotheses of expert review is that experts' judgements are trustworthy and reliable. The presumption is that these persons have judgement, experience and a professional ethos. Nonetheless, evaluators' decisions may be affected by personal relationships, and this may prevent the evaluation process from being impartial and objective. This is why it is necessary to manage potential or existing conflicts of interest in the expert review process. In many reviews, reviewers must sign a conflict-of-interest form prior to the beginning of the review process. In addition, reviewers agree to disclose actual or perceived conflicts of interest as soon as they are aware of them during the review.

The United States Office of Management and Budget's Peer Review Standards points out several factors that are relevant to whether an individual is an impartial evaluator. They include whether the individual: *i)* has a financial interest in the matter at issue; *ii)* has, in recent years, advocated a position on the specific matter at issue; *iii)* is currently receiving or seeking substantial funding from the agency through a contract or research grant (either directly or indirectly through another entity, such as a university); or *iv)* has conducted multiple peer reviews for the same agency in recent years or has conducted a peer review for the same agency on the same specific matter in recent years (OMB's Draft Peer Review Standards for Regulatory Science under Executive Order 12866, August 29, 2003).

A way to avoid such a problem is to exclude evaluators who might have interests similar to those of the proposers, although it is nearly impossible to nominate experts to a review panel with absolutely no common interests. A fundamental dilemma is the link between reviewers who are indeed peers and the increased chances of a conflict of interest (Wood and Wessely, 2007). A solution is to include evaluators with conflicting views rather than exclude certain reviewers altogether.

Conflicts of interests may also occur between an evaluation manager and a reviewer. Those conflicts are usually related to the questions such as "who is responsible for the evaluation results?" and "how deeply should a manager and a reviewer be involved in decision making?" For example, both an expert reviewer and a manager may want to make a final decision on the proposals, rather than just support the other's decision. Severe disagreements over resource allocation decisions in the expert review process often come from conflicts of interest among parties involved in the evaluation process. Proper management of conflict resolution would enhance mutual receptivity. Furthermore, it is important to construct review mechanisms that focus on objective evidence.

A possible solution to potential conflicts of interest is to ask evaluators to prepare a declaration of interests (Bozeman, 1993). The UK Research Assessment Exercise requires declaration of interests in order to avoid obvious or potential conflicts of interest. It is even argued that authors of papers should declare their financial interests (RAE, 2001). For instance, the scientific journal *Nature* requires authors of papers to declare their financial interests. Another solution is to use only foreign experts. The Academy of Finland invited British, American, West German and Swedish experts to evaluate the country's progress in inorganic chemistry. The assessment states that it succeeded only because the panel came entirely from beyond the frontiers of Finland (Dixon, 1987; Pouris, 1988).

It is advisable to limit individual reviewers' number of evaluations or the duration of their evaluating activities so that they do not develop relations with the evaluated bodies and become susceptible to lobbying. Moreover, their previous judgements may make it difficult for them to take a fresh look at similar programmes. At the same time, if experts only participate in a single evaluation, they may feel less of a sense of responsibility than if they are expected to participate in future evaluations.

Lastly, it is advisable to prevent experts with expertise in only one domain from judging the quality or value of the entire project, as they might insist on the allocation of more resources to their area of expertise. An expert's opinions should be shared with other experts and used in the joint decision-making process. For example, an expert on biotechnology who evaluates a biotechnological programme may know more about technical aspects of the programme than other experts but may be unable to judge the socioeconomic value of the programme.

Issues 6. *Expert review in the Internet age: What opportunities does the Internet give for improved and enhanced expert review? Could an Internet-based "open evaluation" tool organised by the scientific community be an alternative to the classical approach? Can network-centred expert review replace classical review? Is evaluation possible without expert review panels?*

The Internet provides not only new means and modes of communication, but also opportunities for advanced evaluation. Panel review and mail review are the main types of expert review and are not very efficient in terms of time and expense. It has been shown that using the Internet to construct the panel and evaluate proposals enhances efficiency. Most importantly, evaluation systems based on the Internet dramatically boost efficiency because evaluators can access the information they need at any time and from any place once the relevant information is available. They can also use search engines and alert systems as well as data analysis tools.

The Internet could give rise to a new style of expert review. Whether a panel review or a mail review, traditional peer review consists of a "closed evaluation" by the nominated experts group. An Internet-based "open evaluation" tool organised by the scientific community could bring in additional evaluators from around the world. Moreover, once the results of the evaluation of a project or an evaluation are posted on the Internet, they can be reviewed by people around the world. Open evaluation is also a very powerful tool for addressing data fabrication. In fact, the first people to call attention to data fabrication in Korea's "Hwang affair" were Internet users. Pinpointing errors and data fabrication is almost impossible during the normal, relatively short, panel review process but it is difficult to deceive all potential reviewers on the Internet.

The publishing system of *Journal of Atmospheric Chemistry and Physics* gives a good example, with its Interactive Open Access Publishing (Mehlhorn, 2006). Papers are handled in two phases:

- In a first phase, the author submits a paper to the editor. The paper is published in *Journal of Atmospheric Chemistry and Physics Discussion* as a paper for discussion. The paper is openly reviewed by the scientific community as well as appointed referees with reactions by the author.

- In a second phase, the author is required to submit a revised paper based on comments from referees and the scientific community. The editorial board makes a final decision about publishing the final revised version.

This system has many advantages. It provides authors, referees and readers with rapid publication (authors and readers); direct feedback and public recognition for quality papers (authors); prevention of obstruction or plagiarism (authors, referees); documentation of critical comments, controversial arguments, scientific flaws and complementary information (referees and readers); deterrence of careless, useless and erroneous papers (referees and readers); public discussion and final revision (readers). In short, JACP's publishing system appears to provide maximum quality assurance through public, interactive and collaborative peer review. The system assumes of course that the readers of JACP are informed enough to offer meaningful review of research.

Information technology, such as groupware, also has the potential to significantly improve the efficiency and overall value of the expert review process. It offers the expert review process real-time data entry, screen sharing, data manipulation and statistical analysis capabilities. Individual reviewers can enter review and rating data anonymously, and the review manager can compute summary rating statistics to be shared in a timely manner. This increased information handling can free up time to for important reviewer-to-reviewer or reviewer-to-review manager interactions. Box 2.2 compares a network-centric expert review with the traditional review process.

Box 2.2. Comparing a groupware-based peer review with the traditional review process

Traditional peer review	Network-centric peer review
• Data input is completed via the evaluation form during the Q&A session or shortly thereafter.	• All the members of the on-site audience are linked by groupware information technology. All data input is digitised and instantly recorded.
• Each reviewer completes his/her evaluation during the session, and individual and panel summary results are computed at the end of each presentation day or after the review has concluded.	• Each reviewer completes his/her evaluation during the session using the groupware. During the presentations, the reviewers enter final ratings and any additional comments they believe are important based on last-minute observations or insights. Individual and panel summary results are made available in real time and routed back to each individual for further discussion.
• Statistical analysis of reviewer comments typically is not available instantly or in time for use in on-site panel discussion.	• Statistical analysis of reviewer comments is completed on site to provide useful performance data quickly.
• Reviewers could meet in closed session to discuss their preliminary reviews. However, during closed session discussion, reviewers often do not have access to the full statistical analysis of ratings for the panel.	• To complement the groupware tool, reviewers could meet in closed session to discuss the preliminary reviews, and once the interactive cycle is complete, they may make final changes to their individual review comments and ratings. The groupware technology would enable reviewers to have access to the full statistical analysis of ratings for the panel.

Sources: www.inform.nu/Articles/Vol2/v2n1p11-18.pdf and Ronald N. Kostoff (2001), *Network-centric Peer Review*, Office of Naval Research.

Issue 7. *Expert review for policy, programme and/or PROs: What type of expert review is appropriate for the evaluation of policies, programmes or PROs? Is expert review a relevant tool for evaluating research institutions?*

Peer review is generally used at the project level. It may also be suitable for programmes, policies and institutions. Because the outcome of the review can affect the final decision-making process, it is necessary to consider which type of peer review is most appropriate for upper-level decision making. Three considerations should be taken into account: the intended impact on final decision making, the type of review panel, and whether to open the process to the public.

There are three possible categories of peer review based on the level of impact on the final decision-making process: pre-emptive peer review, traditional peer review and ancillary peer review (Bozeman, 1993). In pre-emptive peer review the final decision depends entirely on the results of the peer review and a programme manager has no right of judgement. Following a pre-determined format, the peer review employs either a scoring or ranking model. NIH's dual review system, mentioned earlier, is an example of this type of review.

Traditional peer review is similar to pre-emptive peer review in that the results are the most important factor in determining the final decision. Unlike pre-emptive peer review, however, the decision in traditional peer review is influenced by other factors, such as the decision of a programme manager. Here, both the academic standards of the researcher's affiliated organisation and the geographic area are considered in addition to the peer review results. NSF typically uses this method.

Ancillary peer review only provides partial information and plays a minor role in the decision-making process.

In the search for the most suitable type of peer review, different aspects of the evaluation process should be considered. Evaluation methodologies view economic and political areas and the geographical distribution of scarce resources differently. These factors are often important when completing a major programme evaluation or for building up a science complex.[15] In such cases, traditional peer review is used rather than pre-emptive review because it allows for the consideration of multiple policy issues.

Beyond selecting the type of review to use, it is also necessary to select the type of review group to be used at the beginning of the review process. Although there are many types of external expert reviews, two types call for special attention: the independent panel and the group of external reviewers (EPA, 2000; Kostoff, 2003; Kostoff, 2004). The independent panel is a group of experts independent of the agency and typically funded under a contract. It has a chairperson, who attempts to reach consensus on relevant issues and write a report containing the results of the review and sometimes recommendations. An independent panel reports to the agency review manager.

In contrast, the external reviewers group does not have a chairperson; the review manager plays this role. The group may engage in technical discussions during the course of the review, but it does not reach a consensus. There may be individual written inputs from each group member, but there is no group report. Instead, the agency review manager writes the review report based on the evaluators' individual written inputs and other considerations. The review manager should have a solid technical background and some

15. Within the general category of expert review, there are a number of sub-types according to the level of specialisation and professionalisation (Gibbons and Georghiou, 1987; Rigby, 2002): traditional peer review (canonical academic review), direct peer review, modified direct peer review, pre-emptive peer review, indirect peer review, merit review (extended form of peer review), ancillary peer review, expert panels, panel review, professional evaluators.

understanding of the subject matter in order to write a credible report, select the appropriate mix of reviews, and conduct all aspects of the review.

Each of the two approaches has value for specific applications. The group of external reviewers is less formal and has fewer restrictions. It is useful for internal reviews when structural programme issues are paramount and require resolution or improvement, and where comparison with other programmes is not a major focus. The independent panel is more formal. It has more specific constraints/requirements regarding reviewers, meetings and audience selection. From the agency's perspective, either group is very useful for addressing the agency's programme improvement needs. From an external perspective, the independent panel has greater credibility because it is independent. This makes the independent panel more appropriate for priority setting.

It is also important to determine whether to make the expert review process publicly accessible. Those in favour of reviews that are open to the public (EERE, 2004) suggest that open review meetings can: help sharpen the questions raised; improve the transparency of the process; help improve or legitimise the technical or management approach; strengthen integration networks for research, deployment, delivery or business management; broaden public learning by providing an opportunity for individuals to learn what others are doing and how they manage their work; and encourage participants to improve their performance given the pressures of public presentation to peers. Making the evaluation process public has a more positive effect for the evaluation of programmes or institutions than for evaluation for priority setting.

Paths to a high-quality expert review

As the issues listed above demonstrate, several considerations should be taken into account when designing an expert review process. This section discusses components that certain experts believe necessary for high-quality expert reviews. It also summarises much of what has been discussed and sketches out certain principles for improving expert review.

Essential requirements for good practice

Chubin (1994) suggests that the quality and credibility of peer review can be enhanced through attention to seven areas: effectiveness, efficiency, accountability, responsiveness, rationality, fairness and validity. In practice, it is all but impossible for peer review to make improvements in all these areas. Moreover, some experts consider that peer review embodies tensions between five value pairs, *i.e.* desirable properties that are in tension with each other (Hackett 1997; Scott 2006):

- *Effectiveness and efficiency*. Increasing the effectiveness of a review will require more work on the part of peer reviewers and will therefore impose greater costs, while greater efficiency usually comes at the expense of thoroughness.

- *Autonomy and accountability*. Wider accountability may reduce autonomy; more autonomy implies less public accountability.

- *Responsiveness and inertia*. Perpetuating research inertia in traditional fields may stifle a peer review’s responsiveness to new issues and research.

- *Meritocracy and fairness*. A poor paper by a respected academic may be published because of his/her reputation.

- *Reliability and validity*. Reliable criteria may be narrow and rigid and may not produce the most valid results.

It is important to find the optimal balance between such contradictory requirements. Trade-offs are inevitable: research funding bodies constantly face the challenge of determining what constitutes a defensible, appropriate and workable balance (Wood and Wessley, 2007). For instance, although pursuit of greater effectiveness can enrich the precision of the evaluation, it requires too much time and resources and reduces cost efficiency. On the other hand, a focus on cost efficiency may lead to a superficial assessment. Evaluation designers need to consider all resources and conditions and attempt to achieve the best possible balance in the evaluation process and method.

Autonomy is a key value of the professional community, but it often conflicts with accountability. Scientists, as experts, like to decide what research to do and how to do it. The general public, however, wants to see scientists whose work is supported by taxpayers accountable for their results and performance. The dissemination of performance-based budgeting has also increased the emphasis on the accountability of public research. It is very important for a successful peer review to find the optimal balance between these values in the evaluation process.

Expert review is arguably one of most flexible methods for determining value, but requires meeting a number of critical pre-conditions. Much of the literature addresses the requirements of expert review (especially for peer review of project evaluation). Rigby (2002) suggests four essential pre-conditions for applying peer/expert review:

- Experts with knowledge of a particular area must be available and willing to participate. It is important for programme evaluators and responsible bodies to maintain access to expert networks because it can be difficult for them to identify relevant peers, as they are not usually part of the relevant scientific, social or professional networks.

- The panel of experts cannot be expected to answer questions that are beyond the scope of the available knowledge. Terms of reference need to be set with a sense of what groups of experts can be reasonably expected to know, infer or judge from their collective knowledge.

- The panel should only be asked to come to a judgement on a single area of knowledge or expertise. Peer review is known to be weak when comparative judgements between different fields of expertise have to be made.

- While the costs of peer review are low, sufficient resources should be made available to facilitate the work of the panel.

By definition, a high-quality peer review should provide an accurate picture of the intrinsic quality of the research being reviewed. The fundamental problem is the lack of absolute standards for measuring research quality. At present, evaluation of intrinsic research quality is still a subjective process, and depends on the reviewers' perspectives and past experiences. A high-quality review under these circumstances occurs when two conditions are fulfilled: *i)* use of highly competent reviewers, and *ii)* lack of distortions in the reviewers' evaluations as a result of biases, conflicts, fraud or insufficient work (Kostoff, 2004). According to Ormala (1989), high-quality expert review processes require at a minimum:

- The method, organisation and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation.

- Different evaluation levels require different evaluation methods.

- Programme and project goals are an important consideration when an evaluation study is carried out.

- The basic motive behind an evaluation and the relation between an evaluation and decision making should be openly communicated to all parties involved.

- The aims of an evaluation should be explicitly formulated.

- The credibility of an evaluation should always be carefully established.

- The prerequisites for the effective utilisation of evaluation results should be taken into consideration in the design of the evaluation.

EERE's *Peer Review Guide* (2004) describes the minimum requirements for expert reviews of EERE's R&D programmes:

- **Scope of review.** On a regular basis, qualified and objective peers will review all EERE programmes and projects in both their technology development and business administration offices. This should typically cover 80-90% of R&D funding and supporting business analysis and management programmes. Earmarked projects will be included in the review and treated on the same basis as other activities.

- **Frequency of review.** All EERE programmes and their key projects will be reviewed, on average, every two years, depending on the characteristics of the programme and needs for information.

- **Timely preparation.** Preparation for a peer review will include designation of a review leader, determination of the purpose of the review and the review agenda, and communication of this information to reviewers and those being reviewed in time for them to prepare for the review.

- **Core evaluation criteria.** Clear standards for judging the programme or projects will be defined prior to the review. This includes the criteria and the kinds of evidence (data) needed to judge those criteria. At a minimum, programmes will be assessed on quality, productivity and accomplishments; relevance of programme success to EERE and programmatic goals; and management.

- **Reviewers.** There will be a minimum of three reviewers for each programme element or smallest unit that is assessed and reported on. Each reviewer will be independent, competent and objective, and be selected by a transparent, credible process that involves external parties. The reviewers will cover the subject matter jointly. Reviewers will sign conflict-of-interest forms prior to the review and nondisclosure agreements if/when proprietary information is presented or discussed.

- **Plan for collecting reviewer data.** Review leaders will plan ahead for how review inputs will be documented, analysed and reported, as well as how individual reviewer comments will be tracked, while at the same time maintaining their public anonymity. The review agenda will allow sufficient time for a rigorous question and answer period for reviewers. Reviewers will be encouraged to support their comments with citations or data wherever possible.

- **Producing the peer review report.** The peer review report will faithfully reflect the full range of reviewer comments. The report should also include all individual inputs from reviewers and will be reviewed by the panel chair and/or the review panel before release.

- **Programme manager review and response.** Before the report is finalised and goes to senior management, the programme manager or office director will add written responses to peer reviewer findings and recommendations, including actions to be taken to improve the programme.

- **Peer review report distribution.** The final peer review report will be promptly communicated to senior management, associated staff and researchers involved with the R&D programme or project, and all persons involved in the review, and the report will be made available publicly.

- **Peer review record and *ex post* evaluation.** A peer review record will be established at the beginning of, and maintained throughout, the review process. The record should contain the final form of all the key documents of the review for all phases of the review. An evaluation of the peer review process is necessary to aid continuous process improvement.

Based on a variety of experiences, the peer review literature, and the management of hundreds of peer reviews, Kostoff (1995, 1997, 2001, 2003, 2004) concludes that the followings factors, in order of importance, are critical to a high-quality programme evaluation:

- **Senior management commitment**: Senior management’s commitment is the most important factor in the quality of an organisation’s S&T evaluations.
- **Evaluation manager’s motivation**: The operational manager must be motivated to perform a technically credible evaluation.
- **Statement of objectives**: There must be clear and unambiguous transmission of the review’s objectives, methodologies, potential impact and consequences to all participants.
- **Competency of technical evaluators**: Technical evaluators must be highly qualified in terms their role, objectivity and competency.
- **Selection of evaluation criteria**: The criteria will depend on the interests of the evaluation audience, the nature of the benefits and impacts, the availability and quality of the underlying data, the accuracy and quality of the results desired, the complementary criteria available, and the diagnostic techniques available for a full analysis, the status of algorithms and analysis techniques, and the capabilities of the evaluation team.
- **Relevance of evaluation criteria to future action**: Every S&T metric, and its associated data, should answer a question that contributes to the basis for a decision.
- **Reliability of evaluation**: The reliability and repeatability of an evaluation is also crucial. To minimise repeatability problems, a diverse and representative segment of the overall competent technical community should be involved in the construction and execution of the evaluation.
- **Evaluation integration**: A sound evaluation process should in general be seamlessly integrated into the organisation’s business operations. Evaluation processes should not be incorporated in management tools as an afterthought (the typical practice today), but should be part of the organisation’s front-end design.
- **Global data awareness**: Data awareness is also important. Placing the technology of interest in the larger context of technology development and availability worldwide is absolutely necessary.

- **Normalisation across technical disciplines**: For evaluations that will be used as a basis for comparison of S&T programmes or projects, the next most important factor is normalisation and standardisation across different S&T areas.
- **Secrecy**: Secrecy is as important as normalisation: reviewer anonymity and reviewer non-anonymity. "Blind reviewing" has been used to providing fairer evaluations.
- **Cost of S&T evaluations**: Cost is a critical factor in the quality of S&T evaluation.
- **Maintenance of high ethical standards**: A final critical factor, and perhaps the foundational factor in any quality S&T evaluation, is maintenance of high ethical standards throughout the process.

Principles and suggestions for successful expert review

Several scholars have suggested some principles or policy recommendations for successful expert review (Bozeman, 1993; Rigby, 2002; Ormala, 1989; EERE, 2004; Kostoff, 1995, 1997, 2001, 2003, 2004; Nightingale and Scott, 2007; Moed, 2007; Donovan, 2007; British Academy, 2007; ESPRC, 2008; Noble, 1974; Gillespie *et al*., 1985; Bodden, 1982; Porter and Rossi, 1985; GACR, 2007, etc.). For example, the UK EPSRC suggests good peer review principles for reviewing research proposals: transparency, appropriateness, managing interests, confidentiality, expert assessment, prioritisation, right to reply, separation of duties and no parallel assessment (Box 2.3). The US OMB provides another example. The OMB 2001 guidelines set general criteria for competent and credible peer review: *i)* peer reviewers should be selected primarily on the basis of necessary technical expertise, *ii)* peer reviewers are expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, *iii)* peer reviewers are expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and *iv)* peer reviews should conducted in an open and rigorous manner.

Box 2.3. EPSRC's peer review principles

Transparency. Publish the criteria for assessing proposals and details of the peer review process before applicants submit proposals, defining how the assessment process will operate and be managed.

Appropriateness. Use a peer review process that is appropriate to the type of proposed research and in proportion with the investment and complexity of the work.

Managing interests. Ask all participants to declare interests when carrying out peer review activities so that any conflicts can be identified and managed.

Confidentiality. Treat proposals in confidence and ask advisers to do the same.

Expert assessment. Use expert peer reviewers, mainly from EPSRC's college of reviewers, to assess the individual merit of all proposals against the published criteria.

Prioritisation. Prioritise proposals for funding by assessing the merit of each proposal against that of others if its expert assessment has been sufficiently supportive.

Right to reply. Give principal investigators the right to reply to the expert reviewers' assessments when proposals are being prioritised.

Separation of duties. Separate peer review of proposals against the assessment criteria from funding decisions. EPSRC staff will make funding decisions based on peer review advice, taking into account budgets available and the tensions between budgets. Those acting as peers will not be responsible for authorising the funding decision.

No parallel assessment. Avoid carrying out multiple parallel assessments of a proposal's relative merit.

Source: www.esprc.ac.uk/ResearchFunding/ReviewingProposals/Principles.htm.

Most of the literature offers suggestions and principles pertaining to the selection of research topics and the publication of scientific papers. Although these suggestions may be very useful for improving the policy-level or programme-level expert review process, they do not help when organising a high-quality expert review. On the basis of the preceding discussion, this section offers some principles and suggestions for organising such a review.

Principle 1. The philosophy, focus and future uses of an evaluation must be understood and agreed upon in advance by the relevant stakeholders. This principle applies not only to expert review, but also to every type of evaluation. Experts need to fully understand and agree upon the reasons for evaluation, the methods and principles guiding the evaluation, and the utility of the evaluation.

- ***High-level policy makers or evaluation managers should clearly define the roles of each actor in the evaluation process and regularly monitor performance.*** In all areas of public management, the encouragement and continuous interest of managers at the highest level is a key factor of success. Evaluation is more complicated than other tasks, not only because there may be a conflict of interest between evaluators and those evaluated but also because it involves third-party experts. It is therefore indispensable for high-level evaluation managers to clearly define the role of each actor and to ensure that agreement is reached well in advance among the relevant actors concerning the objective and the philosophy of the evaluation.

- ***Provide a pre-evaluation training programme for the relevant actors.*** Training relevant actors, experts, evaluation staff (the secretariat), and those evaluated (*e.g.* programme managers) on the evaluation process and relevant criteria can enhance efficiency, effectiveness and receptiveness.

- ***Before the evaluation, select objective and useful evaluation criteria***. Evaluation criteria are important because they determine the focus and scope of the evaluation. It is essential to provide clear evaluation criteria before embarking upon an evaluation.

Principle 2. Qualified experts should be selected as evaluators. Professional competence and objectivity is required of experts participating in the evaluation. The panel chair and other experts should have strong professional competencies in the areas in which they are required to make judgements in order to instil confidence in the evaluation stakeholders.

- ***In addition to technological experts, seek experts from diverse domains, including the social sciences and the economy.*** To judge a programme's rationale and its socioeconomic value, it is desirable to have, as well as technical experts, experts from fields such as economy and business management. This is very important for policy or programme evaluation.
- ***Build a sufficiently large database of experts.*** This calls for regular monitoring of research personnel in various research institutions and universities. The data that should be collected are: past research experience, current research interests, field, affiliation, degree-granting institution, participating academic organisations and other detailed academic activities. On this basis, it would be possible to learn which policy areas particular researchers would be able to evaluate and what contributions they might make if chosen as an evaluator.

Principle 3. The risk of bias or conflicts of interest should be reduced as much as possible.

- ***Provide a bias statement for reviewers.*** Experts should declare their interests to ensure the evaluation panel's reputation for fairness. In principle, the evaluation manager should not appoint an evaluator with a vested interest in the policy/programme or institution to be evaluated.

- ***Avoid "internal evaluators".*** It is useful to include opinions of experts from another field/region or to have them on the panel. In particular, if there is no language barrier or additional cost, it is desirable to include foreign experts in the evaluation. If they are not well aware of the socioeconomic conditions of the country in question, it is best if they focus on the scientific or the technical aspects of the programme.

- ***Limit the number of evaluations or the duration of evaluating activities of participating experts.*** The number of experts' evaluations or the duration of their evaluating activities should be limited so that they do not develop relations with the evaluated bodies and become susceptible to lobbying or become less able to take a fresh look at similar programmes. At the same time, if experts only participate in a single evaluation, they may feel less of a sense of responsibility than if they are to participate in future evaluations.

- ***Prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated.*** An expert's opinions should be shared with other experts on the panel with a view to joint decision making.

Principle 4. The review should be conducted in a credible, fair and transparent manner and with the highest ethical standards.

- ***Ensure that the evaluation process and evaluation results are transparent.*** Introduce transparency in evaluation principles, criteria and processes and make these available to all actors and stakeholders so that they can prepare for the evaluation properly. After the evaluation, diffuse the results to the evaluated bodies and to the general public, with the exception of those that may be confidential for national security reasons.

- ***Maintain high ethical standards***. To ensure evaluators' freedom from personal bias, bias statements can be required. These may include clauses on overcoming personal bias as well as on the prohibition of misuse of information obtained during the evaluation process, such as use of such information for personal reasons or the release of such information without the permission of the relevant authority.

Principle 5. The review should be based on objective evidence and information.

- ***Provide the evaluators, in advance, sufficient information on the policy/programme to be evaluated.*** The expert panel's judgement is based on the information they receive. Sufficient information is as important as selecting qualified experts.

- ***If indicators or ratings are used, test their validity and reliability***. Indicators are an important means of ensuring objectivity, and evaluators should examine their relevance and reliability.

- ***Encourage a maximum amount of dialogue and discussion.*** In an expert review discussions and mutual learning are a source of valuable ideas, and as much discussion as possible among evaluators and the evaluated (programme managers) should be strongly encouraged. Including experts from various fields would be an effective way to generate productive discussions and new ideas.

Principle 6. "One size does not fit all."

- ***Complement the expert review with quantitative methods to increase the objectivity and scientific reliability of the evaluation.*** The objectivity and accuracy of expert reviews can be increased by using quantitative methods such as bibliometrics or econometrics.

- ***Use the type of expert review that is most appropriate for the particular programme/policy.*** The review should be tailored to the aim of the evaluation and the characteristics of the subject of the evaluation. For example, if the primary objective is to set priorities, a scoring method could be used. If the improvement of a programme is the primary objective, opinions of experts would be very important. The evaluation process and the form of the final results should also be tailored to the particularities of programmes.

Principle 7. Efficiency can be increased in various ways.

- ***Increase remote evaluation.*** With the development of the Internet, evaluating institutions can distribute IDs and passwords to evaluators so that they can access data and submit reports on line. To promote efficiency, the evaluation manager should devise technologies that enable evaluators to participate in the evaluation process from a distance.

- ***Build and operate evaluation management systems through Internet-based technologies.*** Evaluation management systems (EMS) should cover fundamental information on the evaluation, including information on the pool of experts, evaluation data, evaluation principles, the evaluation protocol, relevant analytical data and evaluation results. To increase the use of EMS, evaluating organisms, evaluators and those evaluated should be able to use it freely. Restrictions for security reasons may be introduced when necessary.

- ***Minimise the costs borne by the subject of the evaluation***. Simplification of administrative procedures and evaluation formats can decrease costs. The evaluation subject is often asked by evaluators to provide unnecessary administrative or other information. The evaluation body can remove unnecessary information from the basic database provided to evaluators in order to reduce the administrative burden of those who are being evaluated. In addition, if the evaluators provide clear reasons when they ask for additional information, those being evaluated may co-operate more willingly.

- ***Pay attention to hidden indirect costs when designing the evaluation process***. In expert review, indirect costs, such as billable hours, are more important than direct costs, such as venue and travel expenses.

Principle 8. Improve the design of expert panels

- To ensure continuity, it is advisable to appoint someone who has participated in previous panels as the head of a panel.

- One-third to one-half of a review panel should be carried over from one review to the next to ensure both continuity and new perspectives.

6. Conclusion

Some common views on expert review seem to be emerging. First, despite problems such as hollowing out due to time constraints, rising financial costs and the risk of expert reviewers' conflicts of interest, the expert review process remains a fundamental mechanism for all stages of research planning and implementation. Second, ways to improve expert review processes are available: making the process more transparent, providing clear objectives and guidelines to reviewers, extending expert review to include non-scientific stakeholders, and using a variety of metrics and indicators. Third, while indicators can strengthen and inform judgements, these still require careful attention to prevent perverse outcomes. Fourth, the internationalisation of expert review needs to be facilitated and improved because of increased international collaboration, although caution is needed when dealing with different cultures and contexts. There is no single evaluation model, and the design requirements for expert review need to be better understood.

There is no perfect evaluation system; it must be adapted to the environment. What worked in the past may not work in the future. Also, when modifying evaluation systems, it is important to take into account the views of those evaluated as well as the evaluation managers. A client perspective helps to discover problems, as well as solutions. Admittedly, the opinion of those evaluated will depend to some extent on the evaluation results. It is therefore important to strike a proper balance by carrying out regular opinion surveys of participants in expert reviews.

Expert review involves more individual judgement than any other method of evaluation. The review manager, the experts and the stakeholders need to co-operate in order to realise a successful evaluation. This may be said to be main way to ensure a successful expert review.

REFERENCES

Bozeman, B. (1993), "Peer Review and Evaluation of R&D Impacts", in B. Bozeman and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*. Kluwer Academic Publishers, Boston, Massachusetts.

Bozeman, B. and J. Melkers (eds.) (1993), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, Boston, Massachusetts.

Campbell, P. (2006), "Pressures on peer review at *Nature* and Nature journals", Presentation at the Prague Conference on Peer Review: Its Present and Future State, October.

Chubin, D.E. (1994), "Grants Peer Review in Theory and Practice", *Evaluation Review*, Vol. 18, No. 1, pp. 20-30.

Chubin, D.E. and E. Hackett (1990), *Peerless Science: Peer Review and US Science Policy*. SUNY Press, Albany, New York.

COSEPUP (1999), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act.* Washington DC: National Academy Press, *www.nap.edu/catalog/6416.html*.

Dixon, B. (1987), "Follow the Finnish Lead in Peer Review", *The Scientist*, May 4.

Donovan, C. (2007a), "Introduction: Future pathways for science policy and research assessment: metrics vs. peer review, quality vs. impact", *Science and Public Policy* 34(8), October.

Donovan, C. (2007b), "The qualitative future of research evaluation", *Science and Public Policy* 34(8), October.

EERE (2004), *EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review*, August.

Environmental Protection Agency (EPA) (2000), *Peer Review Handbook*, Science Policy Council, Washington, DC.

EPSRC (2008), "Peer Review Principles", *http://www.epsrc.ac.uk.*

Forsdyke, D.R. (1991), "Bicameral grant review: an alternative to conventional peer review", *FASEB J. 5*, pp. 2312-2314.

Forsdyke, D.R. (1993), "Bicameral grant review: how a systems analyst with AIDS would reform research funding", *Accountability in Research* 2, pp. 237-241.

Forsdyke, D. R. (1993), On giraffes and peer review, *FASEB.J.* 7, pp. 619-621.

GACR (2007), *Peer Review: Its Present and Future State*, Czech Science Foundation.

General Accounting Office (1999), *Federal Research: Peer Review Practices at Federal Science Agencies Vary*, GAO/RCED-99-99.

Gibbons, M. and L. Georghiou (1987), *Evaluation of Research: A Selection of Current Practices*, OECD, Paris.

Gillespie, G.W. Jr., D.E. Chubin and G.M. Kurzon (1985), “Experience with NIH Peer Review: Researchers’ Cynicism and Desire for Change”, *Science, Technology and Human Values*, Vol. 10, No. 3, pp. 44-53.

Goodstein, T. (1995), “Ethics and Peer Review”, *Biotechnology*, Vol. 13, No. 6, p. 618.

Gustafson, T. (1975), “The controversy over peer review: recent studies of the peer review system show that its critics have yet to make their case”, *Science*, Vol. 190, pp. 1060-1066.

Guston, D.H. (2000a), “Expanding the role of peer review processes in the United States”, paper presented to the US-European Policy Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany.

Guston, D.H. (2000b), “Regulatory peer review”, Paper presented at the annual meeting of the American Political Science Association, Washington, DC, 2 September.

Guston, D.H. (2000c), *What is Peer Review?,* Report submitted to the American Chemistry Council, May.

Hackett, E.J. (1997), “Peer Review in Science and Science Policy”, in *Evaluating Science and Scientists* M.S. Frankel and J. Cave (eds.), Central European University Press/Oxford University Press, Budapest and Oxford.

Hartmann, I. and F. Neidhardt (1990), “Peer review at the Deutsche Forschungsgemeinschaft”, *Scientometrics*, Vol. 19, No. 5-6, pp. 419-425.

Klahr, D. (1985), “Insiders, outsiders, and efficiency in a National Science Foundation panel”, *American Psychologist*, Vol. 40, No. 2, pp. 148-154.

Kostoff, R.N. (1993), “Evaluating Federal R&D in the United States”, in B. Bozeman and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, Boston, Massachusetts.

Kostoff, R.N. (1997), Research program peer review: principles, practices, protocols, *www.dtic.mil/dtic/kostoff/index.html*.

Kostoff, R.N. (2001), *Network-centric Peer Review*, Office of Naval Research.

Kostoff, R.N. (2003), *Science and Technology Peer Review: GPRA*, Office of Naval Research.

Kostoff, R.N. (2004), *Research Program Peer Review: Purposes, Principles, Practices, Protocols*, Office of Naval Research.

Kruytbosch, C. (1989), “The Role and Effectiveness of Peer Review”, Paper presented at the Ciba Foundation Conference on Research Evaluation.

Lee, M., K. Om and J. Ko (2000), *Theory and Practice of Peer Review*, Pidi, Seoul.

Mehlhorn, K. (2006), “Peer Review in the Digital Age: Opportunities and Examples”, Presentation at the Prague Conference on Peer Review: Its Present and Future State. October.

Merton, R.K. (1973). *The Sociology of Science*. University of Chicago Press, Chicago, Illinois.

Moed, H.F. (2007), “The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review”, *Science and Public Policy* 34(8), October.

Nightingale, P. and A. Scott (2007), “Peer review and the relevance gap: 10 suggestions for policy-makers”, *Science and Public Policy* 34(8), October.

Noble, J.H. (1974), “Peer review: quality control of applied social research”, *Science*, Vol. 185, pp. 916-921.

OECD (1998), *Policy Evaluation in Innovation and Technology: Towards Best practices*, OECD, Paris.

OECD (2007), “Summary of the workshop on rethinking evaluation in science and technology”, internal working document.

Oh, D.H. and Y.J. Kim (2006), “Overall coordination of government-funded research and development programmes in Korea”, *Journal of Multidisciplinary Evaluation*, 5 (September), pp. 153-170.

Ormala, E. (1989), “Nordic Experiences of the Evaluation of Technical Research and Development”, *Research Policy,* Vol. 18, pp. 333-342.

Porter, A. and F. Rossini (1985) Peer review of interdisciplinary proposals”, *Science*, *Technology and Human Values*, Vol. 10, No. 3, pp. 34-42.

Pouris, A. (1988), “Peer review in scientifically small countries”, *R&D Management*, Vol. 18, No. 4, pp. 333-340.

Rigby, J. (2002), “Expert Panels and Peer Review”, in G. Fahrenkrog, W. Polt, J. Rojo, A. Tubke and K. Zinocker (eds.), *RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies*, IPTS Technical Report Series, EUR 20382 EN.

Rossi, P.H., M.W. Lipsey and H.E. Freeman (2004), *Evaluation: A Systematic Approach,* SAGE Publications, Thousand Oaks, California.

Ruegg, R. and G. Jordan (2007), *Overview of Evaluation Methods for R&D Programs*, Office of Energy Efficiency and Renewable Energy, DOE, Washington, DC.

Ruegg, R. and I. Feller (2003), *A Toolkit for Evaluating Public RTD Investment,* NIST GCR 03-875, National Institute of Standards and Technology, Washington, DC.

Scarpa, T. (2006), “Challenges and Opportunities Facing NIH Peer Review: A Vision for Ensuring Its Strategic National Value”, Presentation at the Prague Conference on Peer Review: Its Present and Future State. October.

Scott, Alister (2006), “Peer Review and the Relevance of Science”, *SPRU Electronic Working Paper Series*, No. 145.

British Academy (2007), “Peer review: the challenges for the humanities and social sciences”, www.britac.ac.uk/reports/peer-review/.

Wood, F. and S. Wessley (2007), “Peer Review of Grant Applications: A Systematic Review”, in F. Godlee and T. Jefferson (eds.) (2007), *Peer Review in Health Sciences*. BMJ Books.

CHAPTER 3

USING IMPACT ASSESSMENT TO IMPROVE THE EVALUATION OF OF PUBLIC R&D[16]

> This chapter explores the question of how experts or evaluators are to properly assess and quantify the social impacts of S&T policies or programmes. It reviews recent and emerging impact assessment practices, including the main methodologies, and highlights their assumptions and limitations. It offers suggestions to increase the effectiveness of future impact assessments.

Introduction

Impact assessment is central to the evaluation of public research and development (R&D). It involves more than measuring success in meeting past objectives. It is also about determining where, who and how much to fund research and anticipating what society will get in return. An impact analysis should help determine both the economic effects and the social impacts (*e.g.* better health outcomes) of public investment in R&D.

Impact assessment of public R&D is therefore closely intertwined with the evaluation of public R&D and should provide valuable feedback to the different phases of public policy formulation, including policy design. Public R&D impact assessment assists governments in their decisions to prioritise R&D resources and can help them design their research programmes. Moreover, it enhances public accountability, creates a better-informed society, and raises awareness of public research's contribution to a country's economic and social development.

This chapter first defines the nature and scope of the potential impacts of public R&D and the main challenges practitioners face when identifying and assessing them. It then distinguishes three main levels: *i)* overall public R&D investment in the research system; *ii)* public research organisations (PROs), including funding of research councils for research carried out or funded by specific institutions; and *iii)* research programmes. Finally, it presents practices for assessing the impact of publicly funded and performed research and for assessing systemic impacts (*i.e.* those affecting the economy or society) as well as sector-specific impacts.

Defining the objects and impacts of R&D

Many definitions of "impact" are used by evaluators and policy makers. In general, the definition used depends on: *i)* the nature of the impact: economic, scientific, technological, cultural, societal environmental, etc.; *ii)* the scope of the impact: systemic, organisational, firm-based; and *iii)* the timing of the impact: estimated, contemporary, *ex post*. The academic literature provides various definitions of the types of science and technology (S&T) impacts (Box 3.1).

---

16. This chapter draws and builds on Chapter 4 of OECD (2008), *OECD Science, Technology and Industry Outlook*, the 2008 TIP workshop on impact assessment, and on the report of the Research Institutions and Human Resources (RIHR) project on the evaluation of public research institutions.

Box 3.1. Eleven dimensions of the impacts of science

Science impacts: Research results have an effect on the subsequent progress of knowledge owing to advances in theories, methodologies, models and facts. They affect the formation and development of disciplines and training and can also affect the development of research itself, generating interdisciplinary, crosscutting and international research.

Technology impacts: Product, process and service innovations as well as technical know-how are types of impacts that partly result from research activities. There are few indicators for properly assessing this, other than patents, at least until work based on innovation surveys results in analysis of outputs and impacts as well as innovation activity itself.

Economy impacts: These refer to the impact on an organisation's budgetary situation, operating costs, revenues, profits and sale price of products; on the sources of finance, investments and production activities; and on the development of new markets. At the aggregate level, they can also refer to the economic returns, either through economic growth or productivity growth, of a given geographical unit. It is probably the best-known dimension.

Culture impacts: These relate to what people often call public understanding of science, but above all to four types of knowledge: know-what, know-why, know-how and know-who. In other words, these are the impacts on an individual's knowledge and understanding of ideas and reality, as well as intellectual and practical skills, attitudes, interests, values and beliefs.

Society impacts: Research affects the welfare, behaviour, practices and activities of people and groups, including their well-being and quality of life. It also concerns customs and habits: consumption, work, sexuality, sports and food. Research can contribute to changing society's views and "modernise" ways of doing "business".

Policy impacts: Research influences how policy makers and policies act. It can provide evidence that influences policy decisions and can enhance citizens' participation in scientific and technological decisions.

Organisation impacts: These refer to the effects on the activities of institutions and organisations: planning, organisation of work, administration, human resources, etc.

Health impacts: These relate to impacts on public health, *e.g.* life expectancy, prevention of illnesses and the health-care system.

Environment impacts: These concern management of the environment, notably natural resources and environmental pollution, as well as the impacts of research on climate and meteorology.

Symbolic impacts: These are gains in areas such as credibility due to undertaking R&D or linked to universities or research institutions that offer gains in terms of potential clients, etc.

Training impacts: These are impacts of research on curricula, pedagogical tools, qualifications, entry into the workforce, etc.

All but the first three dimensions are somewhat new to statisticians, as they are less tangible and therefore difficult to measure or evaluate. This typology provides a checklist to remind evaluators that research affects areas other than those usually identified and measured in the economic literature.

Source: OECD STI Outlook 2008 based on Godin and Doré (2006).

The different impacts can be diverse in scope as well as in nature. Impacts may accrue to society as a whole, to a particular group of people, to a research group, or to enterprises or other institutions. Identifying the type of impact to be measured is crucial when deciding on the choice of methodology or methodologies for assessing the impact of public R&D.

Key challenges for assessing the socioeconomic impacts of public R&D

It is difficult to determine and measure the various benefits of R&D investment for society. This is mainly because R&D spillovers and unintended effects are likely, many key scientific discoveries are made by accident ("serendipity"), and many applications of scientific research are found in areas very different from the original intention. Moreover, the time required for public R&D to generate its full benefits may be quite long, so that measurement of impacts may be premature and partial. Finally, the non-economic impacts of public research may be more difficult to identify and measure. For example, the measurement of health outcomes is not straightforward and complicates efforts to link health outcomes to public investment in R&D. Similar difficulties arise for linking investment in defence R&D to security outcomes or investment in energy R&D to energy security. As noted in the *OECD Science, Technology and Industry Outlook 2008* (OECD, 2008), the most important challenges encountered by science policy researchers and policy makers when analysing the impacts of public R&D can be summarised as follows:

- Causality problem: What is the relation between research inputs, outputs, outcomes and impacts?
- Attribution problem: What portion of the benefits should be attributed to initial research and not to other inputs?
- Internationality problem: What is the role of spillovers?
- Evaluation time scale problem: At what point should the impacts be measured?
- Definition of appropriate indicators: What are the appropriate indicators?

Because of these challenges, analysis has traditionally focused on developing and collecting R&D input and output indicators and establishing a direct relationship between them (Table 3.1). Since many of the impacts of R&D only emerge over time, this type of analysis often ignores many of the long-term benefits of public R&D for a country's economy and society.

Table 3.1. Traditional means of measuring "impacts" of R&D

	R&D inputs					**R&D outputs**		
	Total Public R&D (GOVERD + HERD) 2005*	**GOVERD 2005* (% of GDP)**	**HERD 2005* (% GDP)**	**Basic research 2005* (%GDP)**	**Researchers 2004* (per thousand of labour force)**	**Scientific articles per million population, 2003**	**Relative prominence of scientific literature, 2003**	**Share of PCT Patents owned by GOV + HE (2002/04)**
Iceland	1.28	0.66	0.62	0.53	13	701.8		
Sweden	1	0.24	0.76		10.8	1142.8	0.86	0
Finland	0.99	0.33	0.66		15.7	997.9	0.83	0.4
Canada	0.9	0.18	0.72		7.3	783.2	0.85	10.3
France	0.79	0.37	0.42	0.52	7.3	516.2	0.76	10.8
Austria	0.77	0.12	0.65	0.39	6.6	604.4	0.8	1.1
Australia	0.76	0.28	0.48	0.42	7.9	791.2	0.71	10.3
Denmark	0.76	0.18	0.58	0.46	9.1	981.6	0.94	3.2
Germany	0.75	0.34	0.41		6.8	536.9	0.82	1.7
Netherlands	0.74	0.24	0.5		4.5	830.6	0.97	1.4
Japan	0.73	0.28	0.45	0.4	10.2	470.3	0.58	4.4
Norway	0.71	0.24	0.47	0.28	8.9	731.4	0.72	0.5
Switzerland	0.7	0.03	0.67	0.84	5.8	1153.5	1.15	2.2
United States	0.68	0.31	0.37	0.48	9.5	725.6	1.03	10

GOVERD=Government expenditure on R&D; HERD=Higher education expenditure on R&D; GDP=gross domestic product; PCT=Patent Co-operation Treaty.

Source: OECD Main Science and Technology Indicators, 2008.

Moreover, econometric analysis of the relation between R&D and outcomes is typically based on a linear conception of innovation. It presupposes that innovation starts with basic research, followed by applied R&D, and ends with the production and diffusion of new products and processes in the economy. It is widely acknowledged, however, that innovation is more complex, with multiple feedback loops between stages and actors, and that innovation results from the interplay of public and private R&D investment, commercial interests and many other factors (see for example the churn model discussed in Chapter 1). As a result, a fuller understanding of the effects of science and innovation requires a more encompassing approach to measuring and analysing innovation and the economic and social impacts that accrue to society.

Approaches to impact assessment of public research in OECD countries

What methods can be used to assess the impacts of public R&D? Over the past decade, national governments and academics have worked to develop new analytical techniques for assessing the impacts of public R&D investment, such as econometric analysis, data linkage approaches and case studies. The outcomes and robustness of such analyses are heavily influenced by the nature of the methods used, the assumptions on which they rely and their inherent limitations. Impact assessment methodologies are not universally applicable; they are context-specific and depend on the objective of the impact assessment exercise, its timing (*ex ante* and/or *ex post*); and the scope and nature of the R&D.

The study reported in the *OECD Science, Technology and Industry Outlook* (OECD, 2008) found top-down approaches, especially econometric and mathematical models, better suited to assessing impacts affecting the whole research system and dealing with all types of basic and applied research. In particular, mathematical models, such as general equilibrium or similar models, may be a good way to assess systemic impacts *ex ante*. On the other hand, bottom-up approaches are preferable when the subject of the assessment is a research programme and/or institution that aims to develop a specific type of technology with a clear industrial focus.

Some of the most promising and forward-looking practices include general equilibrium models, econometric analyses, data linkages, scientometric methods, survey-based indicators combined with econometric analyses, and case studies. These various methodologies are still evolving, but they have opened new and encouraging lines of investigation, as the currently available techniques cannot capture the full range of the impacts of public R&D on society. The following reviews some of these emerging approaches and their advantages and disadvantages.

Econometrics-based impact assessments

Econometric studies have examined evidence on the contribution of R&D investment to economic growth both in microeconometric studies, which use data on firm and industry productivity to estimate the private and social returns of R&D investments, and in macroeconometric studies, which estimate the contribution of overall R&D investment to aggregate productivity.

Microeconometric studies have analysed productivity growth in private firms in a number of countries and for different periods of time. They have also assessed knowledge spillovers and calculated the social rate of return (*i.e.* the benefits that private R&D investment generates for other firms located inside and outside their own industry). In a seminal study on the effects of private R&D investments on total factor productivity (TFP), Lichtenberg and Siegel (1991) found that the gross rate of return was significant and up to 35% for company-funded R&D. For publicly funded R&D, however, they found little significant impact on productivity. Mamuneas and Nadiri (1994) also explored the social return of publicly funded R&D for US manufacturing firms by estimating the cost reductions associated with an extra dollar of public R&D investment. The results showed returns ranging from 8.7% to 5.8% and thus a positive social return to publicly funded R&D. Griliches (1986) also concluded that publicly funded R&D in industry had positive effects on productivity, although less than privately financed R&D.

In general, microeconometric studies have shown strong returns to private R&D investment and the presence of strong spillover effects that generate substantial economic benefits, although these vary by industry. There is relatively little evidence on the impact of public R&D investments on private productivity growth. Furthermore, the few existing studies provide inconclusive results. This may be because studies at the firm and industry level are unable to account for positive spillovers accruing from public R&D, which may only emerge at the national level. Moreover, as public research often concerns the pre-competitive stage, the link to immediate commercial applications and productivity growth is likely to be less direct.

Macroeconometric studies analyse the effect of overall R&D on national productivity and can capture the full extent of knowledge spillovers to different firms and industries. These cross-country studies also make it possible to take into account benefits that diffuse across firms and industries. Many of these studies investigate both the social returns to national R&D investment and the spillover effects of foreign R&D. Coe and Helpman (1995) calculated the stocks of domestic R&D using the perpetual inventory method with an assumed depreciation rate ranging from 5% to 15%, and calculated the effects on total factor productivity for 22 OECD countries for the period 1971-90. They calculated a marginal rate of social return[17] of 123% for the seven large OECD economies and 85% for the others. In this study, the specific effect of public R&D expenditure on productivity growth was difficult to assess because public and private expenditures were aggregated. A study by Guellec and van Pottelsberghe de la Potterie (2001) later filled in this gap and has been extremely influential.

The conclusions of this line of research, however, have been challenged (Sveikauskas, 2007) owing to the lack of detailed microeconomic evidence on the specific mechanisms through which public science affects productivity growth, such as more rapid growth of high-technology industries. Moreover, Khan and Luintel (2006) introduced a number of other potential variables such as education or public infrastructure that may explain productivity growth. They did not find that public R&D was a significant factor in productivity growth rates, thus suggesting the lack of a direct link between the two. Finally, other macroeconometric studies have provided only limited evidence on the role of public R&D investment in productivity growth. OECD (2003) analysed different contributions to growth rates in different OECD countries which might explain differences over time. Using cross-country regression analysis and a large set of variables that might explain observed differences in growth, the study concluded that private R&D has high social returns and contributes to economic growth, but that there is no evidence of this for government R&D. In general, macroeconometric studies have reported high social rates of return, above 50% in many cases, showing the positive effect of overall R&D investment on productivity growth. These studies also suggest that public R&D does not contribute directly to economic growth, but has an indirect effect via the impact on private R&D.

One limitation of these econometric studies is that they have ignored, at least until recently, the relations among different R&D actors that can provide insight into innovation processes resulting from R&D investment. Although econometric studies that take a linear view of innovation demonstrate associations between variables, they seldom demonstrate a causal link. Moreover, they focus on the relation between R&D and increased output or productivity. Other objectives of research, such as national security, energy security, environmental protection, health or social cohesion, are excluded from the analysis, as they are not captured by measures of economic growth. These objectives, however, need to be kept in mind when assessing the impacts of specific public R&D investment.

17. These estimates are calculated for the lower rate of capital depreciation of 5%.

Capitalisation of R&D

Currently, econometric work is being complemented by growth accounting analysis, which explicitly considers public and private investment in R&D as a source of productive investment. Inclusion of R&D in national accounts stems from the need to move from a traditional view of R&D as current spending to recognition that R&D should be seen as an investment in intangible capital that expands a nation's knowledge stock, while also providing benefits over a number of years. Although R&D capital is commonly used to approximate knowledge stocks, its relationship to growth has not been a focus of national accounts.

It seems conceptually sound to treat R&D as investment, in the sense that it generates an asset, knowledge capital, which can be drawn on in the future to realise benefits in the form of new products or improved processes that reduce production costs. Therefore, treating R&D as investment may provide a consistent accounting link between the expenditure and the corresponding asset. However, R&D is not a straightforward investment, as R&D entails risks and its economic returns are not assured. However, a number of issues need to be addressed to ensure credible estimates of R&D capital formation.

Preliminary analysis for some OECD countries suggests that R&D investment may account for substantial shares of productivity growth. For the United Kingdom, Edworthy and Wallis (2006) give an estimated elasticity of 0.095% for R&D capital, which implies that a 10% increase in R&D capital is associated with a productivity increase of 0.95%. In the United States, a recent study carried out by the Bureau of Economic Analysis and the NSF (2007) estimated that R&D capitalisation resulted in an average increase in GDP of 2.9% between 1959 and 2004, and that current dollar private domestic investment in 2004 would be 10.6% higher than the currently published estimate. The results are more modest for the Netherlands, as de Haan and van Rooijen-Horsten (2005) conclude that the effect of capitalisation of R&D adjusts total gross domestic product (GDP) upwards by 1.1% to 1.2%. Equally, economic growth, measured by the volume increase in GDP, is scarcely affected. Consequently, adjustments of net national income are also quite modest since upward adjustments of gross fixed capital formation (GFCF) are counterbalanced by negative adjustments from consumption of fixed capital. In principle, the capitalisation of R&D in the national accounts will also show the contribution of public investment in R&D to growth of GDP, to the extent that public investment leads to goods and services that can be sold in the market.

Impact assessment of research councils and public research organisations

Detailed assessments of the impact of public R&D, at the level of individual institutions and programmes, have typically been more successful at identifying impacts (Box 3.2). Sometimes the evaluation is built into routine policy processes (Box 3.3). Research councils and public research organisations can be differentiated according to their functions in the research system and the type of research they carry out. The national research council's (*e.g.* the Australian Research Council) mainly fund the research performed in a country, while public research organisations (*e.g.* Belgium's federal scientific institutes [EFS]) carry out research activities. Hybrids, which have aspects of a national research council and a public research organisation, both fund and perform research (*e.g.* the US National Institutes of Health [NIH]). Some focus on basic research while others are industry-oriented. For example, the Australian Research Council focuses on basic research, the NIH on health, and the EFS on space. The next section gives examples of funding and performing institutions engaged in general or sector-specific research, with or without an industry orientation.

Box 3.2. Evaluating PRIs: Insights from the Research Institutions and Human Resources (RIHR) project

Public research institutions (PRIs), broadly defined as higher education or government research institutes, make a vital contribution to national innovation systems and their performance. They contribute to the formation of a skilled scientific and technological workforce, extend the boundaries of knowledge, and act as an important source of knowledge transfer for the innovation activities of firms. Understanding the exact nature and size of their impact on society and the economy has become an increasingly central question for policy makers in recent years. As governments seek to be more rigorous in their decisions about how much and where to invest in R&D, they require more information about the contribution to growth and the social impacts of various programmes and institutions.

The RIHR project draws on evaluations of PRIs (excluding pure university institutes) to compare methodologies, highlight lessons learned regarding PRI policy and evaluation processes, and examine how evaluation results are used in practice.

In general, the goals of evaluation of PRIs are to better understand the scale, nature and determinants of the return to investment in these institutions and to learn about any unintended effects. This information can be used to improve steering and funding decisions, as lessons are drawn from successes and failures. Some key questions that are (or should be) asked in such evaluations relate to the efficiency and effectiveness of institutions and their programmes, the rationale for government intervention and whether the original drivers are still valid, and the level of additionality achieved through government funding.

The impacts can be widespread, and PRI evaluations have the potential to cover a broad range of less tangible issues, such as impacts on culture, societal views and organisations. These aspects are more difficult to measure, although advances have been made in some areas. For this reason, evaluation can only serve as a guide to policy makers. As it cannot provide definitive answers, value judgements will continue to play a role in steering and funding decisions.

The evaluations examined by RIHR addressed a variety of levels of PRI activity, ranging from the sector to individual projects. Impact or value added was the most common evaluation issue to be explored, although the exact meaning of these terms was not always clear. The second most common evaluation issue dealt with scientific outputs. The method most commonly used was qualitative assessment based on interviews and questionnaires to stakeholders. Some evaluations used indicators to inform their assessment of performance. These were generally "backward looking", in terms of summing up past performance, and few evaluations attempted to capture the potential future impacts of research in their assessments. Most of the evaluations judged that the sector, institute or programme/project in question had been of value or had performed adequately; consistent with the choice of methodologies; this was often based on an overall judgement as to the costs, benefits and influences of the initiative, rather than a quantitative assessment.

There is scope for improving the use of evaluations for policy-making purposes. There was little information on how the findings and recommendations of the evaluations were used, which raises broader questions about why evaluations may be overlooked in steering and funding decisions. A useful step would be to improve the relevance of evaluations to decision makers, by ensuring that evaluation methods and indicators keep pace with the changing environment in which PRIs operate, in particular by capturing the increasing numbers of stakeholders and level of cross-sector activity. Raising stakeholder confidence in, and acceptance of, evaluation activities is also important, and could be improved by consciously involving them earlier in the process. A number of countries have some degree of "built-in" evaluation in their policy processes which might be reviewed with these issues in mind.

Evaluations need to explicitly address the issue of an underlying rationale for intervention. Circumstances change and evaluations need to consider whether a market failure still exists, how strong it is, and to what extent government intervention is still appropriate. In addition, since intervention can gradually change the behaviour of stakeholders in negative as well as positive ways, it is important to review whether current approaches are still appropriate. Clearly, this is difficult, and the methodological challenges of creating a counterfactual are real. However, gauging the attitudes of firms and assessing their behaviour can provide some clues. This supports the use of qualitative information as well as quantitative methods of evaluation.

The evaluations highlighted some interesting issues and suggestions on how to improve the operation of PRIs and their programmes. While every country and innovation system is unique, and approaches cannot necessarily be replicated successfully in everywhere, it may be useful to note some of the issues and solutions that have been identified. For instance, many evaluations point to the difficulties PRIs have in meeting the expectations of all their stakeholders. This situation is becoming more acute as the environment becomes more complex and the numbers of stakeholders grow. Misalignment of stakeholder goals could be a reason why some initiatives fail or produce suboptimal results, and a better understanding of stakeholders' motivations and likely behaviour would be a crucial tool in better policy design.

The setting of research agendas, and the related issue of levels of core funding, continues to be a difficult balancing act. The degree to which PRIs undertake fundamental research, and receive core/capability funding to do so, must be set against the needs of other contributing stakeholders, who may have different time horizons and different priorities. One suggestion has been to focus programming on the overall portfolio of research, rather than individual projects. This would likely promote synergy and multidisciplinary work and allow for a longer-term research focus. Allocating core funding at a higher level has also been considered to help promote a strategic research approach and balance the influence of large stakeholders. A related challenge is to allow flexibility in terms of changes to research agendas and stakeholder investments, while maintaining long-term funding stability. A number of evaluations pointed to the dangers of "lock-in", without noting any clear solution.

An important issue highlighted by several evaluations was that research alone does not necessarily add value. Thinking about how results will be converted into further research advances or innovations must be an important part of the design of PRIs and their programmes. Dissemination strategies were observed to be inadequate in several cases and it appeared that specific activities were necessary to stimulate the application of results. The exact approach is likely to differ according to the goals of the institute – for instance, creating "value" through commercialisation may not be appropriate for PRIs with an explicit goal of serving industry. Ensuring that industry has the absorptive capacity to utilise research results is also crucial. Some institutes directly engaged with firms to demonstrate research results and build capacity for future knowledge transfer. Further analysis would be useful in this area. Moreover, although the path to economic impact and effects on innovation is particularly difficult to measure, this issue should not be ignored because of methodological difficulties. One way forward would be to take a longer-term view and include more in-depth analyses in evaluations.

There continues to be room for improvements in the methodology and design of evaluation procedures, including the incorporation of potential future streams of costs and benefits arising from current and completed research. Advances have been made in measuring social and environmental outcomes and further work will be valuable. However, evaluations are not costless and quantification of outcomes can only be undertaken to the point at which it remains cost-effective to do so. Regardless of the value of and need for evaluation, the scope and frequency of evaluations should be justified. Combining a variety of methodological approaches and including a range of stakeholders in the evaluation process can help to overcome the individual shortcomings of various approaches and may be a useful way forward while new approaches are being developed. Not all evaluations used clear and time-consistent indicators of performance, and the integration of field specificities was not transparent in most cases. As a general observation, clear goal setting at the start of initiatives will aid in collecting data and indicators that will later help in evaluation exercises.

Finally, evaluation is now taking place in a more complex environment and new demands are being placed on evaluation exercises. Ensuring that evaluation approaches can take account of overlapping roles and responsibilities of stakeholders, multidisciplinary efforts, globalisation and more complex funding arrangements will be essential if evaluation is to remain a useful tool for policy makers.

Source: OECD (2009), Research Institutions and Human Resources (RIHR) project: Strengthening the Impact of Public Research Institutions.

Box 3.3. Building evaluation into policy processes

A certain degree of evaluation and/or performance monitoring of programmes and institutions is sometimes built into routine policy processes. The following are some examples from public research institutions:

Belgium: The Flemish government concludes a multi-annual management agreement with all public research organisations, policy-related research centres, and special institutes. Every five years, an external partner, supported by an international panel of experts, evaluates the execution of the agreement and the linked results. Following an in-depth evaluation, new agreements were concluded with all PROs in 2006-07. These contained results-oriented criteria such as patents, spin-off companies and publications, in return for which the PROs receive a yearly financial grant.

Finland: The performance management model designed by the Ministry of Finance seeks to improve accountability of public officials, including in publicly funded research organisations. As part of the model, basic criteria for performance are defined and included in legislation: policy effectiveness (or societal impact); operational efficiency; outputs and quality management; and management of human resources. In response, research institutions have developed methods for assessing and monitoring organisational impacts.

Italy: Reforms over the last decade have addressed the framework for evaluation of research. In 1999, reforms aimed to develop a governing structure for the research system that included national research policy evaluation and assessment. As part of this, a Committee for the Evaluation of Research (CIVR) was established. In 2003, an integrated system of research quality assessment was created and in 2007 the National Agency for the Evaluation of Universities and Research was assigned the task of assessing the quality of the research results produced by institutions as well as the efficacy and efficiency of their institutional activities.

Japan: Independent administrative institutions, which perform a diverse range of R&D activities to meet policy challenges, are designed to establish a medium-term goal of three to five years. At the end of this period, the institutions are required to conduct an overall review of the organisation and its operations. This is assessed by an evaluation committee within the supervising ministry and by the minister in charge. The institutions are also required to submit a report of their annual performance/results to the committee. Many national testing and research institutions, another category of PRI, also undergo organisational reviews every few years, in accordance with the General Guidelines on Evaluating R&D.

New Zealand: Under the Crown Research Institutes (CRIs) Act 1992, CRIs should promote and facilitate the application of the results of their research and technological developments. To shed light on this and the broader overall performance of CRIs, three groups of indicators are used: research application metrics, related to the transfer of results (applied to all CRIs); relations/influencing role, related to how well CRIs are engaged in their sector (specific to each CRI); and measure of impact, related to the impact of selected research results or technologies that CRIs have applied or transferred over the previous five years (specific to each CRI). The Crown Company Monitoring and Advisory Unit (CCMAU) provides advice to shareholding ministers on CRI performance, and monitors performance against targets.

Norway: A new core funding system for the research institute sector, to be administered by the Research Council of Norway, was introduced as of 1 January 2009. This new scheme incorporates a tranche of performance-based basic funding (around 10%) based on institutes' production of scientific publications, co-operation with the higher education sector, income from the Research Council of Norway, income from abroad, and income from national research commissions.

Poland: Recent laws, particularly those enacted since 2001, have strengthened the importance of research/science/innovation policy evaluation and put more emphasis on the effectiveness of research, especially for socioeconomic development. The Act of Law on the Financing of Science, which describes the main mechanisms for steering the activities of public research institutions, specifically includes the activities of the Evaluation Committee of Research Units.

United Kingdom: Every five years, the parent Research Council reviews the Research Council Institutes, in terms both of their research portfolio and of their operational effectiveness.

Source: OECD (2009), Information provided by RIHR country delegates for the Research Institutions and Human Resources (RIHR) project: Strengthening the Impact of Public Research Institutions.

Impact assessment of research programmes

Research programmes are one of the main instruments used by OECD countries to implement research and innovation policies. They may aim at funding basic or more applied research in a general or a specific sectoral context, with or without a commercial objective. Two of the most important research programmes in terms of resources are the European Union (EU) Framework Programme (FP) and the United States Advanced Technology Program (ATP). The nature and scope of the research carried out under these two programmes are very different.

The EU 7th RTD Framework Programme

The EU Research and Technological Development (RTD) Framework Programme is the main multi-annual R&D funding programme at European level. Its aim is to help the EU meet its main goals. Since 1984, the FPs have played a leading role in multidisciplinary research and co-operative activities in Europe and beyond. The seventh Framework Programme (FP7) continues this task, and is both larger and more comprehensive than earlier ones (Box 3.4).

FP7 bundles all research-related EU initiatives together under a common umbrella and plays a crucial role in reaching the EU's goals of growth, competitiveness and employment. Running from 2007 to 2013, the programme has a budget of EUR 53.2 billion over its seven-year lifespan, the largest funding allocation yet. It funds both basic and applied research and seeks to improve the research capacities and results of all stakeholders (*i.e.* private companies, individual researchers, universities, public research institutions and foreign actors).

The European Commission has attempted to assess the wider impacts of the FPs on the economy and society. The most significant studies have used mathematical modelling to calculate impacts on the economy. For example, a study by the United Kingdom's Department of Trade and Industry[18] (DTI) analysed the impact on the United Kingdom's total factor productivity, using the model developed by Guellec and van Pottelsberghe de la Potterie (2001). According to this study, the estimated annual contribution to UK industrial output would be GBP 3 billion, a very large economic return on UK Framework activity. Similarly, a study by the European Commission's Joint Research Centre at Ispra, using the same methodology as the UK study, calculated the impacts of the FP on industry. The results seem to indicate that the effects are significant. For example, for Finland, the estimates suggest that 0.9% of annual industry value added is attributable to FP funding. Many member states record even higher contributions. On average, depending on the assumption and parameters of the study, it is estimated that EUR 1 of FP funding leads to a (long-term) increase in industry value added of between EUR 7 and EUR 14. The increase is spread over a number of years because of time lags before R&D spending produces its economic effects.

In addition, the FP7 has introduced an *ex ante* or prospective calculation of the impacts of expenditure. To do so, it uses a general equilibrium model called NEMESIS. This venture, while subject to further improvements, represents a qualitative jump in the *ex ante* impact assessment of research programmes and allows for estimating investment benefits before they occur.

In order to assess the impacts of the new FP, the European Commission drafted three scenarios:

- The "do-nothing option" serves to analyse whether without EU intervention it is possible to reach the same objectives.

- The "business as usual option" continues the previous FP, with the same budget allocations, objectives, instruments, priorities and institutional actors.

- The "enhanced Framework Programme option" doubles the resources of the previous FP and is designed to better respond to the Lisbon Agenda objectives.

For these scenarios, the NEMESIS model can calculate the different sets of benefits that would accrue. As with all econometric forecasts, of course, the results must be interpreted with caution because it is hard to establish a linear causal relationship between specific policies and particular effects. Moreover, it is very difficult to quantify many predominately qualitative effects, such as increased networking, improved absorptive capacity, strengthened research competencies of firms, or changes in behaviour. In addition to the economic gains, the FP could also have large social impacts.

18. Its name has recently changed to the Department of Business, Enterprise and Regulatory Reform (BERR).

Box 3.4. Priority Setting through the European Commission Framework Programmes

The 7th Framework Programme for Research and Technological Development (FP7) is the EU's main instrument for funding and steering research policy. FP7 differs from the previous Framework Programme in that it was explicitly designed to support the Lisbon strategy and as such is focused "on innovation and knowledge for growth" in the context of the European Research Area (ERA), the internal market for knowledge in Europe. The FP7 is organised into four specific programmes, corresponding to four major objectives of European research policy, namely *i)* co-operation in research and discovery; *ii)* ideas, which refers to the establishment of the European Research Council that will support individual research grants; *iii)* people, which refers to Marie Curie actions to improve human resources in S&T, and *iv)* capacities, which refers to research infrastructure, regional and international co-operation.

6th EC Framework Programme (2002-2006)	7th EC Framework Programme (2007-2013)
Budget: EUR 17 billion	**Budget: EUR 53.2 billion**
1. Life science 2. Information society and technologies 3. Nanotechnology, materials process 4. Aeronautics, space 5. Food quality and safety 6. Sustainable development, global changes, ecosystems 7. Citizens and society	1. Health (EUR 6 billion) 2. Food, agriculture and biotechnology (EUR 1.9 billion) 3. Information and communication technologies (EUR 9.1 billion) 4. Nanoproduction (EUR 3.5 billion) 5. Energy (EUR 2.3 billion) 6. Environment (including climate change) (EUR 1.8 billion) 7. Transport (including aeronautics) (EUR 4.1 billion) 8 Socioeconomic sciences and humanities (EUR 0.6 billion). 9. Security (EUR 1.4 billion) 10. Space (EUR 1.3 billion)

The thematic research programmes are subsumed in the "co-operation" pillar (EUR 32.3 billion), which refers to gaining leadership in key scientific and technology areas by supporting co-operation between universities, industry, research centres and public authorities across the EU and with the rest of the world. Trans-national co-operation will remain the main instrument for carrying out research activities. This programme consists of ten different thematic research areas as listed above. Despite the greater number of priorities in FP7, there is a great deal of continuity between the thematic focus of the two programmes. However, the FP7 represents a 63% increase in public spending compared to FP6.

Source: European Commission (2008).

The US also has institutions in place that perform impact assessments to measure a research programme's success and progress. The US Economic Assessment Office (EAO) tracks the progress of funded projects for several years after ATP funding ends, and identifies the direct and indirect benefits delivered by ATP award recipients. Direct benefits are achieved when projects accelerate technology development and commercialisation processes, leading to private returns and market spillovers. Indirect benefits are delivered through publications, conference presentations, patents and other means of dissemination of knowledge.

The EAO uses a variety of methods to “measure against mission” the results and impacts of the ATP’s investment. The methods range from early surveys used to generate immediate information to detailed case studies, statistical analyses, tracking of knowledge created and disseminated through patents and citation of patents, and informed judgements. While current evaluation of emerging technologies occurs, existing tools are modified, new tools are developed, and/or existing methods are combined in new ways. The changing environment in which evaluation takes place makes it difficult to use any one method to measure the results and impacts of R&D investment.

One of the EAO’s main methods, used on nearly 30 projects to date, is in-depth cost-benefit analysis. The case studies are based on interviews of funded companies, their customers and industry experts, and on other primary data collection activities, such as the Business Reporting System Survey (Box 3.5). In the case studies, different stakeholders estimate the benefits directly accruing from the ATP. Here, the time at which the analysis is carried out is important. In general, *ex post* measurement of results already achieved (*e.g.* commercialised technology, sales of innovative products, and reduction of costs due to process improvements) needs to be combined with *ex ante* prospective analysis of the potential commercial benefits of the project.

Box 3.5. The Business Reporting System Survey

In early 1994, ATP implemented the Business Reporting System (BRS), a comprehensive data collection tool for tracking the progress of its portfolio of projects and individual participants, from project baseline through closeout and into the post-ATP period, against business plans, projected economic goals and the ATP’s economic criteria.

The survey is designed to capture economic and organisational changes that are expected in the award recipient population if progress is made towards the expected goals. The themes and topics defined by the goals are reflected in multiple lines of questions that vary in a logical progression over the survey period. Baseline information is collected from the initial survey, and follow-up questions in each area are included at the appropriate anniversary, closeout or post-project survey. Several variants of the surveys are used for different types of organisations. For example, participating non-profit organisations or universities are given a slightly different survey from that given to companies to reflect their specific roles in a project and their different organisational structures.

Intended for immediate use in project management and ATP evaluation, the data are also expected to support analysis of R&D behaviour and outcomes beyond ATP in the longer run.

Source: ATP, 2005.

It is difficult to identify the appropriate time to conduct an impact study. Prospective studies of project outcomes, particularly if performed before technical risks and uncertainties have been overcome and business risks carefully considered, may not generate credible or useful estimates of programme impacts. This is true even if the studies meet high standards of economic modelling and rigour. Also, if prospective studies are undertaken very early, it is extremely difficult to estimate the probability distributions of long-term advanced technology project outcomes. Given the uncertainties about outcomes, some combination of retrospective and prospective analysis is appropriate as long as the analysis includes direct evidence of commercialised products or processes that incorporate the project-funded technology.

Sometimes, a project that achieves quantifiable economic benefits requires funding from multiple external sources. A conservative approach to assessing the impacts of ATP funding is to allocate benefits in some equitable way among funding sources. Identification and attribution of benefits requires the matching of programme-funded projects to direct project outcomes. This matching is completed by tracing product outcomes back from company products to their origin in an R&D project and forward from the ATP-funded projects through the product development stages.

These studies are consistent with the Office of Management and Budget Circular A-94 recommendations for the use of cost-benefit analysis in general and of cash flow analysis, of net present value (NPV), of cost-benefit ratios, and of internal rate of returns. These are key metrics of programme outcomes. Some studies employ other quantitative methodologies, such as hedonic index models.

The results of individual cost-benefit studies can be aggregated to see the impact (usually prospective estimates) across ATP. The net social benefits from about 40 ATP projects, for which ATP provided USD 2.2 billion and industry provided USD 2.1 billion, are estimated at USD 18 billion. As these projects were funded and studied at different times, however, the impacts computed in the different studies are not strictly comparable and their aggregation presents methodological problems.

Non-economic impacts

Beyond economic gain, the aim of a substantial share of public R&D is to increase the well-being of citizens. R&D can both positively and negatively affect the environment, health, social development and cohesion. Although some R&D causes societal harm, much of it produces benefits for society. Cozzens (2007) classifies these benefits into two broad categories: the "what" and the "how" benefits. The "what" benefits concern the overall status of individuals, such as health, education and environmental quality. The "how" benefits relate to the way people live their lives. Equity, democracy and community development are examples. Public research is conducted in a wide range of disciplines, such as health and environmental research, social science research, humanities, etc., that can increase the well-being of citizens.

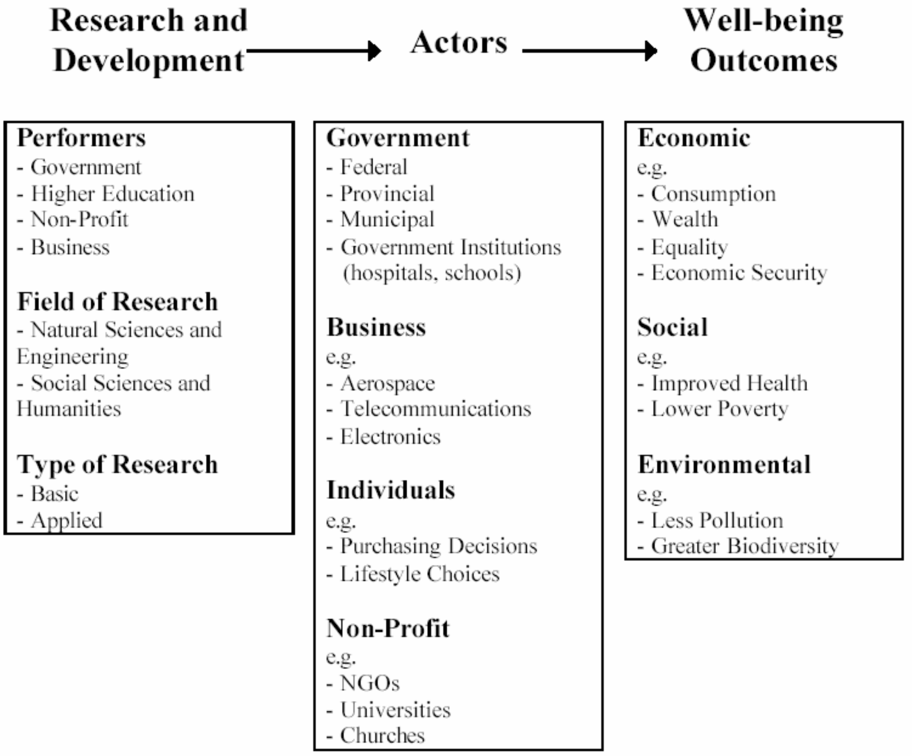
Unfortunately, the literature on the non-economic impacts of science is much less abundant and robust than studies of economic impacts. Godin and Doré (2006) identify three main reasons for the scarce production of non-economic impacts studies.[19] The first is that most measurement of science and research has been undertaken in an economic context. The second is that the economic dimension is often easier to measure than social impacts for the reasons discussed earlier. Third, most of the outputs and impacts of science are intangible, diffuse and often occur with important lags.

Nevertheless, in recent years, researchers and governments have started to be interested in the non-economic impacts of public R&D. There is some consensus among researchers that one of the first steps towards better understanding the non-economic impacts of public R&D is to define a framework that links research investment and well-being (Sharpe and Smith, 2005). Cozzens (2007) argues that social outcome indicators of research are neither difficult nor rare and that dozens of indicators relate to the public goals of research. In her view, what is lacking is not outcome indicators but the logic that connects them to research and innovation.

Sharpe and Smith (2005) develop a basic general framework for assessing the impact of research on well-being. This basic framework (Figure 3.1) links research investment with well-being via the uses made by social actors of the increased knowledge generated by research. In principle, this general framework can capture the impact of many different types of research investments used by different social actors to affect various dimensions of well-being.

19. Godin and Doré (2006) use the concept of science and technology, which is broader than public R&D. However, the problem of impact assessment is the same for both.

Figure 3.1. Framework for analysing the effects of research on well-being



Source: Sharpe, A. and J. Smith (2005), "Measuring the Impacts of Research on Well-being: A Survey of Indicators of Well-being" Centre for the Study of Living Standards- Centre d'étude des niveaux de vie", Ottawa, Canada.

This model requires adopting a four-step approach in order to measure impacts on well-being and establish their connection to public research:

1. Define the broad domains of well-being (social, economic, environmental, etc.) that are of particular interest, as well as sub-domains within the broad domains (*e.g.* within the social domain, the sub-domains child well-being, education, etc.).

2. Choose concrete indicators that can capture the domains or sub-domains.

3. Identify research investments that influence or determine the chosen indicators and specify the paths through which these investments and the knowledge created affect the indicators.

4. Quantify the impact of particular research investments on the indicators of interest.

The model then should be able to use a mix of indicators to track changes in the desired outcome area and should also make it possible to attribute the proportions of the changes to the research effort. The attribution of impacts is not easy, especially given the diverse factors affecting the final outcome and the time that may elapse between the public investment and perception of an impact. Such attributions, however, should be made possible by the use of expert judgements, the timing of change, or direct causal connections (Cozzens, 2007).

In health and environmental sciences, the development of metrics of social impacts is probably more advanced than in other fields, mainly because the causal relationship between investment and impact tends to be clearer as is the attribution of benefits**.** In other cases, however, as the Allen Consulting Group (2005) recognises, it is very difficult to express the primary social benefits[20] by using a common expression of value such as the social rate of return. In general, the most that can be done is to highlight where these impacts occur and articulate qualitatively the "value" of these impacts on society. To do this comprehensively, it would be necessary to "tell the story" of the impacts, and that is why the case study approach is often adopted.

As a result of the problems mentioned above, it remains necessary to improve the models that link public R&D with well-being in order to overcome some of the difficulties inherent in this type of analysis. In particular, these models should emphasise the need to specify the specific research investments and dimensions of well-being that are of interest before undertaking empirical work to estimate the impacts. These models should also deal with the problems of attributing the credit for impacts on well-being to public R&D, despite the difficulties involved. An alternative valuation model is discussed in Chapter 1 and includes some of the suggestions mentioned above. Further work, however, is still needed to overcome many of these difficulties and obtain better estimates.

8. Conclusions

This chapter stresses the importance of understanding and measuring the impacts of public R&D investments in order to evaluate the efficiency of public spending, assess its contribution to achieving social and economic objectives, and legitimise public intervention by enhancing public accountability. It presents some of the most promising and forward-looking practices adopted in this respect, including: general equilibrium models, econometric analyses, data linkages, scientometrics methods, survey-based indicators combined with econometric analyses, and case studies. These are a few of the analytical techniques that governments can use to assess the impacts of their spending on R&D. Other techniques, such as the use of experts (*e.g.* peer reviews), Delphi methods, technological foresight, sociological and socioeconomic, longitudinal, and historical methods are also options in the toolkit available for impact assessment.

The choice of methodology, or methodologies, must be made in the context of a specific research evaluation. An impact assessment exercise requires a deliberate selection of the dimensions that will shape the exercise. These are the timing (*e.g.* *ex ante*, monitoring, *ex post*), the object to be assessed (*e.g.* a research programme, public research organisation or a research system), and the specific nature of the research (*i.e.* whether it is basic science or technology development, and whether or not it is primarily industry-oriented).

When deciding which methodology to apply, it is also important to consider the scope of the impacts to be measured. Public R&D may have impacts at different levels of the economy or society and public R&D impact assessment exercises may focus on assessing the impacts of that investment on a specific sector or on the overall economy or society. As a result, no single analytical method can be used in all contexts. In fact, methodologies tend to be quite context-specific and specific factors determine their appropriateness in a given situation.

This review focuses mainly on top-down approaches, especially econometric and mathematical models, which are likely better suited to assess impacts affecting the whole research system and dealing with all types of research, both basic and applied. In particular, mathematical models, such as general equilibrium or similar models, may be a good way to assess systemic impacts *ex ante*.

20. The Allen Consulting Group's classification uses "the human, environmental and social dimension of benefits" as the equivalent of what is here called the social or non-economic benefits of public R&D.

On the other hand, when the subject of the assessment is a research programme and/or institution that aims at developing a specific type of technology with a clear industrial focus, bottom-up approaches are preferred. For large research programmes or institutions carrying out a wide range of research activities that are not particularly focused on specific technologies or industries, case study analyses that identify and quantify benefits and track them back to the original sources seem to be an option. Case studies describing the main benefits, together with a narrative about these benefits, seem to be the best option for assessing the non-economic impacts of public R&D at present. In general, these methods seem to work better for *ex post* assessment. In the case of *ex ante* impact assessments, uncertainty about the type and nature of the benefits that may accrue and the time required for them to appear make these methods less accurate. As yet there are few *ex ante* impact assessments dealing with the specific impacts of research programmes or institutions. Most *ex ante* studies have focused on assessing systemic macroeconomic effects deriving from the research investment. Accurate *ex ante* identification of specific benefits and potential users is still limited.

When assessing the impacts of public R&D, it is also important to distinguish between publicly funded and publicly performed R&D. The objectives and scope of the activities differ, which may explain differences in returns to public resources. Publicly funded but privately performed R&D may have a more targeted objective and achieve more immediate results. On the other hand, publicly performed R&D may focus on basic research that might otherwise not be carried out and may take a long time to produce visible impacts, which may be more difficult to attribute to the original research. Therefore, distinguishing between publicly funded and publicly performed R&D when evaluating the impact of investments may provide a better picture of the returns.

This chapter has also shown that the various methodologies are still evolving and based on a series of working assumptions that must be borne in mind when drawing conclusions. Because of the many types of public R&D undertaken and the many different dimensions of well-being affected by these activities, it is very difficult to develop a framework that captures all the possible impacts of public R&D. As a result, until now, none of the available techniques has been able to capture the full range of impacts of public R&D on society. The search for new techniques, however, has opened new and encouraging lines of investigation.

In practice, since socioeconomic impacts are complex and very different in nature, it is recommended to use a variety of methods to assess them. Where systematic and continuous assessments have been carried out using a range of methods, the coverage of impacts is better and the overall effectiveness and efficiency of the public investment can be better analysed.

Further work is needed on integrating different approaches and methodologies to create coherent impact assessment practices. More integrated frameworks using a combination of complementary methods should be explored. For now, no common framework for developing and using these analytical techniques has been agreed and international collaboration in this field is still scarce. The scope, nature and objectives of public R&D vary across OECD countries, as do national socioeconomic demands for public research. Therefore, it may be difficult, owing notably to data and methodological limitations, to achieve full international comparability and benchmarks. This should not imply, however, that countries should or can give up.

Finally, although methodologies for impact assessment remain a challenge, it is crucial to recognise that some important values of scientific research will remain hard to quantify. Investment in some areas of basic science primarily seeks to satisfy human curiosity and deepen our understanding of the universe. In some cases, such research may prove to have benefits beyond pure knowledge and the satisfaction of curiosity; in others, it may not. A related and perhaps larger challenge remains the fact that in many cases the results do not feed back into the policy debate.

REFERENCES

Allen Consulting Group (2005), “Measuring the Impacts of Publicly Funded Research”, commissioned by the Australian Government, Department of Education, Science and Training.

Bureau of Economic Analysis and National Science Foundation (2007), “2007 Research and Development Satellite Account”, mimeo.

Coe, D.T. and E. Helpman (1995), “International R&D Spillovers”, *European Economic Review,* 39(5), pp. 859-887.

Cozzens, S.E. (2007), “Maximizing Social Impacts through Science and Technology: Carson’s Quadrant”, Presentation at the “Impacts of Science”, 10-11 May, Vienna.

Edworthy, E. and G. Wallis (2006), “Research and Development as a Value Creating Asset”, UK Office for National Statistics (ONS), mimeo.

Godin, B. and C. Doré (2006), “Measuring the Impacts of Science: beyond the Economic Dimension”, Working Paper, mimeo.

Griliches, Z. (1979), “Issues in Assessing the Contribution of R&D to Productivity”, *Bell Journal of Economics,* 10(1), pp. 92-116.

Guellec, D. and B. van Pottelsberghe de la Potterie (2001), “R&D and Productivity Growth: Panel Data Analysis for 16 OECD Countries”, *Oxford Bulletin of Economics and Statistics*, *OECD Economic Studies*, pp. 103-125.

Haan, M. de and M. van Rooijen-Horsten (2005), “Measuring R&D Output and Capital Formation in Open Economies”, Presentation at the workshop on “The Measurement and Analysis of Innovation and Productivity Growth”, Groningen, November.

Khan, M. and K.B. Luintel (2006), “Sources of Knowledge and Productivity: How Robust is the Relationship?”, DSTI Working Papers 2006/6, OECD Paris.

Lichtenberg, F. and D. Siegel (1991), “The Impact of R&D Investment on Productivity: New Evidence using Linked R&D-LRD Data” *Economic Inquiry,* 29, pp. 203-228.

Mamuneas, T.P. and M.I. Nadiri (1996), “Public R&D Policies and Cost Behaviour of the US Manufacturing Industries”, *Journal of Public Economics* 63, pp. 57-81.

OECD (2003), *The Sources of Economic Growth in OECD Countries,* OECD, Paris.

OECD (2008), *Science, Technology and Industry Outlook,* OECD, Paris.

Sharpe, A. and J. Smith (2005), "Measuring the Impacts of Research on Well-being: A Survey of Indicators of Well-being" Centre for the Study of Living Standards/Centre d'étude des niveaux de vie", Ottawa, Canada.

Sveikauskas, L. (2007), "R&D and Productivity Growth: A Review of the Literature", BLS Working Papers 408.

CHAPTER 4
SETTING PRIORITIES IN SCIENCE AND TECHNOLOGY: THE ROLE OF EVALUATION

> Evaluations are used in many contexts. One is S&T priority setting. This chapter presents some of the main policy issues surrounding priority setting. It discusses the role of the different actors in the system and the different tools used to help S&T decision makers set priorities. It then seeks to identify best practices for improving both the quality of *ex ante* evaluations and their usefulness in the policy making process. It also assesses the process of priority setting in S&T itself and identifies structural weaknesses as well as best practice solutions.

Introduction

As research and innovation take on a more central role in economic development, the setting of priorities for public R&D and innovation has become a more complex and urgent challenge. Priority setting, however, is no easy task and requires not only political vision and clear societal goals, but also tools and mechanisms that can help governments set and implement priorities. Evaluation, in particular *ex ante* evaluation, is one such tool.

Priority setting and evaluation are two distinct issues with their own dimensions. Priority setting is the conscious selection of activities at the expense of others with an intended impact on resource allocation. This type of priority setting contrasts with the type that takes place in a self-organising system. Priority setting is concerned with questions such as: Shall we invest more in basic research or innovation? What technologies have greater private and social returns? In an institution, shall we invest in an Earth observation platform or a particle collider? Emphasis in the priority setting process varies over time. Historically, thematic priorities such as technology dominated, followed by mission-oriented priorities to respond to societal demands. More recently, "functional" priorities or those that affect the functioning of the system (*e.g.* Shall we focus on policy strands instead of technologies?) have drawn the attention of policy makers.

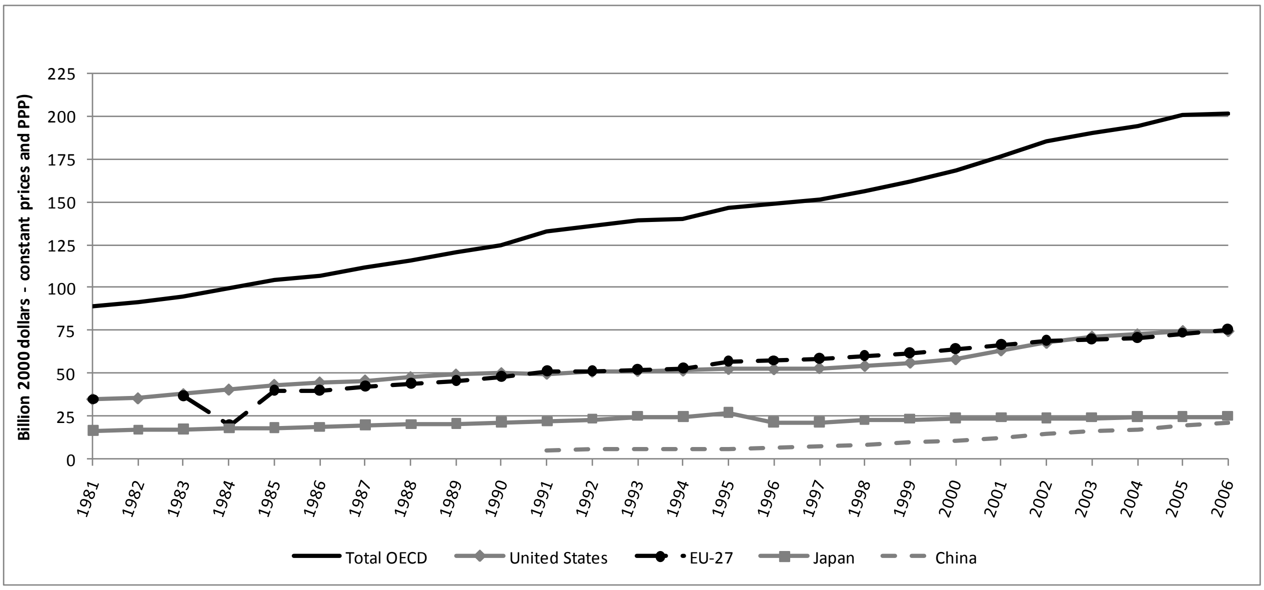
The rationale for priority setting

The rationale for setting priorities for public investment in S&T continues to evolve. While scientific excellence continues to be the dominant rationale, in particular for funding of basic research, the focus on accountability and the social and economic benefits of research has become more pronounced over the past decades. Indeed, over much of the post-World War II period, it was largely taken as given that science should be performed without thought about its practical use, in order to expand knowledge and understanding of the natural world. Science would, through the mechanism of technology transfer, drive technological innovation and contribute to economic and social welfare (Bush, 1945). Implicit in this understanding was the idea that researchers should set the priorities for basic research.

In practice, this situation involved an implicit social contract between scientists and society. Over the past decades, the terms of this contract have changed as technological progress and a range of reforms in the government of democratic societies has resulted in greater calls for accountability of publicly funded research. These demands focus not only on achieving payoffs from public investment in research, *e.g.* in terms of "national competitiveness", better health outcomes, etc., but also on accountability of science regarding the potential negative implications of certain scientific activities.

The pressure for greater accountability and economic benefit from R&D has increased alongside a rapid and absolute rise in public support to R&D.[21] Moreover, the combined amount of public R&D expenditures (*i.e*. the sum of higher education R&D and government R&D) in the OECD area rose from USD 89 billion in 1981 to USD 202 billion (in constant 2000 PPP USD) in 2006 (Figure 4.1). Given the large amounts of public expenditures there is demand for evidence on the outcomes and impacts of such investment and for more robust mechanisms for setting research priorities by governments or their delegated authorities. Today, many countries are struggling with the challenge of allocating resources to improve economic performance without restricting the freedom of individual scientists to set their own research directions.

Figure 4.1. Trends in total public R&D expenditures in the Triad and China, 1981-2006



Source: OECD Main Science and Technology Indicators, 2008.

3. Defining priority setting

Priority setting for S&T can be defined as "the selection of certain activities at the expense of others with an impact on the allocation of public resources" (Polt, 2007). In terms of R&D, it is the decision on who and what to fund and how much and for how long. Priority setting, however, is not a straightforward process: it is a democratic process involving political bargaining and compromise among the different actors in the system (*e.g.* researchers, institutions, funding agencies, ministries). However, there have been attempts since the 1980s and 1990s to improve the conceptual underpinnings of priority setting by greater use of technology foresight and technology assessment tools as well as greater stakeholder involvement.

187. Priority setting is highly context-specific and largely influenced, if not determined, by the institutional settings that govern S&T as well as by the technological specialisation of different countries. Barré (2008) suggests that the basis for setting "national priorities" is to establish a rationale and a

21. This is true even if the share of government R&D has tended to fall relative to business R&D and higher education R&D.

discourse about national S&T policy and to give it political visibility. A secondary aim is to identify and highlight policy measures to improve the functioning of the national innovation system (*e.g.* framework conditions for innovation policy). The final aim is to send signals about longer-term shifts in relative funding among broad sectors. In contrast to national priority setting, priority setting at the sectoral level allows governments to give special attention to a few politically significant issues/sectors (challenges) and to design an integrated set of actions to address them.

Polt (2007) categorises three main dimensions of the priority setting processes, namely:

- types of priorities: thematic priorities or functional/generic priorities;
- levels of priority setting: national priority setting exercises, institutional priority setting, etc.;
- nature of the priority setting process: top-down/expert-based *vs.* bottom-up/participatory, degree of formalisation, mechanisms for implementation, evaluation, etc.

The process of priority setting

Priority setting for research themes or areas can be divided into top-down and bottom-up approaches. The former include governmental priorities expressed by government ministries that reflect strategic priorities (*e.g.* economic development) or public missions (*e.g.* health). The latter essentially reflect the priorities of research producers: researchers themselves, research institutions and funding agents.

With the exception of mission-oriented research institutions, bottom-up approaches have dominated thematic priorities for longer-term and fundamental research. Moreover, these represent the core funding of public research in many countries. Peer review of research publications has been the main basis for assessing research to determine priorities. In practice, *ex post* and *ex ante* evaluations of research policies and instruments have had less impact on the priority setting process. Although research-performing institutions enjoy a high degree of autonomy and set priorities according to their own criteria, the priorities of the public research funding agencies are inevitably reflected in the priorities of the performing institutions. For example, programme or project funding is often tied to the funding agencies' priorities. National priorities for societal objectives, such as health and environment, and broader objectives, such as a "knowledge-based" or "information" society, can influence thematic research priorities at the operational level.

During the 1990s, many OECD countries began addressing gaps in their innovation system by making "functional" or structural issues a priority for policy action and additional funding. Functional priorities include increasing research funding, strengthening university research, promoting basic research, increasing women's participation, promoting sustainable development, or strengthening specific technology areas (ICT and biotechnology). By identifying and setting these functional priorities, policy makers aim to enhance the functioning of the national innovation system.

National priority setting can have many very different forms. At the macroeconomic level it can be expressed in government White Papers, national innovation strategies or national S&T plans. At the operational level, priorities can be expressed via the missions of institutions or through more flexible structures such as centres of excellence. More recently, governments have increasingly used instruments such as research and technology programmes, performance-based contracting and public-private partnerships as more flexible ways of influencing the research agenda of research institutions (which still have quite a high degree of autonomy in setting their research agenda) Moreover, funding instruments also serve to adjust or set national priorities. Industry financing of public research or public/private partnerships can also shift public priorities for research or align them with business strategies, over both the long and short term.

Table 4.1. Forms of priority setting

Strategic policy instruments	Institutional instruments	Funding instruments
Government White Papers Policies addressing specific industrial sectors, clusters (*e.g.* the technology platforms established by the EU) Government procurement for specific sectors/technologies Strategic research agendas of research teams	Targeted research and technology programmes Research institutions with specific profiles and technological orientation Centres of excellence	Budget plans and allocations Performance-based contracting for public research institutions Public-private partnerships Industry funding of public research

These (and other) means of priority setting have been used in varying degrees over time and in different innovation systems. For example, in countries in which R&D has a strong military focus, government procurement of R&D has typically played an important role. Countries with a strong focus on other large technological systems (*e.g.* nuclear energy) have also made significant use of government procurement. Also, the setting up of dedicated public research institutions (sometimes with a quite narrow technological focus) was an early trend in the definition and implementation of S&T policy priorities. These means are not confined to governments. Agencies and research institutions have also built up capacities to engage in priority setting and the implementation of priorities.

Actors in priority setting

Priority setting is a complex decision-making process involving not only the scientific community, but also stakeholders outside of science (patient groups, industry, agriculture, etc.). These stakeholders are the actors in research funding and performance who take part in setting priorities: sponsors of research, intermediary agencies, research-performing institutions and researchers. Depending on the institutional context, industry, the social partners and civil society directly or indirectly play an important role in the process. Industry and civil society often participate in formal consultative mechanisms such as advisory councils or university boards. Sometimes civil society actors such as private non-profit foundations (*e.g.* the Bill and Melinda Gates Foundation) influence government priorities by making their research priorities highly visible.[22]

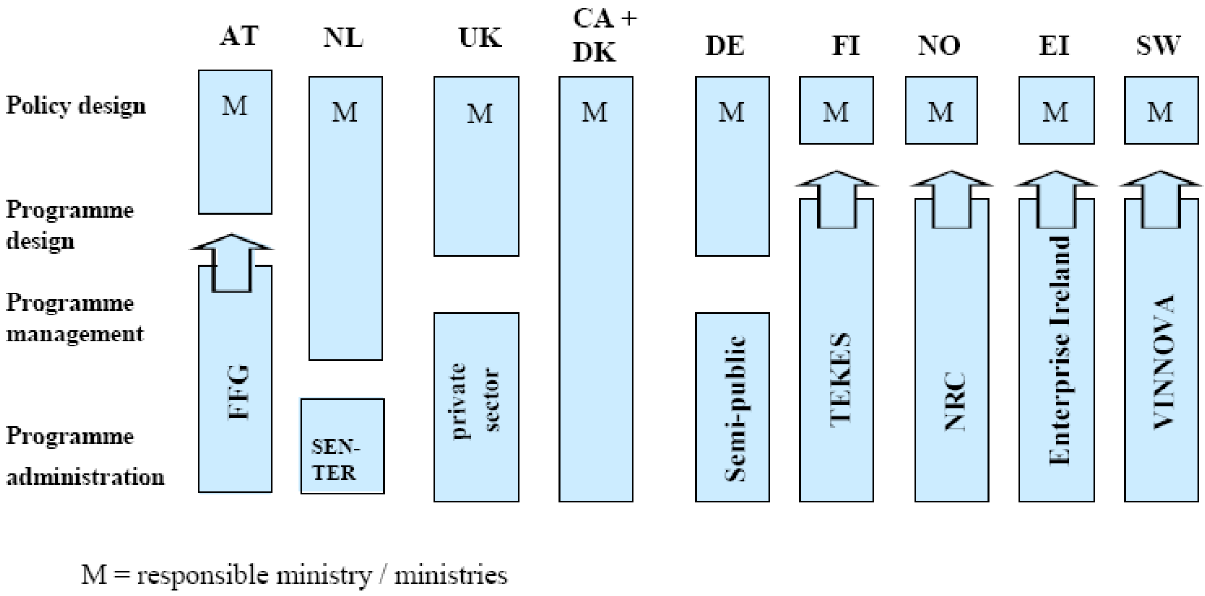
In most countries, priority setting is not centralised and does not take place primarily at the level of federal/central government. Rather, with a growing division of labour in policy systems, a number of actors are involved in STI policy making, ranging from agencies (research councils, funding agencies) to regional governments (Figure 4.2). Various approaches to setting priorities can be observed, but nearly all include stakeholder involvement of some kind, including consultations with enterprises or business associations, as well as non-governmental organisations (NGOs).

In a typical division of labour, government (quite often with the help of S&T policy councils) formulates broad policy orientations through budget frameworks and general goals in STI policy strategy documents, while individual ministries and funding agencies set more concrete priorities.

22. The role of such charities and foundations is growing, but remains poorly documented. the OECD's Main Science and Technology Indicators data for the OECD as a whole show that the share of gross expenditure on R&D funded by the private non-profit sector rose from 2.3% in 1981 to 2.6% in 2006. However, there is considerable cross-country variation in this indicator, with the US share being the highest, at 4.3% in 2006.

As a result of the trend towards devolution of competences to specialised agencies, they build up their competences in evaluation and in priority setting. Examples include: Finland's technology promotion agency (Tekes), Belgium's innovation promotion agency (IWT), and Austria's research promotion agency (FFG). Priority setting in such complex policy systems involves close interaction and bargaining between the principals (governments, ministries) and the agents (funding agencies, public research institutes).

Figure 4.2. Governance structure of technology policy in selected countries



Note: This figure does not necessarily represent the current state of division of labour between actors. It is shown here as an illustration of the different possible configurations between actors.

Source: Arnold (2004), Innovation Governance: Typologies and Principles, European Trend Chart Policy Workshop, European Commission. Brussels.

Institutional features and mechanisms for priority setting: some examples

Although priority setting is typically not centralised, some attempts have been made to centralise and especially to co-ordinate priority setting. This may be done through research funding decision-making mechanisms. Broadly speaking, there are countries in which the top-down approach dominates and others in which the bottom-up approach is more important. In all countries, both top-down and bottom-up forces exist, and some countries attempt to integrate the two. In many countries there seem to be increasing tensions and shifts in this balance, making priority setting a major policy issue.

In countries in which the top-down approach dominates, the central government adopts explicit strategies, policies, or plans that specify priority areas of research (*e.g.* Austria, Japan, Norway). Most of these countries, as well as some others (*e.g.* Netherlands, Denmark, Germany, Korea), have some kind of central advisory body that makes recommendations about priorities.

In Japan, the Council on S&T Policy (CSTP) is chaired by the Prime Minister and meets every month. It is responsible for preparing the Plural-annual Basic Plan. The CSTP assesses the compatibility of each ministerial research programme or initiative with the Basic Plan. The plan has mostly transversal objectives, but also includes large sectoral priorities. Japan also produces a White Paper more oriented towards societal issues. Ministries have a role similar to that of agencies. The CSTP also assures policy guidance and co-ordination.

In France, the situation is evolving as the country moves from a system of dominant public research organisations relative to universities, towards a more functionally driven system. Traditionally the steering and execution of research in France has been divided as follows:

- steering of research (government);
- programming of research (intermediate organisations, funding agencies);
- research performance (institutes, universities and their departments/units).

In practice, however, the "national S&T policy" has largely been the *ex post* sum of the sectorally defined policies and strategies of the PROs. A major implication of the 2006 Law on Research is the need for an explicit *ex ante* national S&T policy. This has arisen because of the need for performance indicators in the national budget, the demand for political transparency and rationale, and the fact that new players enter the steering and programming process (*i.e.* the *Agence Nationale pour la Recherche,* ANR).

In July 2008, France launched its "National Strategy for Research and Innovation" (NSRI). It is to be updated every four years. The NSRI serves as a "steering" document. It sets the stage for an overall vision and multi-annual perspective that is coherent with the European Union's research and innovation priorities. The demand for priority setting is really a demand to formulate a national policy that makes political sense by highlighting the societal challenges the nation addresses in setting priorities in the budgetary process.

Other countries use bottom-up, decentralised approaches to priority setting. In the United States and Canada, the government advisory bodies for research are decentralised and serve different government agencies. In countries in which no central advisory body exists, such as Sweden, priority setting is left to individual government ministries and agencies.

In the United States, federal priority setting for research occurs at three levels: *i)* in setting federal goals for research; *ii)* in the budget allocation processes for research within the White House and the Congress that in the aggregate produces the federal research portfolio; and *iii)* in federal agencies and departments focused on carrying out their missions in line with the Administration's research priorities. The agencies' advice on priorities comes from federal advisory committees, which are set up by different agencies that fund research. These committees make recommendations based on reports from the President's Committee on S&T, the National Academies of Sciences, the president's science advisor, workshops organised by the agencies, and advice from professional societies. The committee is frequently composed of various stakeholders, including industry. The US federal government-wide budget cycle allows agencies to co-ordinate their proposals and receive funding for identified priority areas.

In some countries, top-down and bottom-up approaches are explicitly integrated. Germany, for example, has a decentralised research system with autonomous public research institutions and universities. Priorities are set at the level of individual institutions following discussions between the government and the scientific community. Despite its decentralised structure, Germany has a Science Council, an independent advisory body consisting of representatives from the scientific community, government, business and civil society. This body plays an important part in making recommendations on priority areas and conducts evaluations of research institutions and programmes. National action plans and priorities are implemented mostly through the programmes managed by the federal ministries. A significant effort is made to co-ordinate ministries and instruments. The Research-Industry alliance (FWW), a committee of high-level industrialists and heads of PROs, also plays an important role.

In the United Kingdom, national S&T priorities are articulated through the spending review of the science budget and the Technology Strategy Board (TSB). Research councils ensure co-ordination. The total science budget addresses both research and innovation priorities, and the process allows for interactions and input from the bottom-up. A key trend is the emphasis on stakeholder input and expert advice. Bibliometric tools are also used to monitor evaluation outcomes, and socioeconomic effects are considered before the research is carried out. The challenge for the United Kingdom is to support decision making in strategic areas. Evaluations can play a role by providing input.

In many countries, research priorities are directly linked to annual funding decisions; this can create tensions between longer-term and shorter-term objectives. In some cases, funding decisions focus on increasing investment in new initiatives or priority areas. Some are inter-agency initiatives (*e.g.* the US National Nanotechnology Initiative), while others involve the creation of new structures, such as the Canadian Innovation Foundation which focuses on infrastructure and Genomics Canada.

Strategic policy intelligence and priority setting

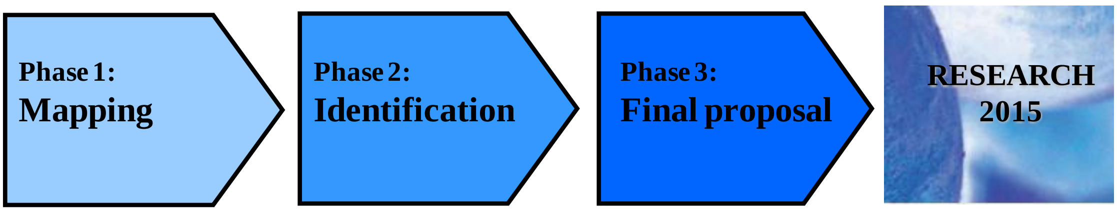
Priority setting has always mainly been a reflection of political priorities and the political bargaining process. Yet from the onset, policy makers have also sought the support of the studies and tools that are now called strategic policy intelligence (Boxes 4.1 and 4.2). Strategic policy intelligence can be defined as "the set of activities to search, process, diffuse and protect information in order to make it available to the right persons at the right time, so that they can make the right decisions". It is related to research and innovation policies and includes such policy support instruments such as foresight and technology assessment, monitoring, benchmarking, regional innovation auditing, technology road mapping, horizon scanning, specialisation indices, and strategic evaluation (Acheson, 2008).

Box 4.1. Prioritising strategic research in Denmark – RESEARCH 2015

Background: The Danish parliament (*Folketing*) decided that the foundation for setting priorities for strategic research – research within prioritised areas of society – should be improved. Strategic research can be both fundamental and applied but is essentially problem-oriented and interdisciplinary.

Focus/goal: Mandated by the parliament, the Danish Agency for Science, Technology and Innovation carried out the exercise and involved broad range of stakeholders.

Structure/approach: A mapping was carried out between March and October 2007 and involved an international literature scan by the OECD, broad consultations with societal stakeholders, as well as input from government ministries on strategic themes. Benchmarking tools were used to assess Danish research's capacity to pursue research priorities.



Result: 21 themes were identified as addressing important societal challenges for which research-based knowledge is important, broad enough to ensure competition among research institutions and concrete enough to form the basis for coherent research programmes. The final set of options will be used in political negotiations to set priorities for strategic research.

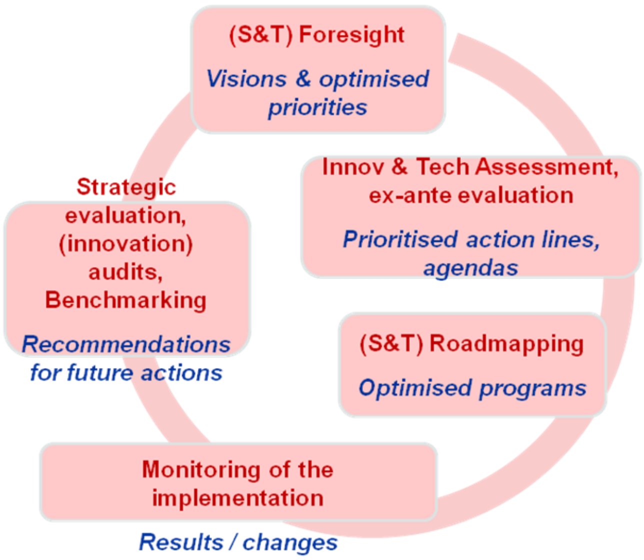
Implementation: the Danish Research Council for Strategic Research will directly implement the proposals. The RESEARCH2015 catalogue proposals do not aim to set priorities for allocating core funding (*i.e.* general university funds), but should serve to inspire the direction of research in universities. It is expected that the exercise will be repeated in four years.

Lessons: It is important to balance short- and long-term perspectives. Political drivers must also be weighted again challenge/opportunity drivers. In addition, there is a risk of reproducing existing priorities rather than identifying new themes. Better documentation and data are also needed to assess current and potential strengths. Finally, the process requires clear political commitment.

Source: OECD based on Hoff (2008).

Box 4.2. Strategic Policy Intelligence in Ireland

Like many countries, Ireland has turned to technology foresight exercises to help set priorities. Before the mid-1980s, spending for science, technology and innovation was derived from the wider policy objectives of each department and no clear system of relating STI expenditure to these policy objectives was discernable. The results of the first foresight exercise highlighted the significant investment needed for basic research in niche areas of information and communication technology (ICT) and biotechnology, two generic technologies underpinning strategically important sectors in the Irish economy. Since the first foresight study in the 1990s, the conceptual and empirical base for helping policy makers set priorities has evolved to include other tools such as technology assessment and technology road mapping.



Source: Acheson (2008).

Many governments use foresight processes, a type of "strategic policy intelligence", as part of their priority setting procedures or to stimulate dialogue. For example, Canada uses different types of foresight analysis adapted for various priority-setting needs. The United Kingdom has had a government-level foresight programme since the 1990s and government departments are obliged to take foresight into account when developing their science and innovation strategies. Japan has been conducting periodic technology forecasting exercises using the Delphi method since the 1970s. Korea also conducts foresight accounting and implicitly uses the results of experts who are involved in evaluation and pre-budget review to set national priorities.

Both *ex ante* and *ex post* evaluation is an element of strategic policy intelligence. Until recently, policy makers did not systematically use evaluations when setting priorities. This is partly due to the fact that past evaluations were better able to assess the rationale, implementation, and assessment of goal attainment of programmes/projects than to assess outcomes and impacts. Also, most evaluations have been *ex post*, while *ex ante* evaluations would lend themselves better to priority setting.

Foresight and evaluation are two major aspects of strategic policy intelligence. Others also deserve attention, such as technology road mapping, horizon scanning and specialisation indices.

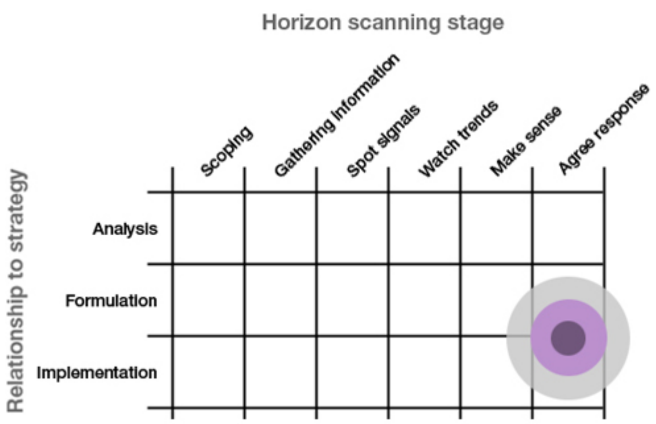
Technology road mapping

Prior to 1980, strategic economic and societal goals were included in the S&T policy agenda. After 1980, new tools appeared in the form of technology assessments and technology road mapping. Technology roadmaping is the process by which technologists try to determine the trajectory of a technology and the developments needed to maintain that trajectory. Most roadmaps have been organised around existing industries or technologies, such as semiconductors, optoelectronics, aluminium, pulp and paper, and electronic packaging. Some have included an assessment of future market needs, as well as technology developments. Many have been done by and for industry, but industries have often received government funding to support the road mapping process. Technology road mapping is especially important in complex technology areas, in which many components and subsystems have to be co-ordinated. Roadmaps identify the minimum performance needed for future technologies to be part of the system. They appear to be most useful in sectors in which established technologies evolve incrementally. However, they usually do not consider the social effects of the technology (Cheney, 2003). At the OECD, an important effort to develop a technology roadmap concerns energy (IEA, 2008).

Horizon scanning

In recent years, horizon-scanning exercises have become a mainstream activity as a follow-up to futures and foresight exercises. These activities were first developed in the United Kingdom (Figure 4.3) and then were emulated in countries such as Australia, Denmark and the Netherlands. Horizon scanning is a distinct futures methodology that researches and draws out key trends on the margins of current thinking that will affect people’s lives in the future. Most horizon scanning exercises aim to provide advance notice of significant new and emerging risks and opportunities, to exchange information, and to evaluate potential impacts. This involves the review of a broad spectrum of information beyond the usual timescales and sources and the participation of various sectors of society. Smaller economies have perhaps been the most active with regard to using foresight and other future-oriented studies to inform priority setting because of the need to focus and get returns from relatively small investments.

Figure 4.3. Horizon scanning for policy making



Source: UK Foresight Centre 2008.

Specialisation indices

Specialisation indices are a more recent development. They help policy makers take stock of a country's absolute and relative strengths and weaknesses in a given research or technological field. They are useful for assessing characteristic patterns of distribution in national and regional research and innovation systems. Different national specialisation patterns reflect R&D intensity, position in global R&D networks, etc. Ideally these indices can help identify priorities that can increase critical mass and create areas of comparative advantage, but their interpretation must take into account factors such as political decisions (*e.g.* on nuclear energy) or public choices.

Overall, the circle of people involved in these forward-looking exercises has expanded considerably. This focus now is on involving not just experts in the respective technology field but also the broader public. Yet, there has also been a clear trend towards more "expertise-supported consultation mechanisms" for priority setting.

Data requirements for priority setting

Improving the empirical basis of priority setting, especially for budgetary choices, depends on timely quantitative and qualitative data of high quality. On the input side, this implies knowing the amount of R&D funded by public budgets. While countries report aggregate data, there are gaps in the understanding of socioeconomic objectives of public R&D in scientific and technological fields. The multidisciplinary nature of research complicates this further. For instance, mathematical research can advance research and innovation in nanotechnologies or life sciences but may not be accounted for as such.

On the output side, evidence of the outcomes and impacts of public support to R&D in various fields is important to determine whether to invest more in research area "A" instead of research area "B". Over the years, countries have sought to improve the measurement of scientific outputs such as scientific publications, graduates or patents in order to identify areas of national strengths and weaknesses. Using such benchmarking to make funding decisions, however, has inherent limitations. For instance, it is extremely difficult to determine with any accuracy where the high impact and high return to investments lie (see OECD, 2008, Chapter 4). Improving data and analysis on both the input and output side will necessarily require work to develop up-to-date definitions and taxonomies, as well as greater co-operation among agents (funders and performers) in collecting data. The report of the OECD Blue Sky II conference discusses this issue in more detail (OECD, 2007).

International dimensions of priority setting

There are also important international dimensions to national priority setting. Foreign priority setting has both direct and indirect impacts on national priority-setting exercises. The most direct impact is competition among scientists to discover and create new knowledge. Strength in knowledge production (specialisation) in one country may influence the direction of specialisation in another country with different financial and intellectual resource endowments. Foreign funding for research via multinational firms or public research organisations may also have an indirect impact on the direction of research in the receiving country by signalling user demand in a given area. EU Structural Funds and Regional Funds have arguably shaped the direction of research priorities in the new member states.

There is, however, a risk that such external priorities may shift internal and national priority setting processes and resources towards areas of knowledge production that are of less concern to "national" stakeholders but that are important in a globally linked research community. Regulations in one country that limit certain areas of research (*e.g.* stem cells) can also create incentives and opportunities for boosting research in that area in a country with different regulations. While science has always been international, globalisation and the Internet have accelerated the movement and exchange of ideas and people. Priority-setting exercises, like the evaluation of research itself, increasingly involve taking into account global trends as well as direct inputs from foreign scientists and experts.

Another international dimension of the priority setting process concerns priorities for regional or global challenges or for large research infrastructures which require international co-operation to shoulder the high development costs. With the creation of the European Research Area, there has been an increasing focus on ways to better co-ordinate the national research programmes of the member states. This ranges from the definition of agreed common priorities (common vision), as set out in the Lisbon Strategy and European Framework programmes, to the implementation and common monitoring or evaluation of national and common programmes. The challenge in setting priorities for international collaboration is to balance a global, co-ordinated approach with a differentiated approach in relation to the different types of research. Furthermore, international priority setting must take into account the different technological specialisations of partnering countries, the need to foster both competition and co-operation among research teams, and the need to balance bottom-up research initiatives with top-down strategic guidance.

Feedback of evaluation and into policy design

Against the background of increasing support for research and innovation as well as increasing demands for accountability and socioeconomic impacts, the question of how *ex ante* and *ex post* evaluations relate to the selection of research and innovation priorities has become important for policy makers.

Use of evaluation for priority setting

To what extent can evaluations improve priority setting? Evaluation is an instrument that can be used for priority setting and help improve its quality. Evaluation can be *ex ante* and look at the potential impacts up front or during the life of a project/programme or *ex post* and look at impact and attainment of goals. However, evaluation has not been used for priority setting. The main challenge is to compare alternatives. Most evaluations have focused on single policy measures. Today, they address appropriateness, quality, efficiency of implementation, and assessment of additionality. There has been some progress in assessing the quality and efficiency of interventions. In terms of appropriateness there has been some progress using *ex ante* assessments (*e.g.* programme logic, rationale). Also there has been some progress on impact and behavioural additionality, but less on output additionality.

Of course, to prioritise, it is necessary to know the potential impacts of different measures (*e.g.* the net present value of alternative investments). Meso- and micro-level estimates have been made of the impacts on different measures, such as productivity, on the basis of the relation between R&D and impact at various levels. Progress has been slower on the meso level owing to time lag effects. In short, current methodologies for measuring the social rate of return of individual projects or subsidies to specific technologies are unsatisfactory. For example, studies of the US ATP programme find that on average the return is positive, but the range is so large that the conclusions cannot be used to select specific investments. While there are limits to the use of *ex ante* evaluation in priority setting, it can give a good idea about the rationale of programmes. Evaluation may also be more appropriate for legitimising policy interventions than for setting priorities because there are few examples of studies of policy measures in context (*e.g.* compared to other measures).

Gap between evaluation methodology and practice

Because most evaluations do not use quantitative tools, comparisons are difficult. Policy makers should have realistic expectations as to what evaluations can and cannot do. First, the information requirements of evaluations with respect to behavioural additionality far exceed what is available. As a way forward, policy makers should:

- Push the envelope on evaluation. For example, some researchers used instruments widely used in economics (*e.g.* option value approaches or microeconomic modelling using CIS data). The most promising avenue would be evaluation in context in a systemic perspective, for example of subsidies in relation to R&D tax credits to see whether they are complements or substitutes.

- Another option is to evaluate the priority setting process itself. It would be worthwhile expanding qualitative analysis to see which of the priority setting processes would be good practice for other countries. The benchmark would be a priority setting process that actually has an impact on the direction of technology and on specialisation.

Ireland demonstrates possible gaps between evaluation methodology and practice. In Ireland many actors are involved in priority setting. Nine ministries deal with S&T; they have different budgets and their own agencies. The number of actors, budgets and agencies leads to gaps between methodology and practice.

Ireland's top-down and bottom-up technology foresight exercise led to establishment of Science Foundation Ireland, but served mainly as a technology intelligence tool. In 2005, the Science Foundation Ireland's foresight exercise once again served a similar purpose. The Foundation is seen as a key pillar of Ireland's national innovation system and funding increased to EUR 1.4 billion in the latest seven-year national plan. Policy makers are, however, asking questions about the return on investment, especially the funding for basic research. Moreover, because evaluations of business R&D support measures were found unsatisfactory, Forfas, Ireland's national policy advisory body, is attempting to establish a framework for understanding and evaluating policy goals and performance (*e.g.* inputs, outcomes, outputs, etc.). It has also put together a schedule for a ministry-agency consultation. Technology Ireland has agreed to the schedule and its implementation in the agencies. A key lesson is that evaluations require "buy-in" so that the right indicators are collected to inform future policy. Actors need to be educated about evaluation and territorial issues need to be resolved: 9 out of 15 ministries and 40-50 agencies protect their S&T activities. On the positive side, there has been at least more verbal support for evaluation at the administration level, led by the department of finance, which is more focused on getting "more bang for the buck".

One of the challenges of thinking about evaluation in a "systemic" perspective is the need for better definitions of innovation systems. This also requires the establishment of relations between different policy mixes and policy outcomes. With more modest targets, it might be possible to make modest steps towards evaluations that consider this context.

One contentious issue is that priority setting is often seen from an economic perspective and assumes the generation of social returns. However, priorities are often set with goals such as strengthening existing capabilities. Sweden and Finland, for example, have focused priority research on topics that reflect resource-based or industrial comparative advantage (*e.g.* wood and fisheries). An important question is whether priorities should reflect socioeconomic needs such as ageing or global warming or whether priorities should first be explicitly formulated.

Another issue is the relation between changing priorities and spending. Priority accorded to one field often has consequences for spending in other fields; for example, greater focus on nanotechnology may result in greater attention to a certain area of mathematics. In practice, priority setting reflects what is considered important. This may be relevance, political choices or a perceived information gap. Evaluations can provide lessons in this respect because it is possible to analyse a country's national innovation system to determine gaps, such as poor industry-science relations or low rates of new firm creation. Policy analysis can contribute to some extent, but cannot help for making portfolio investment decisions. No form of *ex ante* evaluation can reach an investment decision of the type reached by investment bankers, policy makers' wishes to the contrary. For instance, UK science councils are increasingly asking about the economic return to basic research. However, it is unclear how priority setting can be approached from this perspective and some experts suggest that priority setting should not be viewed in terms of a tool to determine economic returns on investment.

Evaluation and feedback into policy design

The relation between evaluation and the development of policy measures is perhaps the most difficult challenge facing policy makers (Box 4.3). In this area, policy makers rely much more on expert opinion than on the results of quantitative econometric models. Although S&T policies have not been developed in the same way as policies for which *ex ante* policy evaluation is the norm, they also require a systems approach. *Ex ante* evaluation of policy options are most often based on qualitative measures, which aim to determine which programmes provide the greatest benefit to stakeholders. Expert opinion and review are also used in *ex ante* evaluation. Foresight and analysis of past experiences are useful for developing policy alternatives.

Among quantitative models, "logic models" are increasingly used for evaluations by academics and experts, but these are little used by policy makers. Other quantitative methodologies include cost-benefit, totally quantitative or qualitative tools, or mixture of the two. Often these are not completely carried out owing to time constraints and their inability to project the long-term benefits. By the time analysts determine the impacts, policy makers have moved on to new policy measures. The cost of implementation is also a consideration when a policy decision requires a new administrative body such as an agency. For example, in Canada two new agencies were created to address specific policy concerns (*e.g.* the Canadian Innovation Foundation and Genomics Canada) (Box 4.3).

Box 4.3. Feedback of evaluation in setting national priorities in Canada

The development of Canada's 2007 Science and Technology Strategy, which drew on expert opinion and foresight exercises, provides an interesting case study to illustrate some of the challenges in linking evaluations to policy making. The Council for S&T identified key areas in which Canada had strengths and this was reflected in the priorities of the final strategy. A number of surveys were conducted to determine the efficiency of current programmes and of programme delivery mechanisms in view of the fragmentation of access to research dollars (*e.g.* applicants having to apply for research funding at one council and then to the Canadian Innovation Foundation for infrastructure funding). In addition, an analysis of best practices in programme elements was carried out. The Council for S&T used macro-level indicators such as OECD S&T indicators to compare Canada to other OECD countries. There was also extensive use of key stakeholder roundtables, involving industry or the scientific community as well as calls for written submissions. Programme evaluation reports did influence some of the programme elements attached to the 2007 strategy. There was, however, no direct influence of evaluation on policy making. Attempts to evaluate previous strategies (*i.e.* the 1996 and 2002 strategies) failed due to lack of "buy-in" to collect data, although there is some hope that it can be done for the 2007 strategy. In conjunction with the focus on evaluation, there will be a greater focus on *ex post* programme evaluation, but there will be feed-in to improve policy decisions. *Ex ante* evaluation remains weak however. The Council administration will strictly enforce new rules to require cost- benefit analysis of all decisions in light of performance measure frameworks of regulatory decisions.

Thus, although there has been progress in developing conceptual and theoretical frameworks for providing analytical input to policy actions, in practice the take up has been slow. The interest of the international community and the OECD in particular, is essential. If this holds, it is hoped that the next national S&T strategy will have a greater focus on *ex ante* evaluations of policy options.

Source: Newell (2007).

In practice, and despite the emergence of new tools, policy analysis lacks robust measures. While qualitative analysis is important and needs to be undertaken, there is a need for more robust measures in order to evaluate a "portfolio" of measures. When selecting the goals against which to measure policy alternatives, different quantitative and qualitative tools are needed to estimate the predicted impact of the alternatives. For multi-goal analysis it is important to determine the policy criteria against which a programme or policy will be measured. The set of criteria to be used usually depends on policy goals and the relative weight of a criterion should be determined by the extent to which it will influence a policy outcome. This requires the combination of different elements some of which will entail more risks (*e.g.* cost, public perception or ability to implement) than others. This makes it easier to define policy options to central agencies.

What is the value of *ex ante* evaluation besides determining differences in policy options? First, it formalises the policy rationale and provides a basis for *ex post* evaluations of the programme. When *ex ante* and *ex post* evaluations are combined, it becomes possible to combine policy evaluation at the macro level with programme evaluation. What is often missing is a rationale for choosing a particular programme in relation to a defined policy goal. The issue is dual: evaluation and the engineering of policy making. The coherence between policy tools, values and objectives, and the involvement of actors are important. A look at *ex ante* evaluation of policy tools shows that the logic is based on resource intensity. There is also a question of the coherence of the policy mix of instruments used to address a given target group. There are intrinsic weaknesses in national strategies given the complexity of national innovation systems, as well as the new modes of knowledge production and new modes of governance. There is also a need for balance between new public management tools and national innovation systems. The main challenge for *ex ante* evaluation is to be developed as part of policy making. *Ex ante* evaluation requires indicators and data that can reduce information asymmetries and help legitimise policy interventions.

Good practices in the *ex ante* evaluation of policy measures are those that: *i)* identify the policy context in which the new policy is to be created; *ii)* analyse the new balance of the policy mix that would emerge for each policy alternative; and *iii)* make explicit the assumptions under which the new balance of the policy mix would be justified.

Conclusion

The key findings of TIP work on evaluation and priority setting are:

- Although priority setting and evaluation interact in policy making, they remain distinct dimensions of policy making with their own characteristics and internal processes.

- In line with the greater number of actors in S&T policy making (*e.g.* regional governments, separate funding agencies, etc.) and a greater variety of approaches and methodologies, there is greater need for mechanisms to ensure coherence between policy making and priority setting.

- Despite the emergence of new quantitative tools for evaluation, the conceptual underpinnings of priority setting remain quite weak and expert opinion continues to predominate in the evaluations used by policy makers to make policy decisions.

- Improving the process of priority setting through the use of *ex ante* evaluations will require political buy-in from the various stakeholders and commitment to invest in resources and skills, including the creation of indicators to monitor policy effectiveness.

- The process of priority setting in S&T could itself be the subject of evaluation in order to identify structural weaknesses as well as best practices.

- The interest of the international community and of the OECD in particular, is essential to improve the ability of countries to develop and use *ex ante* evaluation in policy making and priority setting in general.

REFERENCES

Acheson, H. (2008), Presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting.

Arnold, E. (2004), "Innovation Governance: Typologies and Principles", European Trend Chart Policy Workshop, European Commission. Brussels.

Barré, R. (2008), Presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting.

Bush, V. (1945), *Science, the Endless Frontier: A Report to the President,*, Public Affairs Press, Washington, DC, www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

Cheney, D.W. (2003), "The Emergence of S&T Priorities in the United States", Presentation at the Second International Conference on Technology Foresight, SRI International United States, Tokyo, 27-28 February.

Hoff, A. (2008), Presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting.

IEA (2008), *Energy Technology Perspectives*, IEA, Paris.

OECD (2007), *Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs*, OECD, Paris.

Polt, W. (2007), "Issues in Evaluation and Priority Setting", Room Document for the TIP/RIHR Workshop on Evaluation and Priority Setting.

PART II

THE ARCHITECTURE OF EVALUATION SYSTEMS

CHAPTER 5: CHINA: EVALUATION OF GOVERNMENT-FUNDED NATIONAL R&D PROGRAMMES[23]

The case study analysis begins with China. Although the country lacked formal laws and regulations pertaining to evaluation until 2000, the country states that R&D evaluation has become a high priority item. This case study offers a historical description of China's R&D and evaluation evolution, followed by a description of the methodology used to conduct this case study. It then describes and analyses the institutional framework and methodology for the evaluation of national programmes. Finally, it presents key challenges to the system.

The need for an effective government-funded R&D evaluation system has become a high priority for the Chinese government. It is also attracting growing public attention. Since evaluation is a relatively new concept in China, the institutional framework for R&D evaluation has not yet been established and the international evaluation community knows little about R&D evaluation in China. As one of the volunteer countries, China participated in this study as an effort to map the institutional frameworks, actors, regulations and practices of public R&D evaluation in different countries. This section presents[24] the findings of the Chinese case study.

1. China's evolving public R&D system

As a basis for understanding the R&D evaluation system in China, this section describes the key milestones in China's R&D system in the last 30 years and the major funding organisations at the central level.

Key milestones in the history of China's R&D system

China's R&D system[25] was arguably created in the 1950s following the Soviet Union model. However, within the scope of this study, it is more important to focus on its evolution from the late 1970s.[26] 1978 was a milestone year in the history of China's R&D system. The National Conference on Science and Technology, which was held in 1978, was the starting point of a major shift in China's R&D system. Since then, the Chinese government has continually taken steps to move from a planned to a market-oriented economy.

---

23. Prepared by CHEN Zhaoying, HAN Jun and SHI Xiaoyong, National Center for Science and Technology Evaluation of China.

24. The opinions, findings and observations are those of the authors and do not necessarily reflect the views of the National Center for Science and Technology Evaluation (NCSTE) of China.

25. The discussion only covers civilian R&D.

26. Most studies of China's R&D system take their point of departure in the late 1970s.

From the mid-1980s to the early 2000s, national government-funded R&D programmes were included in the sixth, seventh, ninth and tenth five-year plans. These programmes addressed the priorities in each five-year plan period and were also the Chinese government's policy tools to promote innovation.

The year 2006 can be regarded as another milestone in the history of China's R&D system. At the beginning of 2006, China initiated the National Medium- and Long-term Science and Technology Development Plan (2006-2020) (the MLP). According to the MLP, China will invest 2.5% of GDP in R&D by 2020, up from 1.3% in 2005, raise the contributions of technological advances to economic growth to more than 60%, and reduce its dependence on imported technology from 50% to less than 30%.[27] The MLP also calls for China to become one of the world's top five countries in terms of number of invention patents granted to Chinese citizens, and for Chinese-authored scientific papers to become among the most cited in the world. For many observers inside and outside China, the MLP can be viewed as an important effort to shift China's current growth model to a more sustainable one and to build an innovation-based economy by fostering indigenous innovation capability.

Governance at the central level

The State Council Steering Group for Science, Technology and Education is a top-level co-ordination mechanism for dealing with strategic innovation issues. The Ministry of Science and Technology (MOST), the National Natural Science Foundation of China (NNSFC), and the Chinese Academy of Sciences (CAS) are the main organisations that administer civilian R&D in China (Figure 5.1).

The main missions of MOST, under the direction of the State Council, are the formulation of innovation strategy and policies, promotion of the national innovation system, identification of R&D priorities, and the design and implementation of government-funded R&D programmes at national level.

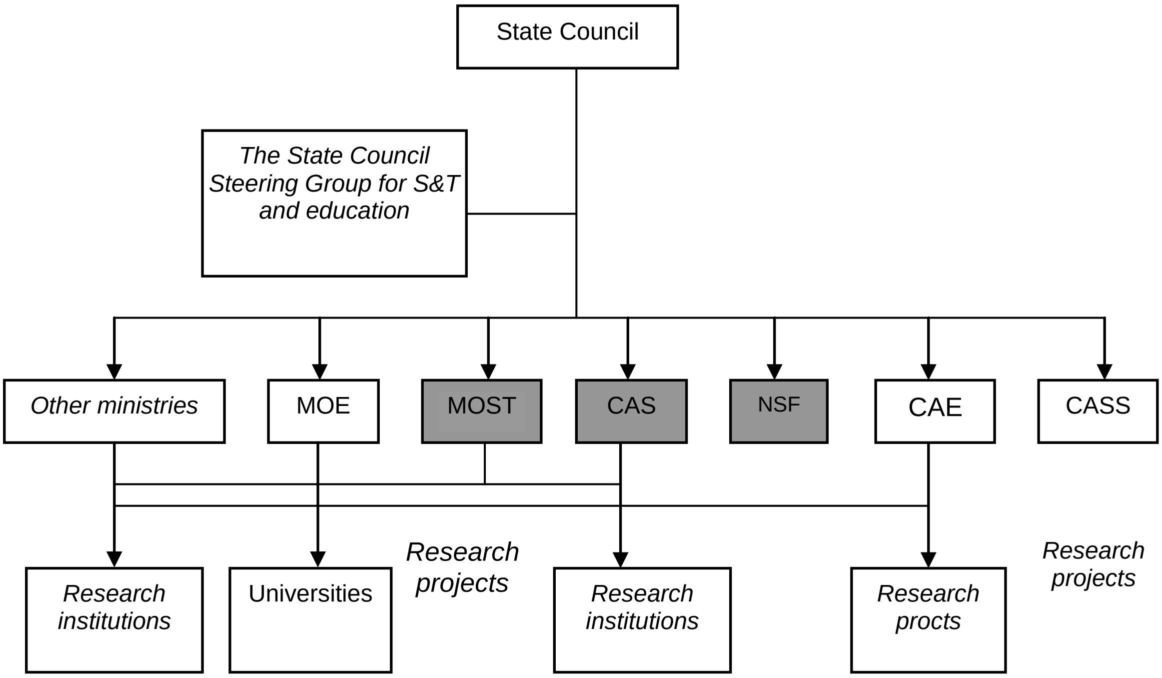
The NNSFC aims to promote and finance basic research in China. It mainly funds research in the sciences, such as physics, chemistry and life sciences. Funds are allocated on the basis of proposals, which are subject to peer review. The principal recipients are Chinese universities and CAS research institutes.

CAS is essentially an R&D complex composed of about 120 institutes located around China. Major R&D funding for CAS comes from a line item in the government budget, projects supported by the national R&D programme, and some funding from the NNSFC.

A number of line ministries such as the Ministry of Agriculture (MOA), the Ministry of Education (MOE), the Ministry of Health (MOH) and the State Forestry Administration (SFA) also have R&D operations under their direct management.

27. These goals are part of the MLP documents. However the authors think that the measurement of these indicators, the policy implications and the utilisation made of them require further study.

Figure 5.1. Governance of public R&D at the central level



2. The Chinese case study

This section briefly describes the case study, including background on the evaluation of government-funded R&D in China, the reasons for the focus on the evaluation of the national R&D programme, and the methods used.

The development of and demand for evaluation of government-funded R&D in China

Evaluation of government-funded R&D in China can be said to have started in the 1990s when MOST carried out a programme evaluation.[28] Largely through MOST's initiative, some regulations pertaining to evaluation were released and the importance of evaluation was increasingly recognised. MOST and a few other ministries then established internal bodies or specific staff with responsibility for managing evaluation. Since the early 1990s, MOST has carried out evaluations of several national R&D programmes. A few line ministries have also carried out evaluation activities, primarily at the project level. Since 2005, China's leaders have called for government departments to be accountable for the results of public expenditure.[29] Since the call for more accountability, new evaluation requirements of government-funded R&D are currently being established.

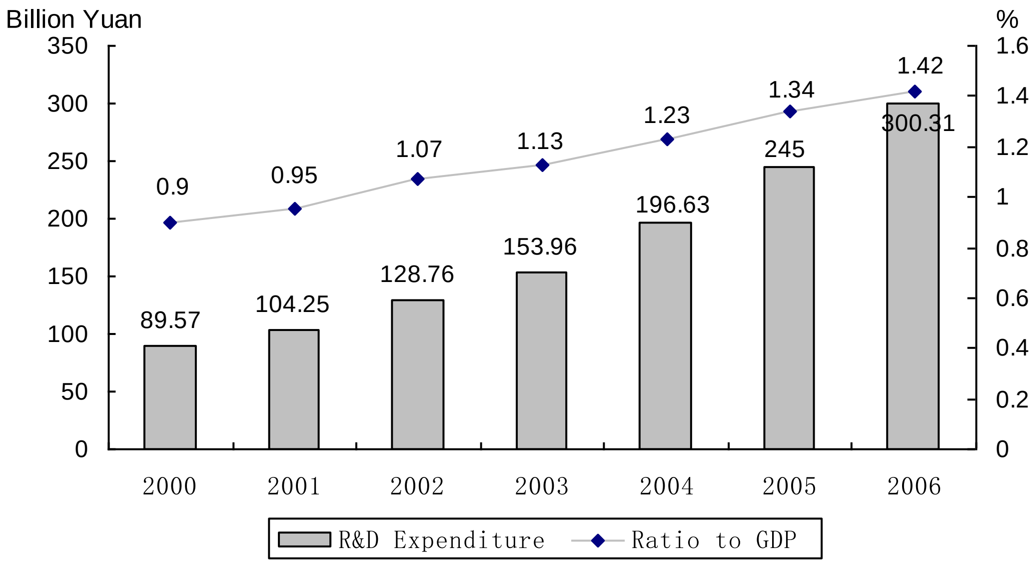
---

28. Some activities, such as policy analysis, management studies, surveys and programme reviews, were sometimes broadly termed "evaluation", but they differ a great deal from evaluation as understood by the international community in terms of design, implementation and the report presented.

29. Premier Wen Jiabao and the President Hu Jintao have called many times for government departments to be accountable for the results of public expenditure. The country's leaders require a government performance

The significant expansion of government funding of R&D has drawn more public attention to performance. Since 1999, China's spending on R&D has increased more than 20% annually. In 2006, it reached RMB 300.3 billion and 1.42% of gross domestic product (GDP), of which RMB 71.6 billion from the central government. Figure 5.2 shows the dramatic increase in R&D expenditure in China. This has raised concerns about the performance of R&D funding, and the government has come under pressure to establish an effective evaluation system for public R&D.

Figure 5.2. China's R&D expenditure and intensity 2000-06



Source: The Department of Development Planning, MOST, *Science and Technology Statistics 2007*, February 2008.

In its eleventh five-year plan the Chinese government proposes to improve administrative procedures and the management of public expenditure. While China's S&T policy already seeks to enforce accountability and improve the management of government-funded R&D through the introduction of an evaluation system, the implementation of the MLP will increase the pressure to evaluate the performance of government-funded R&D.

Since its accession to the World Trade Organization (WTO), China has become more active in bilateral and multilateral R&D programmes and projects and engages in a certain amount of joint activity. Examples include Galileo, ITER, and the Framework Programme of the European Union. Under the EU Sixth Framework Programme, China was the second largest country in terms of the number of projects in which it participated. As China gradually becomes a major player in global innovation, its R&D system will become more global. This also requires Chinese government departments and relevant institutions to engage in the evaluation of international co-operation programmes or projects. However, the evaluation system is presently relatively weak in China, as evidenced by China's inability to meet the evaluation requirements set by its co-operation partners.

evaluation system to be set up to provide objective assessments of the government's policies, projects and programmes based on professional methods of evaluation and to improve administrative efficiency and ensure sound decision making.

Reasons for focusing on the evaluation of national R&D programmes

There are several reasons for focusing on the evaluation of national R&D programmes, specifically government-funded R&D, when introducing China's system for evaluation. Owing to the different duties and responsibilities of MOST, CAS and NNSFC, the type and focus of their evaluation processes have evolved quite differently. MOST is responsible for the evaluation of government-funded R&D. It also establishes relevant policies and manages evaluation activities, which are set out in its mission statement approved by the State Council. The main focus of evaluation at NNSFC is on project selection (*ex ante* evaluation on the basis of peer reviews). At CAS, evaluations mainly involve internal R&D activities, primarily of the R&D labs, key projects and the research institutes. Since MOST deals exclusively with government-funded R&D and carries out formal evaluations, it is a logical place to start the discussion of China's evaluation system.

As stated above, MOST carries out the evaluations of national R&D programmes. It is widely recognised in China that national R&D programmes play a significant role in the country's R&D system. These programmes are the country's most important policy tools for innovation. They allocate public resources to national priorities identified by the government and determine the most important government-funded R&D activities. For example, the National Hi-tech R&D Programme (the 863 Programme) and the National Key Basic Research Development Programme (the 973 Programme) have been the most important means of concentrating public resources on priority areas for S&T development to meet China's social and economic development needs.

MOST also carries out formal evaluations, which rely heavily on evidence and systematic design and implementation tools. Such evaluations have been relatively rare in China and tend to concentrate on the national R&D programmes. While some R&D programme evaluations can be compared with those of other countries, they are relatively few in number. The evaluations described here are recognised as good practice in China. For policy evaluation, China is still at a preliminary stage and it is not easy to find suitable examples. For instance, the subject of e*x ante* evaluation (proposal review for project selection on the basis of peer review) is not addressed.

Methods of the case study

The case study follows the general TIP guidelines and the analytical framework, so that the results can be compared with those of other countries. Starting from the general guidelines and focusing on the key questions to be addressed by each case study, the study team based its work on the actual situation in China and established an analytical framework and a list of key issues.

The case study is based on desk study and expert review. Researchers examined existing information including relevant policy documents, statistics and evaluation reports. Interviews and discussions were held with a number of government officials, R&D researchers and evaluators. Specific evaluations, institutes and events with a strong influence on the development of the evaluation of R&D in China were also reviewed. Some of these are described below in boxes. For some issues mentioned in the general guidelines (mainly about goals, strategy and planning for evaluation), it is difficult to make detailed statements at present, as most evaluations have not addressed them. They are therefore only mentioned briefly.

3. Institutional framework for the evaluation of R&D at MOST

R&D evaluation is a new concept in China, and prior to 2000, there were no relevant laws or regulations. Through an initiative of MOST, regulations on R&D evaluation were released in 2000 and

2002.[30] They aim at promoting the development of R&D evaluation. They do not deal with strategy, planning, implementation or budgeting of the evaluations.

In December 2007, the People's Congress revised the Law on Science and Technology Advancement, which entered into force on 1 July 2008. The new law stipulates that the state will establish and improve the S&T evaluation system, which favours local innovation. It requires the evaluation of investments in S&T. While the principle has been established, the follow-up policies and regulations have yet to be determined.

This section focuses on the institutional framework for the evaluation of R&D programmes at MOST and briefly discusses strategy and planning for national R&D programme evaluation.

Evaluation requirements in regulations on the management of national R&D programmes

Recently, in order to improve the management of national R&D programmes and in response to the general public's demand for accountability, the ministers and officials of MOST have accorded more importance to evaluation. For example, evaluation requirements have been included in regulations on the management of national R&D programmes. During the eleventh five-year period (2006-10), the regulations for three major national R&D programmes – the National Hi-tech R&D Programme (863 Programme), the National Key Basic Research Development Programme (973 Programme) and the National Key Technologies R&D Programme – stipulate the evaluation of projects and sectors. They do not mention overall programme evaluation but focus on projects implemented under the programme's framework. Therefore, regular evaluations of programmes are not planned and evaluation is conducted on a case-by-case basis. Table 5.1 summarises the evaluation requirements for projects and sectors.[31] It is not clear which programmes are to be evaluated, when evaluations should be done and who is responsible.

30. They include the Provisional Regulation of S&T Evaluation Management (MOST, 2000); the Regulation of S&T Evaluation (MOST, CEPD, CET and MOF, 2002); the Policy Statement of Improving Activities for S&T Evaluation (MOST, MOE, CAS, CAE and NSFCC, 2002).

31. A national R&D programme is often structured with three levels: programme, sector and project. For example, the 863 Programme has ten sectors and each sector has its goals and implementation plan. The goals of the sector are to be achieved through projects.

Table 5.1. Evaluation requirements in management regulations of three major national R&D programmes

Programme name	Appraisal	Mid-term evaluation	Completion evaluation	Performance evaluation
863 Programme	All projects to be appraised through peer review or expert panel review.	Mid-term evaluation conducted by a professional evaluation organisation to assess the implementation of megaprojects or sectors.	All projects to be assessed when completed. This is organised by the MOST programme management offices.	Performance evaluation of megaprojects and sectors to be conducted after completion.
973 Programme	All projects to be appraised through peer review or expert panel review.	All projects to be evaluated after two years by an expert consultant group, with a focus on the project's status and prospect.	Project completion evaluation to be conducted by an expert panel commissioned by MOST. Evaluation of sub-projects to be conducted by the chief scientist of the project. The focus is on the achievement of objectives, effectiveness and nurturing of talent.	No requirement.
National Key Technologies R&D Programme	All projects to be appraised through peer review or expert panel review.	All projects of more than 3 years of duration to be evaluated at mid-term, organised by MOST.	Evaluation of completed projects to be organised by MOST	Performance evaluation to should be conducted for all projects. Can be combined with mid-term evaluation or completion evaluation.

Management structure for the evaluation of national R&D programmes

At MOST, the main actors for the evaluation of national R&D programmes are the Division for S&T Evaluation and Statistics of the Department of Development and Planning (DDP), different operational departments (ODs), and the National Centre for Science and Technology Evaluation (NCSTE).

The DDP is responsible for co-ordinating national R&D programmes and acts as manager of R&D evaluations. Its functions relating to evaluation are: *i)* to conduct research on theories, norms and standards of R&D evaluation; *ii)* to put forward regulations for managing evaluations; and *iii)* to organise and co-ordinate the evaluation of R&D policies, R&D strategies, R&D programmes and megaprojects.

Different ODs are responsible for the preparation and organisation of R&D programmes. For example, the Department of Basic Research is responsible for the National Key Basic Research Programme (973 Programme). Therefore, its evaluation is usually commissioned by the DDP and the relevant ODs. In some cases, MOST leadership will commission an evaluation to address issues of concern to the OD leadership. The commissioners of the evaluations generally state the purpose of the evaluation and identify issues to be addressed.

Since its establishment in 1997, the NCSTE has the main responsibility for implementing evaluations of national R&D programmes and projects and of science and technology policies. As a professional evaluation agency, the NCSTE provides solid evidence for R&D decision-making and makes recommendations for improving the management of R&D through objective, impartial and independent evaluations. Box 5.1 specifies its functions.

Box 5.1. The Chinese National Centre for Science and Technology Evaluation

Founded in 1997 with the approval of the Ministry of Science and Technology, the NCSTE is one of the leading organisations in the field of evaluations in China. NCSTE is responsible for implementing evaluations of R&D programmes, policies, institutes, as well as megaprojects. The NCSTE's responsibilities are as follows:

1. Evaluation of major science and technology (S&T) development strategy.
2. Evaluation of various national R&D programmes.
3. Performance evaluation of national R&D institutes.
4. National R&D project and/or programme budget appraisal.
5. R&D human resource evaluation.
6. Regional innovation capacity evaluation.
7. Provision of technical support and evaluation quality control for other MOST agencies.
8. Research on R&D evaluation norms, standards and methodologies.
9. International co-operation on R&D evaluation.

The NCSTE's human resources include: *i)* 25 permanent staff specialised in management consulting, public R&D policy research, technological and economic analysis, R&D evaluation and development evaluation, etc.; *ii)* about 25 contracted senior experts and advisors who are either senior specialists or retired senior officials in various fields; and *iii)* an expert database with approximately 3 000 registered experts.

In order to learn about international practice and experience with evaluation, the NCSTE has established close links with partners in various countries and international organisations, such as the United States, Canada, Japan, Korea, the United Kingdom, Germany, Denmark, the Netherlands, UNDP, the OECD, and the World Bank. In the past ten years, NCSTE has held a number of evaluation workshops, seminars and training courses in Beijing in co-operation with international organisations and foreign ministries.

Standards for R&D evaluation in China

Mandated by the MOST, NCSTE has drafted the Uniform Standards for Science & Technology Evaluation[32] (the standards), which was published in 2001. In the same year, MOST made this the reference document for government regulation on evaluation management.

As the first and only R&D evaluation standards in China, the standards have three objectives. First, they give guidance on ethical conduct for evaluators and other actors in science and technology (S&T) evaluation. Second, they provide standards to define the professional practice of various S&T evaluation

32. In China, especially at MOST, the term S&T evaluation is commonly used. However, judging from the purpose, content and implementation of evaluation, evaluation of R&D activities would be more appropriate. This is therefore the term used here except when referring to published documents or reports, when the term S&T evaluation is used as in the documents or reports themselves. Therefore, when introducing the standards, the term S&T evaluation is used.

activities to enhance the quality of evaluation processes, and to improve the utility of evaluation results. Third, they can be used as fundamental materials for training for S&T evaluation in China.

In order to achieve these objectives, the standards are divided into two parts: core content and reference content (Box 5.2). The former includes principles on ethical conduct in evaluations and the standards for professional practices. The latter includes further explanations of the core content and discussions of typical evaluations.

Box 5.2. Structure and content of the standards

The standards are divided into core content and reference content, each of which is subdivided into two sections.

Core content

Section1. Guiding principles on ethical conduct of evaluation

It covers rules of behaviour for evaluators and evaluators' relationship with evaluated bodies, clients and end users, with the focus on the evaluator's behaviour. It also provides the terminology on science and technology evaluation used in the standards.

Section 2. Standards for professional practices

It addresses technical issues regarding evaluations, including the major evaluation procedures, the roles of the actors in evaluations, key steps and critical issues in the design and implementation of evaluations, detailed requirements for evaluation reports, and commonly used methodology and tools.

Reference content

Section 3. Explanations of the core content of the standards

It gives further clarification, explanations and supplements to the core content. It is mainly targeted at the key issues and issues that are easy to misunderstand or about which there is some debate.

Section 4. Discussions on typical cases

On the basis of current S&T evaluation practices in China, this section provides ten cases of evaluation design and implementation as a reference to facilitate the readers' better understanding of the core content.

Since MOST and NCSTE issued the document, the standards have been used in different types of R&D evaluations nationwide. They also provide a basis for evaluators from various regions and institutions to share R&D evaluation experience, and for evaluators to standardise their work. Moreover, it constitutes the basic material for training in science and technology evaluation. More than 800 evaluators from 70 evaluation institutions across China participated in S&T evaluation training courses organised by NCSTE between 2001 and 2003.

Strategy and planning for national R&D programme evaluation

Although MOST increasingly recognises the importance of evaluation, it does not presently have either a strategy or a schedule for evaluations of national R&D programmes. Rather these evaluations are conducted on a case-by-case basis. There is no regular budget allocation and no timing requirement.

Although the purpose of individual evaluations is clear, it varies to some extent depending on the evaluation. On the whole, the purpose of evaluations of national R&D programmes can be summarised as follows: *i)* to objectively assess the appropriateness of the goals, implementation and management, and the effectiveness and impacts of the programme; *ii)* to learn from past experience and practices and identify weaknesses in the programme in order to improve its management; and *iii)* to provide evidence to the programme management and leadership of MOST for decision-making purposes.

4. Design and implementation of an evaluation of national R&D programmes

This section describes the design and implementation of an evaluation of national R&D programmes. It also discusses the role and activities of decision makers, programme managers, external experts, and other stakeholders.

Design and organisation

When the NCSTE is asked to carry out an evaluation, it organises an evaluation team. The team is composed of two types of experts: evaluators from NCSTE and external experts from specific technological fields who provide opinions. An NCSTE evaluator acts as team leader. This ensures that the evaluation respects the evaluation standards and fully considers the characteristics of the specific R&D programme. In the case of a comprehensive programme evaluation, the evaluation team is divided into groups to address different themes.

In addition to the evaluation team, a steering committee (SC) is also organised to assist in the evaluation. The SC is composed of officials including commissioners, programme managers and staff from the implementing agencies of the national R&D programme. Sometimes the SC also includes representatives from other stakeholders, including industry. The SC is responsible for co-ordination and makes decisions on major issues that arise during the evaluation. In order to assess the progress of the evaluation process, the SC holds periodic meetings

The design of an evaluation of a national R&D programme is mainly the responsibility of the evaluation team, but the final decision is generally reached by consulting with the commissioners. The evaluation team prepares a draft design for the evaluation based on the commissioner’s needs and the desk study of the related programme documents. The draft is then submitted to the commissioner. The draft document covers the objective, scope methodology, process, and work plan of the evaluation. The commissioner discusses the document with the evaluation team and makes some suggestions. On the basis of the discussion, the evaluation team revises the document and resubmits it to the commissioner. In general, the commissioner and the evaluation team confirm this document as final. After the evaluation has been completed, the team leader and a couple of evaluation team core members are responsible for preparing the evaluation report. The final evaluation report is submitted in the name of NCSTE, which is responsible for the evaluation results.

Methodology

The NCSTE has developed a relatively mature evaluation framework that covers programme goals and objectives, programme management and implementation, and programme effectiveness and impacts. Each dimension is examined with the use of some key questions (Table 5.2). The evaluation team can specify the questions to develop indicators that reflect the features of a given programme. For example, in the evaluation of the National Key Basic Research Programme (973 Programme), the evaluation of the programme’s effectiveness and impact took into account major national development needs, progress of pioneering basic research, nurturing of human resources, academic communication and co-operation, and building of the research base.

Table 5.2. Framework for national R&D programme evaluation

Criteria	Key questions
Goals and objectives	• Are the programme's goals and objectives clear? • Is the programme designed so that it is not redundant or duplicative of any other programmes? • Is the programme planned in an efficient and effective way to achieve its goals and objectives?
Management and implementation	• Is the management model suitable for the implementation of the programme? • Is the programme and its projects organised and implemented in an effective and efficient way? • Is the distribution of programme funds appropriate? • Are the project implementers competent to carry out the projects?
Effectiveness and impacts	• What are the outputs of the programme, such as papers, patents, etc.? • Has the programme nurtured qualified human resources for R&D? • Has the programme enhanced the research infrastructure in its field? • Has the programme addressed bottlenecks of technology development in its field? • Has the programme facilitated the development of high-technology industry? • Has the programme promoted co-operation between industry, university and research?

Evaluations make use of both quantitative and qualitative indicators, with more emphasis on qualitative indicators. The current information management system (IMS) for national R&D programmes impedes accurate evaluation because some key data are missing, such as data on the effectiveness and socioeconomic impact of the programme. The quantitative indicators are based on data provided by the IMS or by questionnaire surveys, while qualitative indicators are obtained from evidence collected at workshops, interviews, and questionnaire survey.

Methods used in evaluations are desk studies, questionnaire surveys, field visits, focus group meetings, statistics and crosscutting analyses. The desk study reviews programme and project documents and related documents to understand the nature of programme and to collect the evidence needed for the evaluation. The self-administered questionnaire survey usually covers three types of actors involved in the programme: management experts, scientists and organisations.[33] Focus group meetings are an important way to collect information on the programme and the opinions of the three types of stakeholders.

Peer review is used in different types of evaluations. At MOST, peer review is used in project appraisal and largely determines the approval of projects. In evaluation of national R&D programmes, peer review is usually in the form of peer panel review, mainly for consultation about technical issues. During an evaluation, the evaluation team organises several workshops and invites a group of experts in certain S&T fields to participate. In the workshops, the evaluation team consults the experts about the quality of R&D results produced by the programme. If their judgements concur, their view is adopted in the evaluation. If their judgements differ, all views are presented in the evaluation report.

33. "Organisations" includes the universities, research institutes and enterprises for which the scientists work.

Process

When the design of the evaluation has been finalised, evaluators associated with the NCSTE carry it out. It should rigorously follow the procedure set by NCSTE. Generally speaking, there are five steps in the process of evaluating national R&D programmes: evidence collection and analysis, thematic evaluation, crosscutting evaluation, report preparation and interaction with commissioners, and finalisation of the report.

The evidence collected and analysed includes R&D project data, programme statistics, programme-related documents such as project appraisal documents, mid-term reviews and final reports, related policy documents, and opinions of stakeholders. Evidence is collected from statistics, questionnaire surveys, field visits, focus group meetings and interviews.

Poor information management systems for national R&D programmes make it difficult to collect evidence. In 2006, MOST introduced a uniform IMS for national R&D programmes and projects. At present, data on project appraisals have been computerised, while data on project implementation and completion have only been partly entered into the database. The quality of the data, particularly on project effectiveness and the application of results, is particularly weak. This has made it difficult to evaluate national R&D programmes.

Thematic evaluations are conducted to address leading issues at the time of the evaluation, such as participation by enterprises in national R&D programmes, industrialisation of R&D results and structure of R&D project implementers. These timely issues are usually proposed by programme management and sometimes by the MOST leadership.

Based on analysis of thematic evaluations, a crosscutting evaluation is carried out to draw conclusions. The conclusions are classified according to the evaluation framework described above: programme goals and objectives, programme management and implementation, and effectiveness and impacts of programme. Once the draft evaluation report has been prepared, the evaluation team consults with the commissioner and then amends and finalises the report.

During the evaluation process, the evaluation team interacts with those who were evaluated. These are the officials and staff of operational departments, implementers of R&D projects, and the programme's management experts. The precise role of those who are evaluated is discussed below. The evaluators and the evaluated mainly interact through focus group workshops. The evaluators present the purpose and design of evaluation to those evaluated to help them understand the evaluation. The evaluators also inform them on the progress of the evaluation. Those evaluated give their opinions and recommendations on the evaluation as well as on the R&D programme. This interaction helps both sides better understand each other.

Role of stakeholders in the implementation of evaluation

While the evaluation is being carried out, the DDP gives the evaluation all necessary support, such as stakeholder co-ordination. Operational departments usually participate in discussing the evaluation design and are responsible for providing the necessary data and material on the national R&D programme. The project implementers under a programme are interviewed by the evaluation team and provide information and their opinions on their project.

External experts mainly consult on important technical aspects of R&D projects and give their opinions on the goal, priority setting, management and implementation of the programme. They also fill in the self-administered questionnaire provided by evaluators. Other stakeholders fill in the self-administered questionnaire, participate in focus group workshops and are interviewed by evaluators.

Box 5.3.The evaluation of China's National Key Basic Research Development Programme (973 Programme)

In March 1997, The National Key Basic Research Development Programme (973 Programme) was introduced to strengthen basic research in line with national strategic targets. The 973 Programme covers six sectors: agriculture, energy sources, information, resources and environment, population and health, materials, plus cross-disciplines and frontier sciences. Projects in the 973 Programme generally have a five-year implementation period.

The main tasks of the 973 Programme are to strengthen and support research on a number of major scientific issues of importance to national socioeconomic development, to consolidate a highly qualified contingent for basic research and cultivate personnel with innovative capabilities, and to improve and perfect programme management to create a sound environment for innovation.

From April 2005 to March 2006, the NCSTE conducted the first evaluation of the 973 Programme. The evaluation adopted a framework based on NCSTE's past experience and practice, which covers goals and arrangement, management and implementation, and effectiveness and impacts. The methods used included policy analysis, statistics analysis, a questionnaire survey, case studies, field visits and focus group meetings. It further aimed to promote lessons learned and to provide evidence for decision making.

The main findings from the evaluation were as follows. First, the 973 Programme was launched in the 1990s by the Chinese government to enhance basis research and to improve China's indigenous innovation capacity, At the time, there was a lack of national needs-oriented basic research; and the programme has significantly improved China's basic research system and the integration of basic research and national needs. Second, the management mode of the 973 Programme is generally suitable for the programme's characteristics and its implementation. It works efficiently towards achieving the programme's goals, but needs to be improved in some respects. Third, owing to the relatively short period of implementation, the overall effectiveness, and impact of the programme have not been fully demonstrated and breakthroughs to resolve important national needs or to reach the scientific frontier are still on the horizon. The evaluation recommended the establishment of a stable funding mechanism for the 973 Programme so as to increase total funding for the whole programme and the intensity of project funding.

Box 5.4. The evaluations of China's National Hi-tech R&D Programme (863 Programme)

The Hi-tech Research & Development Programme (863 Programme) is China's largest R&D programme. It is committed to addressing strategic, advanced, and forward-looking high-technology issues that are crucial to the nation's future development and security. It plays a leading role in the future development of emerging industries by developing, integrating and applying proprietary high technologies.

So far, three evaluations of the 863 Programme have been conducted, in September 1995, August 2000 and May 2006. All three have adopted a combination of comprehensive and thematic evaluations. The comprehensive evaluation covers programme goals, programme management, and effectiveness and impacts. The thematic evaluation covers that are relevant at the time. For example, the first two evaluations mainly focused on the adjustment of programme goals, the management model of the programme, accountability in the programme management, and the impacts of the programme on high-technology industrialisation. The third emphasised issues such as the organisation and implementation of megaprojects, the participation if enterprises in the programme, patents produced by the programme, etc.

The usual methods were followed: desk studies, field studies, surveys using questionnaires, and data and information from the 863 Programme management, and information collected directly during the evaluation. Probably the most striking and original aspect of this evaluation was the "stakeholder dialogue approach", with the organisation of several roundtable workshops. These included programme managers or persons with a direct interest in the programme such as project managers, conductors of 863 projects and S&T experts not participating directly in 863 projects. Debates were led by the NCSTE professional evaluation staff, which afterwards summarised them in the form of reports.

Although the evaluation results have not yet directly led to changes in the programme, some results received the attention of 863 Programme and MOST management and lead to follow-up measures. For example, the third evaluation found that during the tenth five-year period (2001-05), enterprises became major project implementers.

They took charge of or participated in 50% of projects and received 60% of central funds, and the number of patents produced was 3.8 times that of the preceding 15 years. Yet much information about the operation and financial status of these enterprises was lacking or of poor quality, which made it difficult to appraise and control risks that could emerge during project implementation. Thus, the evaluation recommended that the DDP improve its MIS and supplement its information on the operation and financial data of enterprises. The DDP and the programme management recognised the importance of these findings and took measures to track and improve data about enterprise development and patents.

Source: Chinese National Centre for Science and Technology Evaluation, Evaluation Report of the Tenth Five-year National Hi-tech Research & Development Programme (863 Programme), November 2007; Chinese National Centre for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-2000), March 2001; Chinese National Centre for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-95), January 1996.

5. Utilisation of national R&D programme evaluations

This section discusses the utilisation of the evaluation by programme management and the response to the evaluation results by the leadership. At MOST, feedback from evaluation results is weak and unstable owing to the lack of any institutional mechanism. Evaluations have had little effect on priority setting and on budget allocation and co-ordination. They do, however, have an indirect influence on programme management. There is much room for improving the use of evaluations.

Circulation of evaluation reports

The evaluation reports on national R&D programmes have not all been made public. This is mainly because MOST considers the reports as being for internal use. Summaries of the early programme evaluations were published in some newspapers, but later evaluations were not. MOST does, however, circulate the reports within the organisation and occasionally to some external stakeholders. The commissioner of the evaluation determines what is circulated. In sum, there is no standard procedure for circulating evaluation reports and practices differ.

Utilisation of evaluations by programme management

The primary users of the evaluation findings are the commissioners, DDP and operational departments. They are informed about the implementation and effectiveness of the programmes and thus take measures to improve the management of the programme. In some cases, the evaluation reports are also circulated to implementing agencies of the programme at MOST.

More generally, the evaluation's role in project management takes place during the evaluation process. The evaluation team informs programme management about problems identified in the management process and discusses the causes and possible remedies with the management. This allows for resolving some problems before the evaluation is completed. The evaluations also make recommendations for improvements that help managers enhance their management and planning capacities.

Leadership's response to the evaluation results

The leadership's response to evaluation results can affect the importance accorded to evaluations. Early evaluations received more attention from the leadership than more recent ones. Currently, when an evaluation is completed, the evaluation team briefs the minister responsible for the programme. Sometimes the evaluation team or the DDP also briefs the leadership on the findings, in which case the leadership attaches more importance to the evaluation and the issues. Usually, the leadership responds and indicates which departments and agencies should address the relevant issues. In such cases, an evaluation may influence decision making and lead to policy improvements.

6. Observation and key challenges

This section offers some observations about the present status and the key challenges for the evaluation of R&D in China.

Current status of R&D evaluation

China's R&D system has developed strongly and the country occupies a special position among developing countries in terms of R&D. The evaluation of R&D, however, does not seem to be keeping pace with other aspects of the system. At present, China's R&D evaluation system is at an early stage of development.

Although the terms "evaluation" and "assessment" are broadly used for R&D management in China, there is no shared understanding of the function, implementation, and utilisation of evaluations. China's R&D evaluation system can be characterised as follows.

Weak institutionalisation*:* Compared to other aspects of the R&D system, the institutional context is weak. Over the past decade, government departments have issued regulations on the regular evaluation of the government's programmes and key projects. The regulations have not clearly stipulated which programmes should be evaluated, when evaluations should take place, or how to carry out the evaluations. There is no annual planning of work and budget for evaluation activities, and there are no specific implementation guidelines for different kinds of evaluation. In consequence, the regulations pertaining to evaluations have not truly been implemented. The evaluation of R&D has not yet been incorporated into the management and decision-making process and is not a regular and obligatory activity.

Imbalance in evaluation capacity*:* Attention is paid to the capacity to undertake an evaluation, but little attention is paid to the capacity to use evaluation results. The gap between China and the developed countries is reflected in the lack of professional personnel to design and implement evaluations. The gap is even more reflected in the lack of competent officials to manage and co-ordinate evaluation activities. There is already a small group of professional evaluators in China. Yet within government departments, only a few officials understand why the government needs evaluation and know how to ensure that evaluation findings are used appropriately.

Some good programme evaluation practices*:* China has already completed some programme evaluations, such as those of the 863 Programme and 973 Programme. Judging from the evaluation framework and methods adopted in these cases, the methods are not very different from those used elsewhere in the world. For instance, the international evaluation community regards the Country-led Joint Evaluation Dutch ORET/MILIEV Programme in China[34] as a success.

Use of evaluations depends on top leaders*:* The influence of evaluation results on decision making depends mostly on top leaders in government departments. Even if two programme evaluations are basically on the same level in terms of design, information collection and analysis, the influence of the

34. The evaluation was jointly launched and completed by IOB (the Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs) and NCSTE (National Centre for Science and Technology Evaluation of China). NCSTE and IOB made a joint presentation on the evaluation at the Sixth Meeting of the OECD DAC Network on Development Evaluation, 27-28 June 2007, Paris. The report of the joint evaluation has been placed on the DAC website.

evaluation on decision-making varies greatly, owing to different officials' understanding of evaluations. In sum, the institutional arrangements and the feedback and learning loops that involve policy makers and programme managers are lacking.

Building China's government-funded R&D evaluation system

The government-funded R&D evaluation system should cover R&D at different levels, starting from projects and then moving to programmes and ministries and finally to the system as a whole. A further question is whether to start building at the project, programme, ministry or national level. China may not yet be in a position to adopt a strategy at the ministry or country level; a mixed approach may be a more realistic choice at present. In terms of what has already been and is being done, the point of departure is usually pilot work for core national R&D programmes and key projects supported by the programmes.

In terms of evaluation priorities, outcomes should perhaps be the first priority. The Chinese government is under increasing pressure to produce and to demonstrate the results of R&D. MOST is held accountable for the national R&D programmes and is required to report on performance. Therefore, the main initial focus of evaluations should be the identifiable changes resulting from the programmes and the key project deliverables.

It is more difficult to measure the impact of R&D than the outcomes. The impacts are usually generated not simply by a single programme/project, but also depend upon numerous factors, including exogenous factors, that are very difficult to attribute directly to the programme/project. The goal of measuring project impacts might be addressed at a subsequent stage.

Peer review in the programme evaluation

Peer review plays different roles in *ex ante* and *ex post* evaluations. Therefore, the problems that arise during peer review as an *ex post* evaluation tool must be addressed differently than those that arise during *ex ante* review. In China, criticism or debate relating to peer review mainly focuses on its role in *ex ante* evaluation. As a fundamental mechanism for project selection, almost all government departments use peer review. Table 5.4 shows that, for the three major S&T programmes, all projects are to be appraised through peer review or expert panel review. Peer review is an important element in the allocation of R&D resources, and the most serious criticism of peer review at present concerns its fairness.

Government departments try to improve the peer review process for project selection in various ways: by setting up a pool of reviewers pre-approved by relevant management departments, by randomly selecting reviewers from the pool and obliging reviewers to declare in advance any possible conflict of interest; by sending feedback of the panel summary to the applicant, etc. Transparency is the most important way to improve peer review and it is increasing in China.

Much less attention is currently given to peer review in *ex post* evaluations. As noted above, peer review usually takes the form of expert panel review in evaluations of national R&D programmes. This primarily involves providing opinions on technical issues to the evaluation team, such as the quality of the programme outputs. Expert panels do not, however, make judgements themselves. In programme evaluations, the evaluation team's judgement is based on careful analysis.

Guidelines can help peer review play a better role in programme evaluation. Before inviting experts to participate in the programme evaluation, the evaluation team usually prepares guidelines, including background information on the programme being evaluated, key issues that experts need to address, the consultation procedures, requirements for the reviewers' report, and so on. Practice shows that the quality of peer reviews is enhanced when guidelines exist and are used.

Programme evaluation and priority setting

As noted above, evaluations are used more for improving existing programmes than for supporting resource allocation. There are no data to show that the results of evaluation affect the setting of priorities for the following reasons:

- First, when the evaluation task is assigned, the proposed objectives of evaluation usually do not include priority setting.
- Second, the key moment for priority setting is the first year of every five-year plan. Evaluation does not fit into programme cycles because of the absence of any institutional arrangements for programme evaluation.
- Third, in terms of timing, evaluations are often conducted when the programme has achieved a certain level of results. Thus, the evaluation of a programme usually starts in the last year of the five-year plan. Moreover, since an efficient monitoring system that would facilitate evaluations is lacking, it takes quite some time to collect the necessary data and information. As a result, for priority-setting decisions are often made before programme evaluations have been finalised.

The challenge of learning from international experience

Government officials and experts in China have tried to draw lessons from current international experience. However, they must consider the extent to which that experience suits the Chinese context, which elements can be especially valuable for establishing a Chinese R&D evaluation system, and which other elements may not be suitable. There are important international challenges that must be faced in order to strengthen evaluations. The Chinese government will need to pay attention to the following issues.

First, the challenges of transferring policy or management tools from developed to developing countries are well known. Since evaluation is a tool from a "Western" management tradition, it may be asked whether the lessons are applicable or applicable to the same extent in China.

Second, since China and "Western" countries are not at the same stage in the development of the economy, society and R&D system, some earlier evaluation practices in Western countries may better fit China's current situation and some current Western practices might be better kept for the future. For example, the evaluation of the impact of R&D programmes has recently attracted much attention in the evaluation community, especially in OECD countries. China however has given less attention to it.

In terms of design and implementation of evaluations, learning from international experience should be comparatively easier, as Chinese evaluators face similar challenges. For example, an indicator framework to guide performance evaluation of R&D programmes has been an important concern for policy makers, programme managers, and evaluators in China. However, no agreement has been reached on the kind of indicator framework that would be suitable and how to develop it.

REFERENCES

Chinese National Center for Science and Technology Evaluation (2006), Evaluation Report of National Key Basic Research Development Programme (973 Programme), March.

Chinese National Center for Science and Technology Evaluation (2007), Evaluation Report of the Tenth Five-Year National Hi-tech Research & Development Programme (863 Programme), November.

Chinese National Center for Science and Technology Evaluation (2001), Evaluation Report of National Hi-tech Research & Development Programme (1986-2000), March.

Chinese National Center for Science and Technology Evaluation (1996), Evaluation Report of National Hi-tech Research & Development Programme (1986-1995), January.

Department of Development Planning, MOST (2008), Science and Technology Statistics 2007, February.

Chinese National Center for Science and Technology Evaluation (2001), *Uniform Standards for Science and Technology Evaluation*, Price Press, China.

Ministry of Science and Technology (2000), Provisional Regulation on the Science and Technology Evaluation (Rg.No. 2000-588), December.

Ministry of Science and Technology and Ministry of Finance (2006), Management Regulation of National Hi-tech R&D Programme (863 Programme) (Rg.No.2006-329), July.

Ministry of Science and Technology and Ministry of Finance (2006), Management Regulation of National Key Basic R&D Programme (973 Programme) (Rg.No.2006-330), July.

Ministry of Science and Technology and Ministry of Finance (2006), Provisional Management Regulation of National Key Technologies R&D Programme (Rg.No.2006-331), July.

Ministry of Science and Technology (2008), Handbook of the Ministry of Science and Technology of the People’s Republic of China.

OECD (2007), *OECD Review of Chinese National Innovation System and Policy – Synthesis Report*, OECD, Paris.

CHAPTER 6. EVALUATION, IMPACT ASSESSMENT AND FORESIGHT IN FINLAND[35]

The Finnish case study looks broadly at the national evaluation system. It examines evaluation activities of a country with a S&T policy in transition. Therefore, this study shows both the current and possible future state of activities. Unlike the previous case study with an emphasis on peer review, this case study focuses on impact assessment of societal impacts with a growing emphasis on foresight analysis.

1. Introduction

The following is a comprehensive case study of the current status of evaluation, impact assessment and foresight in Finland. It also provides an overview of the main results of evaluations carried out primarily since 2005. It is an example of the way in which the Finnish policy regime treats evaluation.

Finland's STI policy is currently in transition. The new policy took its start from a public statement by the Finnish Science and Technology Policy Council (STPC) in August 2007.[36] The policy addresses both the emerging need to evaluate the Finnish innovation system (NIS) as a whole and the need to enhance existing evaluation practices. It also follows to some extent the guidelines set for the OECD project on evaluation systems.

The following discussion also touches upon the general framework for Finnish evaluation activities. Although evaluation is deeply rooted in the Finnish innovation policy system, Finland as of now has no formal or articulate evaluation system. The current system concerns particularly *ex post* evaluation carried out by domestic specialists, often with foreign experts. In this sense, almost everything is evaluated, from projects and single policy measures to programmes and institutions. Even the national innovation support system was evaluated a few years ago. Apart from the ongoing evaluation of the Finnish STI environment,[37] a comprehensive evaluation at the NIS level has not been carried out since the mid-1980s, when it was done by the OECD (*Review of National Science and Technology Policy: Finland*, published in 1987).

35. Prepared by Mr. Kai Husso, Chief Planning Officer, Research and Innovation Council of Finland and Mr. Esko-Olavi Seppälä, Secretary General, Research and Innovation Council of Finland. In January 2008, the name of the Council was changed from Science and Technology Policy Council (STPC) into Research and Innovation Council (RIC). Depending on the context, both names are used. The first version of the manuscript was completed in February 2008. It has not been possible to cover extensively and discuss all the most recent evaluation projects and studies carried out in 2008–09.

36. Much of this material comes from a background document for the STPC statement (24August 2007) on the development of impact assessment and foresight in Finland. While preparing the text for the OECD project on evaluation, the STPC secretariat was greatly aided by the Finnish OECD/TIP delegate, Mr. Pentti Vuorinen, from the Ministry of Employment and the Economy.

37 This evaluation was commissioned jointly by the Ministry of Employment and the Economy and the Ministry of Education and conducted by a group of domestic and international panellists (see www.evaluation.fi).

2. Historical evolution of evaluation

Finland has a long tradition of performance and impact evaluations, as well as of assessments of policy instruments, measures and S&T organisations. This evaluation culture started to emerge in the latter half of the 1980s. It has remained strong and has even strengthened in the past decade. Today, efforts are being designed and implemented in order to fortify connections between the findings of evaluation reports and the development of policy operations. More attention is also being paid to the different needs of evaluation-based information in various stages of the policy cycle. The challenge that remains is to build a firm, interactive connection between evaluation and impact assessment activities and selective decision making.

Owing to the loosely co-ordinated development of evaluation within the innovation system over the years, there are no formal cross-sectoral or horizontal strategies or evaluation practices in place today. In 2008, Tekes[38] and the Academy of Finland[39] launched a still ongoing joint effort to remedy this situation.

Responsibility for improving evaluation and foresight is shared by all key stakeholders of the NIS with the producers of new knowledge and technology, financiers of research and innovation, and the actors benefiting from the results. A particular responsibility for initiating further development measures is carried by public expert and funding organisations (Academy of Finland, Tekes, Sitra) and the providers of innovation support services. This institutional set-up with divided responsibilities emphasises the need to pay increasing attention to co-ordination and steering issues.

Since the early 1990s, all the stakeholders of the Finnish innovation system – ministries, funding organisations, universities, government R&D institutes, public and private intermediaries etc. – have been evaluated, some of them many times. This describes reasonably well the intensity and the depth of the evaluation culture and its evident link to policy making at the different levels of the NIS. Evaluations have usually been linked with methods, procedures and objectives to:

- improve steering of RDI organisations by using the method called "management by objectives and results" or "results-based management and steering;
- increase the volume of public (competitive) R&D funding (channelled through Tekes and the Academy of Finland) and the impact of financing and other means of support;
- ensure that public funding is used effectively and efficiently;

38 TEKES, the Finnish Funding Agency for Technology and Innovation, is the largest public R&D financing and expert organisation in Finland. It finances industrial R&D projects but also projects in universities and R&D institutes. It promotes innovative, risk-intensive projects. In 2009, TEKES had a total budget of EUR 575 million. Together with the Academy, it accounts for over 46% of total government R&D investments. This figure also indicates the share of competitive funding of total public R&D investments. TEKES operates under the auspices of the Ministry of Employment and the Economy. The ministry steers the operations of TEKES through management by objectives and results and uses a set of measurement indicators in this regard.

39 The Academy of Finland's mission is to finance high-quality scientific research, act as a science policy expert, and strengthen the position of science and research in the society. The Academy's operations cover the full spectrum of scientific disciplines. It supports PhD training and careers in research, internationalisation and the practical application of R&D results. In 2009, the Academy had a total budget of EUR 308million. It operates under the auspices of the Ministry of Education, which steers its operations by following the method of management by objectives and results.

- develop the entire NIS, not only individual organisations, fields of research, technology and research programmes, and procedures.

It should be emphasised, however, that political and other decisions are seldom directly derived from the results of evaluations and impact assessments. Instead, the results have often provided a basis for (public) debates on development needs and other issues. Evaluations are expected to give new perspectives and a positive push in an organisation's development or mode of operation.

There are two principal clients or commissioners of evaluations and impact assessments: the Ministry of Employment and the Economy and the Ministry of Education. Furthermore, as mentioned in the government decree on the STPC/RIC, the Council (chaired by the prime minister and comprising seven other ministers relevant to STI policy and ten other members well versed in Council matters) was established to handle important matters concerning research, technology and their utilisation. The Council also addresses issues relating to the development and utilisation of evaluation and impact analysis. In addition, other ministries (*i.e.* those with R&D institutes of their own) have carried out evaluations using different methods.

The two main actors and implementing organisations in evaluation and foresight are the Academy of Finland and Tekes. In addition, universities, polytechnics and government R&D institutes have carried out numerous self-assessments and external expert evaluations on their own initiative since the 1990s. The results of these evaluations and assessments are sometimes used to develop strategies and operations at ministries, research, development and innovation (RDI) policy expert and funding organisations, and within the RIC. However, there is still a lot to do to improve and intensify the utilisation of evaluation findings.

Evaluation and impact assessment at the Academy of Finland focuses on:

1. The research system:

   a. ***Reviews of the state and quality of scientific research in Finland****:* The aim is to provide an extensive review of the state and quality of science and research in Finland once every three years to coincide with the Research Council's term. The next review (the fifth) will be completed in 2009. To date, Academy staff together with the Research Councils usually complete these reviews using variable methods.

   b. ***Evaluations of disciplines and fields of research****:* The aim is to provide information on the societal, technical, and economic impacts of research in a given field. Evaluation criteria may include scientific visibility, researcher training, postdoctoral research careers, international recognition of researchers and societal impacts. The impetus for the evaluation may come from the research community, the Research Councils, other science funding agencies or the authorities. External expert panels complete evaluations.

2. The Academy's own actions and processes

   a. ***Research programmes****:* In evaluating these programmes evaluators consider whether the goals set for the programmes have been achieved, their success in generating new knowledge, the value added and the societal impact. External expert groups complete evaluations, including the societal impact of projects.

   b. ***Centre of excellence programmes****:* In evaluating these programmes the main focus is on their scientific impact, and secondarily their societal impact.

c. ***Reviews of funding application****:* On average, 4 000 different applications for funding are evaluated each year.

3. The organisation of the Academy:

a. ***Evaluations of individual instruments and programmes:*** These evaluations are commissioned by Tekes and usually done by international experts.

b. ***Evaluations, assessments and studies on outcome level****:* These evaluations are done by universities, R&D institutes or consultants independently or commissioned by Tekes).

c. ***National and international indicators and benchmarking on impact level:*** Tekes collects data for indicators.

d. ***System-level evaluations:*** The STPC and ministries commission these evaluations, while Tekes organises the actual evaluations.

In addition to the Academy and Tekes, the Higher Education Evaluation Council plays an important role in the Finnish evaluation field. This Council is an independent expert body assisting institutes of higher education (HEIs) and the Ministry of Education in matters relating to evaluation. Its activities focus on audits of HEIs' quality assurance systems and centres of excellences' evaluation in education.

3. Recent evaluation-related developments in STI policy

Investments in research and innovation have been growing steadily in Finland, particularly since the latter part of the 1990s. Both public and private parties, sometimes in partnership, have increased their efforts to create and use new knowledge, technology and skills. At the same time, the evaluation of the impact of policy measures and funding has become a focus of attention. Evaluation has gradually become an increasingly systematic activity, yielding more reliable and useful results. Evaluation, impact assessment, and foresight are very challenging activities to complete effectively. They require highly developed methods, a range of measures for development, visionary outlook and insight. Hence, there is still a lot to do to develop evaluation and foresight as such and to bridge the gap between results derived from these activities and (political) decision-making.

Since 2005, dozens of evaluation and foresight reports concerning RDI have been published in Finland. The increases in the cost of and investments in R&D, changes in the nature and execution of R&D, and the globalisation of the operating environment have all led to greater need for evaluation and foresight. The Academy of Finland and Tekes have been active in evaluation and foresight, but the Research Institute of the Finnish Economy (ETLA), the Finnish Innovation Fund (Sitra), the Government Institute for Economic Research (VATT), has also produced reports among other important R&D organisations. The key results of these publications are described here.

The Resolution on the Structural Development of the Public Research System (2005)[40] is arguably the most important science, technology and innovation (STI) policy document in recent years. This document points out that the structural development of the public research system is necessary in order to raise the

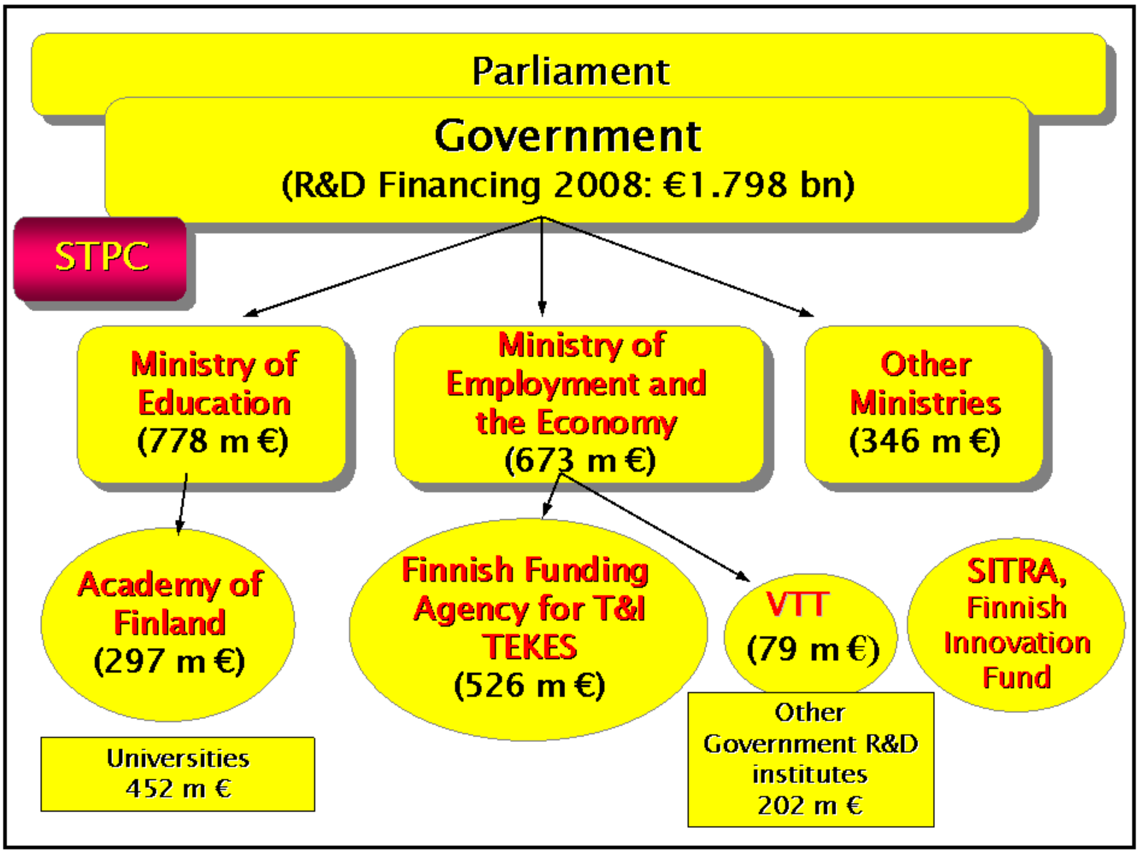
---

40. The Resolution on the Structural Development of the Public Research System of 7 April 2005. The resolution includes the general guidelines and basic principles for structural development measures. Based on the resolution, numerous policy measures (which combine funding and structures) has been initiated and implemented. The resolution emerged from an evaluation project launched in 2003 by the STPC. To a large extent, the resolution is related to a number of separate studies, which shed light on the structural issues of universities, ministries and public research institutes, as well as intermediaries operating at the interface between the public research system and the rest of society.

quality and relevance of R&D. Development measures should be targeted to boosting the prioritisation of activities, the international and national profiling of research organisations, and the selectivity of decision making. Successful implementation requires accurate assessment, high-quality foresight, and a sufficient information base. Along the same lines, the STPC's STI policy review (2006) suggests that the promotion of innovation dynamics, the development of S&T policy tools, and improvements in the productivity and impact of the related measures all require support in terms of the enhancement of competencies related to evaluation and foresight. The role of the Academy and Tekes and the specific objectives of their partnership are to develop evaluation activities, methods and data, and to investigate the impact of measures on the economy and society and the structural development of public RDI systems (Figure 6.1).

Besides being connected to the development of STI, evaluation and foresight are also linked to general societal development. In particular, impact assessments of policy measures and the evaluations of the legislative drafting standards have been strongly emphasised, particularly in the field of sectoral research (*i.e.* research mostly done by government R&D institutes).

Figure 6.1. Structure of Finland's public RDI system

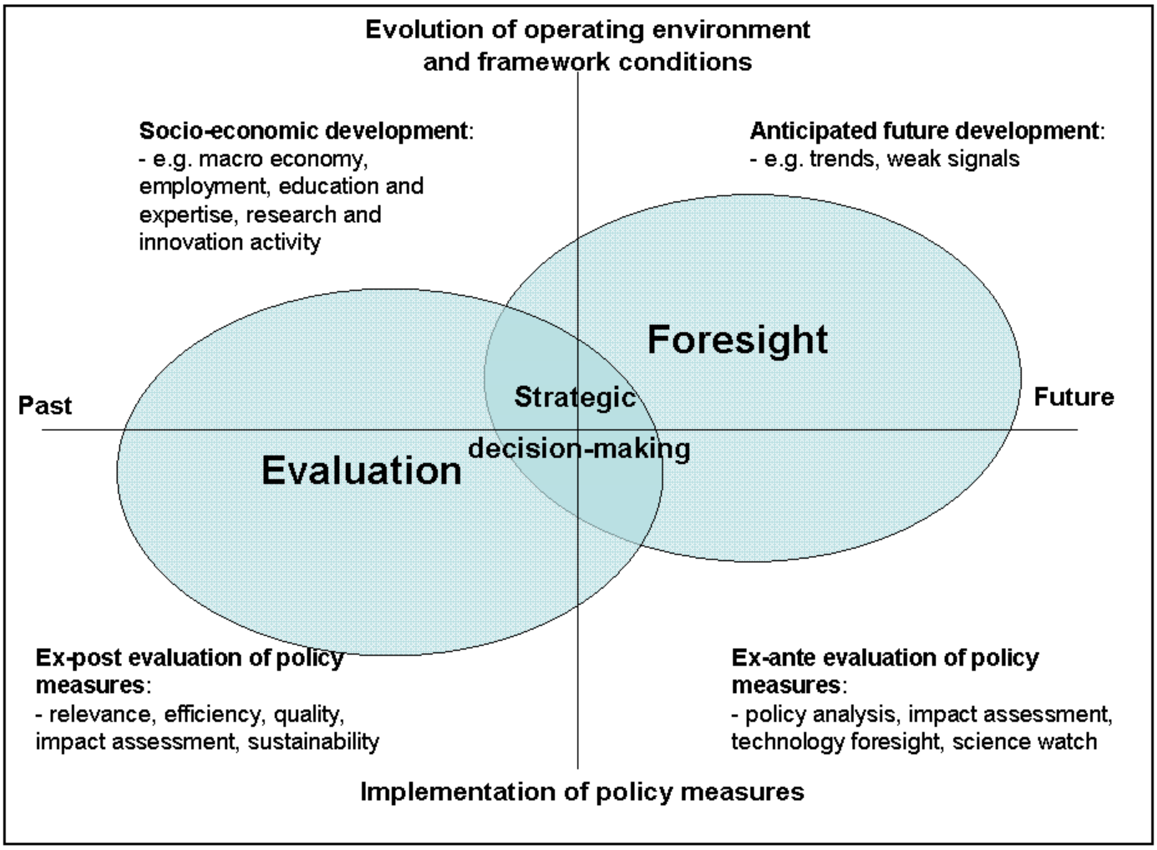


There is a fundamental link between evaluation and foresight: evaluation results are used for foresight purposes and *vice versa*.[41] Accordingly, development measures should target both areas (Figure 6.2). Evaluation and foresight support the continuous development of research and innovation policy and strategic decision making. As such, they are instruments of common learning, understanding and use.

41 For a more in depth discussion, see the section on "strategic policy intelligence" in Chapter 4.

Therefore, it would be valuable to pay special attention to the successful integration of evaluation and foresight and the division of responsibilities for the activities involved.

Figure 6.2. The relationship between evaluation and foresight and their relation to strategic decision-making and the development of policy measures



Source: Based on Valovirta and Hjelt (2005).

4. Definitions of impacts and mechanisms of RDI

Impact is the last phase in a process that begins by organising the input factors for achieving a goal. In a research project, input factors include the necessary material, the intangible resources and the manner of organising the work process. At the macro level, impact refers to a change that has resulted from the perspective of the goals set by society. The impact evaluation process is usually made more difficult because impacts may be harder to predict the further the distance in time and space from the immediate quantitative results of the activity.

Although impact areas can be defined, examining the nature of specific factors and their role in overall societal change is far more difficult. The impact assessments of scientific research implemented by the Academy of Finland (2006) approached the problem of defining impact areas by dividing “impact” into two parts: scientific and social impacts. This is justified for the evaluation of basic research. They found that it is difficult to determine the impacts of STI because the impacts are a result of complex cause and effect chains, which occur over long periods of time. Furthermore, as concluded in the Academy’s review of methods for evaluating the impact of basic research funding, there is no best method for this kind of impact assessment.

Research and its impact manifest themselves as the generation of new information and the growth of information capital. This is more widely reflected in the formation of world views, in social debate, and in the greater understanding of various phenomena and issues. The output and impacts of research are mutually supportive. For instance, PhD training, the growing volume of knowledge and the mobility of experts form a process that can be considered a result of research and a diffusion mechanism for R&D results and competencies, as well as a result of research.

The impact of research depends on the degree of applicability of research results and the extent to which a society can utilise information and skills and transform them into innovations. A large volume of reports published recently in Finland indicate that research may have the following kinds of impact:

- ***Scientific****:* the accumulation and renewal of the knowledge base, including tacit knowledge.
- ***Technical:*** technological solutions; new products and processes; patents; new research equipment; methodological and technical skills.
- ***Economic:*** new R&D and technology-based entrepreneurial activity; technological and social innovations that promote production and foreign trade; the development of the competitiveness, productivity and efficiency of enterprises;
- ***Social:***

  a) societal and cultural impacts, including education (lifelong learning, adult education and in-service training among others);

  b) impacts associated with health and welfare, including safety;

  c) impacts on working life, including new skilled workforce, employment;

  d) impacts on nature and the environment as they apply to people;

  e) political impacts, *e.g.* exploitation of research results and use of the expertise of researchers in decision making;

  f) administrative and organisational impacts, *e.g.* introduction and application of new modes of operation and structures, regulatory environment.

The mechanisms for utilising and distributing the results of research and innovation include:

- partnerships between companies, research organisations, and the administration, as well as measures promoting networking (research and technology programmes, common research centres, cluster projects);
- mobility of researchers and experts;
- new R&D- and technology-based enterprises, including academic entrepreneurship, spin-off companies and activities supporting these (intermediaries, technology centres and science parks, R&D and innovation agents, business incubators);
- transfer of R&D results and expertise from research organisations to those introducing, applying and commercialising these (including patenting, licensing, university companies, IPR transfers against share ownership);
- publications, technical reports, and other documents;
- dissemination and application of tacit knowledge through channels including official, unofficial and personal contacts; education; meetings, workshops and conferences; expert services.

5. Education, research and innovation and their impact

Much of the evaluation of various STI programmes, projects and policies in Finland occurs during the financing process. As this section discusses, this makes it difficult to assess the social impacts of certain decisions. This section describes the scientific and societal impacts of basic research and STI funding from both the Academy of Finland's and Tekes' point of view. Moreover, the impacts of universities and other ministries are reviewed.

Scientific and societal impact of (basic) research

The evaluation of scientific impacts and research standards are a more familiar subject than social impact and its assessment. An example is the Academy of Finland report (2006), which uses bibliometric methods to examine Finnish science and its international status. Given the method used, the results show that the quality of Finnish basic research and its visibility are relatively high compared with international research quality. During the period 1995-2004, for example, Finnish scientific publications received 13% more citations than those of other OECD countries on average.

Some important questions and comments have been raised about the impacts of future research. For instance, Professor Raimo Väyrynen, then President of the Academy, has asked questions that can be summarised as follows*: i)* Is technological research capable of producing sufficient scientific results relative to the public resources allocated for the purpose? *ii)* Can the biosciences generate sufficient social and industrial benefits relative to the significant research resources allocated for the purpose? *iii)* How can the social sciences and humanities justify their social importance in comparison with the relatively modest scientific results they have achieved internationally?

These questions are not easily answered. Bibliometric methods, for example, alone cannot describe the impact of even basic research. Furthermore, bibliometrics do not provide a neutral analysis of various fields of science. Väyrynen's comments illustrate that other measurement methods and qualitative assessments are necessary. The Academy's Research Council for Culture and Society pondered this problem in its impact report, *Civilization Cannot Be Imported* (2006) and made the following proposals:

- It should be emphasised that the most important impact exercised by researchers on society comes through the education of a new generation of experts.
- We should monitor the extent to which researchers influence the development of their own disciplines on a global level.
- It should be accepted that, in most cases, social impacts of research cannot be seen for years, and often for decades.
- Quantitative data should be collected on the concrete effects of research on innovation.
- Researchers should be encouraged to consider the various impacts of their research and to report on them.

In relation to social impacts, the Research Council for Culture and Society considers the researchers to have applied two strategies in the debate:

> "On the one hand, they have pointed out that not all values can be measured in [monetary] terms. The aim is to try and contain the impacts of research within the field of science. Knowledge is considered valuable in and of itself; it is not thought to be necessary [to]

separately… identify any instrumental values. The thinking is that civilised society should support and sponsor science in the same way as it supports and sponsors art and other forms of culture. Science is thus protected from the practices and principles of business. The concepts of accountability, benefits, efficiency, performance targets, competition, measurement and impact are rejected by reference to civilisation and scientific autonomy. Dependence on the outside society is a problem because science has to be able to justify…[its] value…[to] civilisation and [its] autonomy in order to get the funding it needs. Nonetheless, it is accepted that at least within the scientific community, it is necessary to have debate about what is good and worthwhile and what has a positive impact. The other strategy is to accept the challenge presented by society. Questions of impact[s] and… benefits are discussed and debated, but not without weighing and defining the rules of discussion and debate and interpreting the concepts. The utility and benefits of all research must be open to discussion so long as those concepts are understood in broad enough terms".

The Research Council points out that even a strategy with a favourable perspective on the impact debate faces the problem that even a broader concept of economic benefits would be too narrowly defined.

In recent years, the most prominent evaluations of HEIs have been the comprehensive research evaluations implemented by certain universities.[42] With respect to the utilisation of these evaluations, the critical issue is whether they can serve as a basis for the necessary strategic decision making within HEIs or in the ministries and other actors steering these institutions.

The impact of Academy of Finland's research funding

The Academy invited an expert panel to investigate the impact of Academy research funding; their report was published in 2006.[43] As far as impact assessments are concerned, the key findings are as follows: in practice, concrete, detailed impact mechanisms can be described by means of separate innovation-specific and project-specific case studies. "It would be impossible to perform a comprehensive analysis taking into account the cause-and-effect relationships involved in the impact of the Academy of Finland or any other actor in the national innovation system". The report contains some sharp criticisms of the funding operations of the Academy and the instruments they employed. In regards to the key question, the panel has had to be satisfied with somewhat general statements:

"With respect to the Finnish innovation system as a whole, it is important that the Academy of Finland take the goals and needs of the entire innovation system into consideration when setting goals and developing its research funding system.

"The best possible assessment of the effects and impacts of the Academy can be achieved by scrutinising the Academy's operations from different perspectives, combining both quantitative and qualitative research methods".

42. For instance, the University of Helsinki, on its own initiative, has carried out two very extensive research assessment exercises (1999 and 2005). The evaluation in 2005 combined an external assessment by 21 international evaluation panels (a total of 148, mainly European, experts) with an internal self-assessment exercise. The assessment results revealed strengths, weaknesses and opportunities in the university's research activities. The University will spend a total of EUR 15 million of its own funds over the next six years to reward the units that were most successful in the evaluation.

43. Publications of the Academy of Finland 11/2006.

The difficulties of measuring the social impact of scientific research

In the summer of 2006, FinnSight 2015, a joint foresight project of the Academy of Finland and Tekes, was completed. By convening ten expert panels, the project explored the future prospects of science, technology and society. The themes of the panels were: learning and renewing society through learning; services and service innovations; welfare and health; environment and energy; infrastructure and security; bio-competence and bio-society; information and communications; understanding and human interaction; materials; and the global economy. The panels identified over 80 areas of expertise, which could help Finland achieve breakthroughs in S&T and develop new innovations. In other words, the exercise produced a very extensive list of areas of expertise and specialisation. While this does not make prioritisation, choices and decision-making easier, it clearly shows the diversity of alternative lines of development and a situation in which the development of the economy and society requires varied and broad-based skills.

The joint FinnSight2015 project of the Academy of Finland and Tekes identified several development trends with many expectations relating to research. The findings can be summarised as:

- increasing productivity is crucial to Finnish competitiveness;
- global risk management will assume ever greater importance;
- energy and environment issues are of critical importance globally;
- increasing emphasis on basic human and social knowledge and skills will raise the quality of future-oriented development;
- know-how facilitating scientific and technological breakthroughs must be fostered.

The social impact of these trends cannot be easily identified. One reason is that the list mainly contains factors that can be linked to the word *relevance*. The reports of the Academy's research councils, however, focus less on relevance and more on scientific quality and the impact of research. An example of this dichotomy is demonstrated by the difference in possible approaches to funding health research. The Research Council for Health points out that the long-term social effects of medical sciences research can be associated with changes in health policy, health care and health behaviour. The Council for Health goes on to state that free project funding is probably the most effective way of raising scientific quality and the impact of research. This suggestion differs from that of the Research Council for Biosciences and Environment, which holds that promoting genuine demand for knowledge is of crucial importance. On the basis of an inquiry implemented by the Research Council for Biosciences and Environment, it draws the conclusion that the focus of the current debate, which considers the structures of knowledge production and know-how as the problem, is wrong. In reality, structures on the demand side for knowledge would seem to constitute the most significant structural problems.

Several questionnaires have been distributed in hopes of understanding the obstacles to innovation application and the channels of possible technology transfer. The results provide a reason for seeking improved methodological knowledge and better comprehension of significant relationships between basic research and its scientific and social impacts. One such questionnaire was addressed to the directors of environmental research projects. It asked about obstacles to the practical application of R&D results. In order of importance, these were: *i)* users do not understand the opportunities for application of the results (49% of answers); *ii)* lack of suitable actors to utilise the results; *iii)* legislative and political obstacles; *iv)* no opportunities to apply the results in sight; *v)* the development of applications would not be cost-effective; and *vi)* lack of willingness to use the results (5%).

A second report, which was produced by the Research Council for Natural Sciences and Engineering, lists the channels through which technology can be transferred for use in businesses. In order of importance these channels are: *i)* personal contacts; *ii)* research co-operation; *iii)* dissertations; *iv)* scientific publications; *v)* conferences, seminars and lecture events; *vi)* mobility; *vii)* technical reports, documents; *viii)* consultation and expert services; *ix)* licensing, sale of patents; and *x)* businesses specialising in the transfer of technology. The Research Council for Natural Sciences and Engineering also points out that the EUR 1 million of funding provided by the Academy generates, on average, 31 peer-reviewed articles in international journals, 4 doctoral degrees, 1.9 licentiate's degrees and 4.8 master's degrees.

The studies conducted to date only show that the issues of improved methodology for identifying social impacts and better comprehension of significant relationships between basic research and social impact should be approached from various perspectives, applying both qualitative and quantitative analytical methods. In response, Tekes and the Academy have recently launched development projects aimed at establishing a databank on the impacts of RDI (including systematic collection of information), creating a fixed set of impact indicators, and introducing more detailed procedures for carrying out evaluation and impact assessments. Furthermore, in the final reports on Academy-funded projects, concrete attention is now paid to collecting data on the impacts of research.

R&D and innovation and the impact of funding: Tekes' point of view

Tekes' funding is targeted at the R&D of businesses and public research organisations. Its goal is to encourage customers to engage in more R&D than they would otherwise (input additionality). According to recent studies, EUR 1 of Tekes funding increases private R&D investment by EUR 0.4-1.0, for a total increase of EUR 1.4–2.0. Public support seems to increase companies' R&D and innovation activity both nominally and in real terms. Approximately 60% of significant innovations developed in Finland in 1985–2000 received Tekes funding. In Tekes' technology programmes, the product development projects of small companies emphasise the importance of co-operation. In particular, low-technology sectors seem to benefit from the programmes. This underscores the importance of policies that support networking, even in such fields.

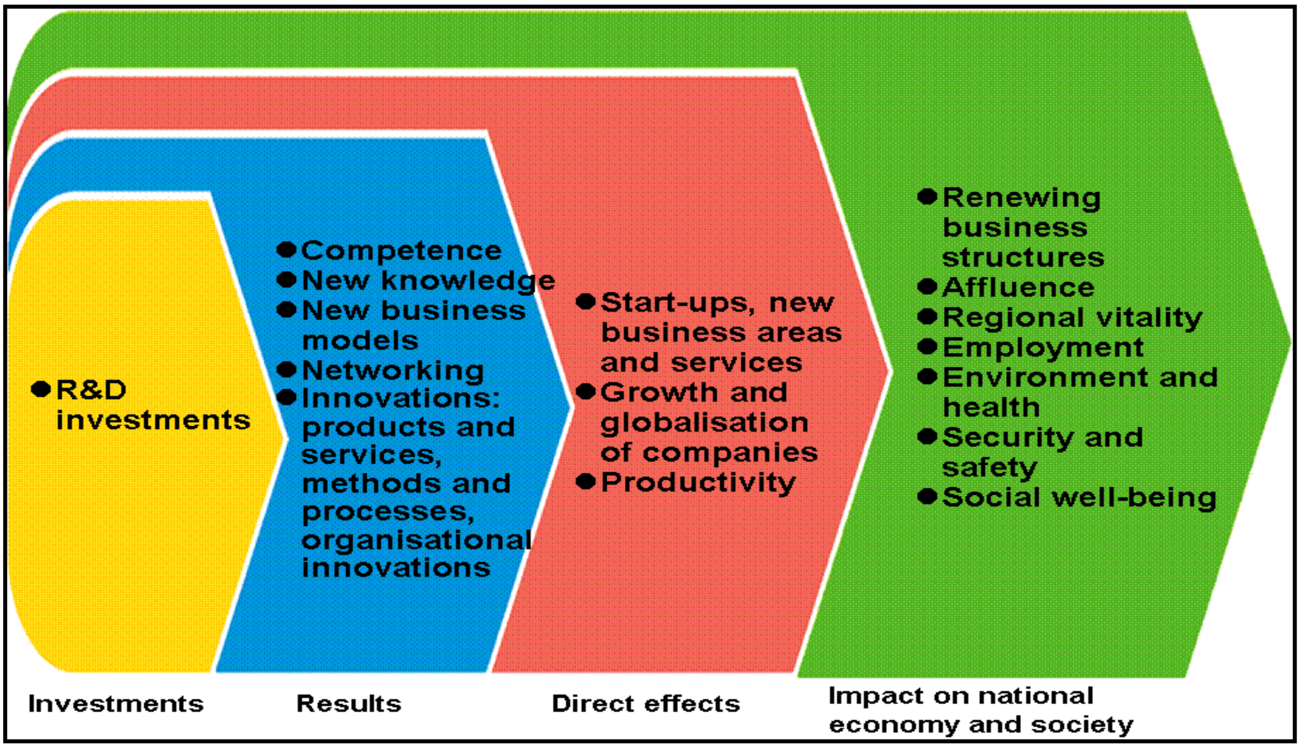
The assessment of Tekes' impacts, including the generation of an impacts database, is a continuous process. Impact assessment is implemented by means of project follow-up evaluations and programme and operation evaluations. For projects, impact information is collected in advance, during the final project stages, and three years after the project has ended. By the end of the three year period a much more comprehensive picture of the results is available. Various parameters form a basis for an overall view of the impacts.

The model Tekes uses for impact analysis describes various effects of the RDI projects based on cause-and-effect relationships. This model is based on the notion that R&D and the development of innovations generate results with immediate effects, which affect the national economy and society in a wider sense. The four-phase model (Figure 6.3) divides funding and results into separate phases, which together form a general view of the impacts of research and innovation activities. The approaches used in the impact model include the concepts of additionality, knowledge spillovers and new growth theory.

The first phase in the model is RDI investments. Tekes gives its stakeholders incentives (through funding) to increase their investment in RDI. This first phase may also be termed an *input* additionality level. The second phase describes the results and behavioural changes that take place within firms, universities and research institutes as a result of public RDI funding and other activities of Tekes. The third phase concerns the renewal of business (*i.e.* whether public RDI funding increases company *outputs* and in turn improves their overall performance). This phase also involves industrial dynamics, such as the utilisation of information technology to enhance productivity and the efficiency of public services. The fourth and final phase analyses the impacts of RDI financing on the economy and society as a whole.

Tekes' recent evaluations (see references) show that R&D and innovation activities as well as their public funding are of great importance to the national economy and welfare. Public support significantly increases the competence and networking of enterprises, generating patents, publications, products, processes and services. Know-how and new modes of action are demonstrated through increased company turnover and productivity, and a higher employment rate.

Figure 6.3. Tekes' impact model: four-phase approach to effects and impacts



Source: *Tekes Technology Review* (2007) 203: 3.

Evaluations show that Tekes' funding has a very positive impact on the company's own R&D investment. This has been one of the key justifications for increasing public funding for RDI, as it does not seem to crowd out or replace private investments. Results have been similar elsewhere. Input additionality has been EUR 1.4–1.9, depending on the country and period in question.

Moreover, evaluations show that Tekes has an impact on the output, commercial results and productivity of business and research organisations. Its funding seems to boost growth, patenting activity and demand for labour in companies, while reducing the probability of business closures and mergers. Tekes' support has enhanced productivity growth and improved the employment rate in companies with R&D staff. The impact on productivity is highest in SMEs and companies operating close to the productivity frontier. The added value of public support is to facilitate implementation of projects and influence how they are executed. Funding has also increased co-operation and interaction between key stakeholders of the innovation system.

In addition to the above financial impacts, companies receiving funding have also often changed their modes of operation and their behaviour (behavioural additionality). These changes have strengthened the importance of education and research and the capability to use new information and competence. Tekes' support of modes of operation and competence (Table 6.1) has had the following impacts: *i)* funding has positively affected innovation activities of companies; *ii)* funding has encouraged companies to build closer partnerships and has promoted the establishment of co-operation networks; *iii)* funding has spurred industry to co-operate with several service companies of various kinds; *iv)* funding has raised the risk-taking level of projects and productivity and raised the level of competence in companies; and funding has affected long-term business strategies.

Table 6.1. The impacts of funding on companies receiving support from Tekes[1]

	Agree	Disagree
Enhanced the **know-how of personnel**	90%	10%
Enabled projects with **more risk and return**	86%	14%
Improved the **credibility** of the company	84%	16%
Helped the company improve sales	77%	23%
Helped build **partnership networks**	68%	32%
Improved the **profitability** of the company	64%	36%
Also aided other projects by improving the quality of **equipment**	63%	37%
Helped **internationalisation** of the company	**60%**	40%
Over time had impacts on **business strategy**	**56%**	44%
Increased the **publicity** of the company	**55%**	45%
Helped to find **potential investors**	14%	86%
Influenced the **location** of R&D facilities	10%	90%
Influenced the **location** of production facilities	7%	93%

1. Figures based on a 2003 Tekes survey on behavioural additionality, including company interviews (survey replies from 200 companies, of which 80 are Tekes customers). "Cannot say" option and blanks are not included in these figures.

Evaluations further show that programme activities have had positive effects. In particular, public co-financed programmes have played an important role in the creation of common strategies and visions. They have enhanced R&D activities in the following ways: *i)* the activities are related to a new area of research; *ii)* the research is justified by benefits to the national economy and business; *iii)* the research theme is important from the viewpoint of political decision making; *iv)* the research is related to problems that are interesting from the viewpoint of enterprises; *v)* the combination of the information generated by different research communities produces something new and significant. Research usually brings added value in situations where there is a genuine will to merge the objectives of various organisations. According to studies, increasing the public resources of a single organisation in the name of co-operation, but without seeking more extensive common goals and consolidation of the different objectives of partner organisations, does not generate added value to the same extent.

The more comprehensive social effects of Tekes' funding are relayed through various channels, as part of the dissemination of expertise and technology. For instance, Tekes' projects have had a positive impact on the turnover of enterprises that are not direct Tekes' customers. They have also had a favourable impact on productivity and number of jobs in other enterprises. The impact of programmes is greater than that of separate projects because information and know-how are transferred more effectively through programme activities and services.

The economic and social impact of RDI

The broader connections between R&D, productivity and economic growth have been analysed by VATT and ETLA, and partly financed by Tekes. They have studied the impact of R&D funding in companies and extended the scope of the impact analyses to cover various sectors and the national economy. In terms of average productivity and gross national product per capita, Finland is close to the OECD average. In comparison to the EU states, Finland's income level and average productivity are above average.

Moreover, in Finland, productivity is among the highest in the world in key industrial export-based sectors. To strengthen productivity at the level of the national economy, however, the productivity of traditionally low-productivity sectors also needs to be raised. Consequently, productivity in these sectors has risen more rapidly in Finland than in other countries, even in the services sector. To promote productivity growth beyond the high performers, special attention must be paid to the comprehensive improvement of competence and the development of technology and conditions for its utilisation.

In addition to technological development, increased productivity requires organisational, administrative and structural reforms. These added factors also enhance the cost-effectiveness of projects and economies of scale. Productivity increases when companies introduce better technology and organise work more efficiently. In the long run, innovations are a key source of productivity and technological development. As far as the development of technology is concerned, the key observation is that the decisive factor is not the use of the newest technology, but rather the most productive. Results also indicate that policies prioritising the generation of new technologies do not always have an optimal outcome.

Recent reviews emphasise the simultaneity of the different phases of the innovation process and the fact that enterprises outsource some of their R&D. In other words, *open innovation* is gaining more ground. Another phenomenon often grouped with open innovation and which has become growingly more recognised in STI policy is *user innovations.* In user innovations, the process involves end users improving products and inventing new uses for them. In this way, the know-how of end users becomes part of the development process and the resulting innovations may become more widespread within society. These end-user innovations draw attention to the failures of certain innovations that do not take the needs of consumers and contexts sufficiently into account. Integrating this aspect of R&D activity into policy measures requires a new kind of thinking and interaction.

Around ten out of more than 300 companies studied as part of the Innovative Growth Companies project were particularly rapidly growing "gazelles" (growth of turnover over 30% during each accounting period in 2001–04). These companies are of major financial importance. In spite of their small number, they were capable of compensating, almost by themselves, for the reduced turnover in the other company groups. Policy measures are particularly necessary for identifying and encouraging growth companies. Moreover, studies show that new companies apply new technologies intensively and their productivity grows faster than that of old ones. Information and communications technology increases productivity, and it has been observed that the impact is greatest in new companies. In other words, the establishment of new, growing companies is important from the point of view both of the national economy and employment.

In research-based enterprises, innovations are based on the latest R&D from universities and research institutes. Typically, both the companies and the innovations take a long time to become profitable. A report commissioned by Sitra on the commercialisation of university research considers it a key problem that, even today, Finland does not have a clear policy on the role universities should play in the commercialisation process.[44] The resources and expertise of R&D service units have increased, but the challenge of clarifying their role in commercialisation activities and developing their management practices remains. Business operations are difficult to steer without clearly specified objectives. In addition, the evaluation of such activity is difficult.

According to studies, the practical implementation and commercialisation of R&D results are not as successful as they could be. In many reports, two factors emerge as obstacles. The first is related to the initial phases of the process: domestic research produces promising early-stage business ideas, but there are problems for identifying them and moving them ahead. The second factor is the lack of financing at the stage at which capital would be needed for long-term R&D and in later stages of product development, when the process is on the verge of production. Accordingly, it would be important to use actual needs as the foundation for the ongoing development of public innovation services. Even the current service structure fails to optimally support the use of the results of R&D and university-company co-operation.

The differences in the innovation activities of various sectors are significant. In certain sectors, new innovations are created every few months, whereas in others product cycles are significantly longer. Since service innovations have not been widely studied, their characteristics, impact and operative mechanisms remain largely unknown. For instance, the impact of service sector R&D is limited. Innovations in information-intensive sectors are often services containing a great deal of silent knowledge. Service concepts, distribution channels and organisational models are typical forms of service innovations that contain a great deal of silent knowledge.

The public sector, on the other hand, has not traditionally been considered innovative. Its public image regarding social innovations is, however, quite different. Public-sector innovations often have a legislative background or are dictated by social needs and developed in collaboration with the private sector. All in all, the measurement and assessment of the impact and productivity of services is difficult. It would be useful to develop comprehensive, reliable methods and improve the accumulation of data for that purpose.

In addition to investments in material and R&D, innovations – and their creation and successful commercialisation – depend on immaterial investments and capital related to a competent workforce, patents, customer relationships, organisational structures, business models, design, digitised data and software. The social and economic importance of immaterial investments is largely unexplored because it is difficult to measure. The challenge of enhancing the quality of research and innovation and its impact is connected with the development and exploitation of competencies within these fields.

Impacts of ministries and research institutes

Increased attention has been paid to the evaluation of the effects and impacts of government R&D. One example is the JYVA project (Assessment of Social Impact in Public Research Organisations), led by Technical Research Centre of Finland (VTT) and completed in 2006, which examined such issues in five

44. Finnish universities produced, on average, 40 reported inventions per EUR 100 million invested in R&D. The corresponding figure in the United Kingdom was 28 and in the United States 87. Finnish universities were granted, on average, seven patents per EUR 100 million used for R&D; the corresponding figures in the United Kingdom and the United States were 15 and 11, respectively. In 2005, licences produced an average income of EUR 250 000 per EUR 100 million spent on R&D for universities. In the United Kingdom, the figure was EUR 1.6 million and in the United States EUR 7 million (in 2004). Finnish universities generated five spin-off companies per EUR 100 million invested in R&D, whereas the UK figure was six and the US figure only one. Taking into consideration all R&D-based companies established in Finland in 2000-05 (171 in all), the turnover of only one exceeded EUR 2 million in 2005. The growth of such companies is relatively slow.

organisations, the MTT Agrifood Research Finland, the Finnish Defence Forces Technical Research Centre, VTT and two polytechnics, Helia and Satakunta University of Applied Sciences. As a continuation of this study, four ministries [Ministry of Finance, Ministry of Agriculture and Forestry, Ministry of Trade and Industry MTI (currently Ministry of Employment and the Economy MEE) and Ministry of Social Affairs and Health] launched a joint project entitled VALO (Impact as Part of Results-based Management and Steering of Research Organisations) and implemented it in 2006-08. Among other issues, the project investigated how the social impact of research organisations might be assessed systematically. This will lay foundations for the development of the impact assessment of sectoral research to a level corresponding to the activities of finance organisations.

Impact of education

It is widely acknowledged that a high level of education is necessary to improve the quality and impact of research and innovation, as well as the favourable development of individuals and society as a whole. Relatively few detailed analytical studies, however, have been conducted on the role and impact of education, particularly with respect to research and innovation activity. VATT's research paper, "Finland 2025" (2005), points out that the rise in the education level of the population of active working age will continue almost automatically and the share of the highly educated will increase over the next 25 years. It is not evident, however, that this will in itself generate growth. This serves as a reminder that productivity is among the key means of securing economic growth.

Nevertheless, Sitra's macro-level analysis, *Economic Impact of Education* (2006), concludes that in many cases investments in education have an undeniably favourable effect on productivity. Still, the information on the primary mechanisms through which investments in education accelerate growth is insufficient. In any case, such investments have a positive effect on the labour market success of individual citizens, particularly their future income. From an individual Finnish citizen's point of view, a longer education is still a good investment. However, the salary benefits of those with an intermediate level of education over those who settle for a basic level of education appear to be negligible in terms of both real differences in salaries and of productivity rates derived from education.

In the context of education, it is also worth mentioning the important work done within the OECD on PISA (Programme for International Student Assessment). This effort has produced more solid measuring methodology and internationally comparable data on the knowledge and skills of (~15-year-old) students and the performance of member countries' education systems. The separate work on extending the study to cover the HEI level is under way. This extension provides us with a better perspective and a solid ground for studying the effects of education.

6. Foresight, its development and utilisation

Foresight brings new information and insight to the creation of research technology and business competency. Finland's position at the top of international STI comparisons requires the country to be independently capable of developing and implementing future prospects and experiments in order to open up new opportunities. Foresight brings different kinds of actors together and creates a common knowledge base. Furthermore, it supports strategy work and selective decision-making. The use of foresight has been helpful when deepening co-operation or launching new projects. The goal is to develop practices and enhance competencies and networking by means of common foresight.

In recent years, Finland has increased its foresight activities. Conscious efforts have been made to reduce the lack of coherence and systematic outlook in foresight activities. Key projects include the Government Foresight Network, MTI's Foresight Forum, and Sitra's National Foresight Network. The Committee for the Future of the Finnish Parliament also partially operates in this arena. The VTT has launched a special Technology Futures Forum. Moreover, several other organisations are implementing their own foresight projects.

FinnSight 2015, mentioned above, has also shown the need to promote foresight. These efforts are being promoted with the help of studies of development trends and so-called weak signals[45] (Figure 6.3, which Tekes' foresight project of 2006 helped to identify. The goal was to identify promising fields and new opportunities for Finland and recognise threats. More than 1 200 experts representing research and innovation companies and communities were involved. The respondents agreed almost unanimously that Finland's success factors were technology and its utilisation, high quality and comprehensive skills, and generation of innovations. Respondents mentioned that the export of methods and techniques and environment-friendly technologies could open up increasing opportunities. Furthermore, they found that international competition and co-operation were closely related. The responses emphasised the need for dynamism and courage.

Beyond the above factors, the participants also considered the utilisation of Finland's special status, competitive edge, its investments in business expertise and the promotion of entrepreneurship to be important. They also said attention should be paid to the creation of infrastructures and the reduction of bureaucracy. The factors they viewed as threats included the sustainability of the economy; demographic factors; developments in neighbouring regions; the purchase of research-based innovations by foreigners; climate change; and the division of society into haves and have-nots. Otherwise the exercise produced wide-ranging results, leaving room for different interpretations. No synthesis report was produced.

Tekes is continuing its foresight activities as part of an ongoing international benchmarking activity. This activity will map Finland's opportunities and threats with the help of foreign experts. Tekes as well as other Finnish funding and expert organisations have a host of international connections and co-operation partnerships in the area of foresight. For instance, Tekes and the Academy of Finland have a co-operation agreement concerning technology foresight and evaluation with NISTEP, which represents Japan's science and technology policy. Sitra, VTT and Helsinki University of Technology TKK have taken advantage of the co-operation opportunities offered by this agreement.

Another way that Finland is developing its foresight is the Government Foresight Network (first term 2004-07; the second term started in 2008). This network has been a co-operation platform consisting of representatives of Finnish ministries, which has co-ordinated foresight activities and acted as a discussion forum. Its aim has been to develop foresight competence, develop foresight at the regional level, and promote the exploitation of foresight results in decision-making. Official have recognised that development in this field should continue.

Ministries carry out the foresight work of the Government Foresight Network. In 2005, it compiled an extensive report on the foresight information produced by various administrative sectors, which then served as a basis for drafting the report of the ministries. The final report of the network considered the current status of foresight and its development. In all foresight practices the competencies and the connection with planning and strategy processes vary greatly from one ministry to another. Foresight has been implemented separately by each ministry, taking the ministry's own strategic needs into account. A foresight process that brings all administrative sectors together is necessary for the government's future report compilation process. The first foresight network fulfilled the task of compilation satisfactorily. In the future, closer commitment to networking by participants, sufficient resources for implementing

45 Weak signals are generally defined in the literature as information on the likelihood of events that have a low probability of occurrence, but have high uncertainty concerning their impact and the trends that can develop afterwards (Medonça *et al*., 2004). Weak signals are usually hard to perceive in the present, but will likely constitute a strong trend in the future (Limola and Kussi, 2006).

common projects, and possibly a separate support organisation/unit for co-ordinating and leading joint efforts will be needed.

Sitra developed a second type of foresight network. The purpose of Sitra's National Foresight Network has been to enhance interaction between experts and decision makers, and to promote the utilisation of foresight. This network has brought together key foresight actors and decision makers who use foresight in their organisations' strategic processes. Social challenges and opportunities have been identified as topics for consideration by decision makers and as the targets of further studies and innovation. The network surveys new trends and signals and organises workshops for those utilising foresight information. In 2007, for instance, the focus was on the utilisation of the results of recently implemented foresight projects. This was considered necessary because insufficient attention has been paid to utilisation of results. The same applies to the utilisation of evaluation results.

7. Conclusions and the rationale for foresight development measures

To date, evaluations of the direct impact of public financing of research and innovation on the operation of companies and research organisations have been mainly positive. They have also, of course, identified shortcomings and development needs. [46]

Despite positive evaluations of the social impacts of RDI funding organisations and of research and innovation activities, many reports fail to provide an organised and sufficiently detailed picture of project impacts. The reasons are well known: the multi-dimensional and indirect nature of the impacts; complex interaction with other factors in the environment; different exploitation mechanisms and dissemination channels; and delays in the emergence of the impacts. These factors have not yet been fully understood, not because of the nature of domestic funding and research organisations, but because impact assessment methodology must be improved everywhere. Furthermore, it is particularly challenging to generalise the findings of detailed case studies.

Evaluation and its development are further impeded by the one-time nature of the activity. Remedying the lack of resources and shortcomings in the information base require concrete financial investments and systematic, long-term development. The strengthening and systematisation of the information base is also a prerequisite for the development of evaluation, impact assessment and foresight in policy measures, and for legislative drafting and legislation.

Tekes has made major investments in impact assessment and in the evaluation of social impacts on its business customers. For research organisations, the challenges relating to impact assessment still exist. In addition to the direct financial impact on business performance, enterprises receiving funding have often changed their modes of operation and their behaviour in a manner that has strengthened the importance of education and research, and the capability to use new information and skills. Alongside the direct impact, the indirect impact raises the estimate of the overall impact of financing. In order to gain a better understanding of the indirect impact, the direct impact of support instruments and incentives should be studied in greater detail, along with the question of the added value generated and whether they produce the desired outcomes.

The evaluations conducted by the Academy indicate that assessing the impact of basic research is difficult because the impact is indirect, complex and unpredictable, and takes a long time to become apparent. In addition, the differences between the various fields of science are great in terms of the nature

46. On the basis of the material presented here, the STPC raised many of these issues in drawing its conclusions on the need to develop evaluation, impact assessment and foresight in Finland.

of research and the form and impact of the results and their dissemination mechanisms. In the future, more attention should be paid to the following areas:

- When should the assessment of the impact of basic research be made?
- What is the significance of the research itself in relation to the observed impact?
- Which parties benefit from funded research?
- By what mechanisms and processes are the impact generated?

In a more comprehensive perspective, the challenges associated with impact assessment and foresight activities include the following: *i)* the relatively limited utilisation of the results of impact assessment and foresight; ii) the lack of clarity in who is responsible for the development of such activities and the division of labour; *iii)* the insufficient and fragmented competence of certain actors. Shortcomings can also be observed in the development of methodology and the generation and strengthening of the information base. In many sectors, impact assessment and foresight are still new concepts and no established methods exist. It would also be worthwhile to pay attention to the means by which evaluation and foresight can be more closely integrated with other development measures and strategic decision-making. Furthermore, evaluations have chiefly been implemented separately within each organisation or policy sector. Thus, the focus has been on the evaluation of a single actor, programme, project, or field of science. Furthermore, impact assessment is often implemented too soon after completion of the activity. The need for increased horizontality is evident. Internationalisation also emerges as a clear challenge.

Foresight has advanced from the earlier focus on technology foresight towards a more comprehensive, social perspective that places the anticipated development of technologies and research into a wider context. Expertise in foresight, however, is somewhat limited in Finland, including among those commissioning the studies. Efforts have been made to improve this situation through network-based co-operation.

Collective, broad-based national foresight should be strengthened. Foresight activities remain fragmented. Various organisations undertake a great deal of foresight, which leads to overlapping actions. In fact, the volume of foresight actions at the national level is comparatively small. Projects implemented until now have not usually been based on broader national starting points or strategic needs.

Foresight is receiving ever more emphasis for many reasons, particularly related to horizontal innovation policy and its development. In the future, the ethical, social and societal aspects of the results and impact of research and innovation will be emphasised in a manner that gives a crucial role to foresight.

In the area of STI policy, in terms of evaluation and foresight in general and impact assessment in particular, the following fields emerge as broader development areas:

- supply of intellectual resources and the renewal of the knowledge base;
- improving the conditions for the growth of productivity and gross national product (GNP);
- internationalisation (associated *e.g.* with research, mobility and business activities);
- quality and productivity of services;
- entrepreneurship, establishment and growth of new companies (including availability of capital and risk funding; development of innovation services; foreign direct investment);
- commercialisation of R&D results and their transfer from research organisations into practical use.

Based on the RDI evaluations discussed here, it has become apparent that more resources should be directed to comprehensive measures designed to improve innovation dynamics, develop the assessment of effectiveness and productivity of public services, and help to include the user perspective in policy measures. In this context, it is also essential to better link impact assessment and foresight with decision making on STI policy issues.

In relation to evaluation and foresight, the STPC has pointed out the increasing need to remedy methodological shortcomings, to bring together fragmented skills scattered throughout Finland, to increase training and to strengthen the knowledge base. It is also necessary to allow access to, and enhance the application of, publicly funded data, statistics and registers. This challenge relates in particular to micro-level data. The ministries should earmark sufficient resources and expand opportunities for joint evaluation and foresight efforts across administrative boundaries. Furthermore, there should be more multilateral international projects on impact assessment and foresight.

In the future external experts will conduct an impact assessment of the entire Finnish innovation system to meet the needs of decision makers and policy planners for more, and more accurate, information on the functioning of the RDI environment and the efficiency of the current STI policy measures

In some ways, these needs are already being addressed. In late 2008, an evaluation of the Finnish NIS was launched. This work was carried out jointly by foreign and domestic experts and the results were published at the end of October 2009. The objectives of the evaluation have been: to assess how well major drivers of change are addressed in innovation policy; to identify ways of addressing the current and future challenges (faced by the NIS); to point out needs for institutional and policy adjustments and reforms, and; to draw conclusions and recommendations for the STI policy governance and steering. The evaluation will certainly facilitate public debate and serve as a fertile ground for all future discussions on the needs and ways to renew the innovation system, with a special view to fortify the international competitiveness of the Finnish STI environment.

REFERENCES

Tekes reports and reviews on the themes of R&D, innovation and impact assessment

Autio, E., K. Miikkulainen and I. Sihvola (2007), "Innovatiiviset kasvuyritykset", *Tekes Technology Review* 201/2007.

Berghäll, E., T. Junka and J. Kiander (2006), "T&K, tuottavuus ja taloudellinen kasvu", Government Institute for Economic Research VATT, Research Reports 121.

Medonça, S. *et al.* (2004), "Wild Cards, Weak Signals, and Organisational Improvisation", *Futures* 36.2 pp. 201-218.

Hjelt, M., M.-L. Niinikoski, M. Syrjänen, V. Valovirta and T. Törmälä (2006). "Julkisten tutkimustulosten kaupallinen hyödyntäminen", *Tekes Technology Review* 192/2006.

Hyvärinen, J, and A.-M. Rautiainen (eds.) (2006), "Innovaatiotoiminnalla kilpailukykyä ja kasvua: tutkimus- ja kehitystoiminnan vaikuttavuus yhteiskunnassa", *Tekes Technology Review* 188/2006.

Innovaatiotoiminta luo menestystä ja kasvua (2007), Tekes Report. Tekes, Helsinki.

Kanninen, S., P. Kutinlahti, T. Luukkonen, J. Oksanen and T. Lemola (2006). "Finnish national evaluation of EUREKA and COST", *Tekes Technology Review* 13/2006.

Limola, L. and O. Kussi (2006), "Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making", *Futures*, 38.8 pp. 908-24.

Rajahonka, M., S. Svahn, M. Tinnilä and M. Valtakari (2006), "Kohti verkostomaista liiketoimintaa. (2000-luvulla päättyneiden teknologiaohjelmien onnistuneisuus eritoten verkostomaisen liiketoiminnan kehittämisessä)", *Tekes Programme Report* 15/2006.

Rilla, N. and J. Saarinen (eds.) (2007), "Tutkimusmatka innovaatioihin", *Tekes Technology Review* 197/2007.

Rouvinen, P. (2007), "Yrityksen tutkimus- ja kehitystoiminnan ulkoisvaikutukset", *Tekes Technology Review* 209/2007.

Syrjänen, M., A. Pathan, S. Ahvenharju and M. Hjelt (2006), "Yhteistyöllä tietoa ja verkostoja. Innovaatiorahoittajien yhteisohjelmien arviointi", *Tekes Programme Report* 20/2006.

"Tekesin vaikuttavuus. Asiakaskyselyjen ja teknologiaohjelmien arviointien analysointi" (2007), *Tekes Technology Review* 203/2007.

"Tilastoanalyysi Tekesin vaikuttavuudesta" (2008), *Tekes Technology Review* 229/2008.

The Academy of Finland’s SIGHT 2006 project produced evaluations and assessments of the state, quality and impact of Finnish scientific research. Seven publications were prepared. Unfortunately, a synthesis report was not made. The publications form a collection of research results, data and views concerning research done in Finland.

Suomen Akatemian rahoittama luonnontieteiden ja tekniikan alojen tutkimus: Arviointi hankkeiden vaikuttavuuksista (2006), Publications of the Academy of Finland 6/2006.

Tutkimuksen vaikuttavuus biotieteiden ja ympäristön tutkimuksen aloilla (2006), Publications of the Academy of Finland 7/2006.

Strategisella rahoituksella vaikuttavampaa tutkimusta? Kolme esimerkkiä vaikutusten ja vaikuttavuuden arvioinnista terveyden tutkimuksen alalta (2006), Publications of the Academy of Finland 8/2006.

Kanninen, S. and T. Lemola (2006), *Methods for Evaluating the Impact of Basic Research Funding: an Analysis of Recent International Evaluation Activity*, Publications of the Academy of Finland 9/2006.

Suomen Akatemian tutkimusrahoituksen vaikuttavuus: arviointiraportti (2006), Publications of the Academy of Finland 11/2006.

Lehvo, A. and A. Nuutinen (2006), *Finnish science in international comparison*, Publications of the Academy of Finland 15/2006.

Civilisation cannot be imported – Researcher commentary on the impact of cultural and social research (2007), Publications of the Academy of Finland 3/2007.

Other relevant publications

Ali-Yrkkö, J. and C. Palmberg (eds.) (2006), “Suomen kilpailukyky t&k-toiminnan sijaintipaikkana”, *ETLA B* 218.

Asplund, R. and M. Maliranta (2006), “Koulutuksen taloudelliset vaikutukset”, *Sitra Reports* 60.

Berghäll, E. (2006), “R&D and productivity growth in Finnish ICT manufacturing”, Government Institute for Economic Research VATT, Discussion Papers 388.

Bibliometristen aineistojen käytettävyys yliopistojen julkaisujen laadun ja tuottavuuden arvioinnissa, Publications of the Ministry of Education 2007:2.

Ebersberger, B. (2005), “The impact of public R&D funding”, *VTT Publications* 588.

Finland’s response to the challenge of globalisation. Report by the secretariat of the Economic Council, Part 2 (2006), Prime Minister’s Office Publications 19/2006.

Globalisation challenges for Europe. Report by the Secretariat of the Economic Council, Part 1 (2006), Prime Minister’s Office Publications 18/2006. (The report comprises 13 articles by European top experts.)

Himanen, P. (2007), *Suomalainen unelma. Innovaatioraportti*. Teknologiateollisuuden 100-vuotissäätiö, Helsinki.

"Impact evaluation of research – from theory to practice" (12.10.2005), Material of the workshop (jointly organised by the Academy of Finland and Tekes) held in Helsinki on 12 October 2005, at *www.aka.fi/*.

Kankaala, K., P. Kutinlahti and T. Törmälä (2007), "Tutkimustulosten kaupallinen hyödyntäminen – kvantitatiivisia tuloksia", *Sitra Reports* 72.

Maliranta, M. and P. Ylä-Anttila (2007), "Kilpailu, innovaatio ja tuottavuus", *ETLA B* 228.

Pekkanen, J. and T. Riipinen with E. Ahola (2004), "Investing in innovation", Presentation at the meeting of the OECD Group on Behavioural Additionality, held in Manchester on 10.–11 May 2004.

Piekkola, H. (2005), "Public funding of R&D and growth. Firm-level evidence from Finland", *ETLA Discussion Papers* 996.

Piekkola, H. (2006), "Knowledge and innovation subsidies as engines for growth. The competitiveness of Finnish regions", *ETLA B* 216.

Piekkola, H. (2006), "Seutukuntien kilpailukyky. Osaamispääoma ja innovaatioiden tukeminen kasvun lähteenä", *ETLA B* 216.

B. Foresight

Publications on foresight are mainly produced in joint projects carried out by the central organisations of the national innovation system and by top experts in the field.

Building on innovation: Priorities for the Future. The focus areas in the Tekes strategy (2005), Tekes, Helsinki.

FinnSight 2015. The outlook for science, technology and society (2006), The Academy of Finland and Tekes, Helsinki. *www.finnsight2015.fi/*.

Hjerppe, R. and J. Honkatukia (eds.) (2005). *Suomi 2025 – kestävän kasvun haasteet*, Publications of the Government Institute for Economic Research 43.

"Hyvä huomen! Introduction of the Foresight Forum of the Ministry of Trade and Industry", www.ennakointifoorumi.fi/index.php?alasivu=etusivu.

Hämäläinen, T. (2006), *Kohti hyvinvoivaa ja kilpailukykyistä yhteiskuntaa: kansallisen ennakointiverkoston näkemyksiä Suomen tulevaisuudesta*, Sitra, Helsinki.

Signals 2006. Results of the foresight project.by Tekes, www.tekes.fi/TilastotJaVaikutukset/tulokset.html.

Ståhle, P. (ed.) (2007), "Five steps for Finland's future", *Tekes Technology Review* 202/2007.

Valovirta, V. and M. Hjelt (2005), "Onnistumista punniten, tulevaa luodaten", *Hallinnon Tutkimus* 2005: 3, 95–111.

CHAPTER 7. ISRAEL: EVALUATION OF THE MAGNETON PROGRAMME[47]

The third case study describes an ongoing evaluation of a programme in Israel. The programme reviewed in this case study, the Magneton programme, is meant to encourage and assist in the transfer of technologies from research bodies to industry. This case study shows how national and foreign evaluators can work together to carry out an effective programmatic evaluation.

1. General history

The Government of Israel, in order to encourage investments in industrial R&D, enacted the R&D law (1984), whose purpose is to allocate funding to industrial companies that perform high-technology R&D. The goal was to make it easier for Israeli companies to invest in R&D because R&D was considered the best way to promote competition and could give companies a relative advantage in international markets.

Under this law a company that obtains an R&D grant and successfully sells the products should return the grant money as a percentage of yearly sales up to the total of the grant. The money received from companies is recycled into the budget of the Chief Scientist in the Ministry of Industry, Trade and Labour in order to finance new R&D activities. If the projects funded do not reach the sales phase or the sales do not yield a return sufficient to return the full grant, there is no requirement to return the balance of the financing.

During the 1990s, it was understood that it was also necessary to invest in the development of technologies in order to lay the technological foundation for additional innovative developments. Israel lacks government research institutions that deal with applied research for industry. Therefore, it was recognised that this task should be performed in research institutions and in industry, preferably in co-operation. Accordingly, in 1994 the Magnet Programme was established within the Chief Scientist's bureau to deal with the advancement of the technological infrastructure of industry. Its objective is to strengthen industrial companies' capacity to draw from a vast and varied pool of research and technology, giving them the capability to develop innovative, high value-added products with significant export potential.

The programme provides financial support to "pre-competitive" R&D projects developed jointly by enterprises and research institutes organised as consortia and governed by "collaborative agreements" among the respective parties. The intellectual property rights derived from technologies developed by a consortium belong to the members that developed it; however, each other member receives a licence to use the technology for further development of its own products free of charge.

As of 2000, the activity was broadened to the Magneton programme and is the case study presented here. Its mission is to encourage and assist in the transfer of technologies from research bodies to industry through co-operation between an enterprise and a research group. The Magneton programme was created

47. Prepared by Dr. Daphna Getz, Mrs. Vered Segal and Mr. Ilan Peled, Ministry of Industry and Trade, Israel.

to make the achievements of Israel's scientific research more readily available to Israeli industry for the benefit of the Israeli market. The goal is to encourage a process that contributes to technological innovativeness, a primary engine of growth in Israel's industry and economy. The time frame allocated for projects within the framework of the Magneton programme is usually two years, during which a technological feasibility study is performed for the product/technology. Upon completion the firm can decide whether to move on to an independent full-scale product or industrial process development or to stop the project if it does not appear technologically feasible or lacks profitability.

The programme started with a pilot that lasted two years. The aim of the pilot was to test the programme's potential, characteristics and ability to enrich industry with advanced technologies. In order to study the main parameters, the pilot phase was constantly supervised and there was continuing dialogue with the research institutions, researchers and companies' representatives who took part in the pilot. Although the pilot included 25 projects, it was not large enough to carry out a statistical/qualitative evaluation because of the intentionally wide variety of the projects, which involved different types of companies (small, medium and large), different fields (communications, electronics, materials, biotechnology, etc.) and different types of research institutions (universities, medical centres, research institutions). The evaluation of the pilot found it successful. The programme's procedures were then updated, and it went into operation immediately following the completion of the pilot.

2. Intermediate conclusions

The programme's administration required the evaluation of the programme during the pilot period to ensure that the programme reached its main goal, that its definitions coincided with the needs of industry, and that the parameters for evaluating the projects reflected Israel's priorities.

The programme is not systematically evaluated each year, but since there are two rounds of submissions a year, data on past performance is made available to the selection committee as background material. Usually this data is qualitative rather than quantitative. In 2008, however, quantitative research began to be regularly included. The findings are presented here.

From 2000 to 2008, 280 requests for financing of projects were submitted, out of which 150 were approved. Currently, there are 24 active projects and 16 additional approved projects that have not yet begun.

In 2004, the Samuel Neaman Institute for Advanced Studies in Science and Technology[48] in the Technion conducted, on its own initiative and financing, a first effort to evaluate the Magneton projects by interviewing ten institute researchers and eight industry managers. The Neaman research team also conducted a comprehensive literature review of the main goals and methods in use for evaluating R&D proposals, models for evaluating the main programmes, and methods for evaluating indirect results. Also, the institute compiled information on programmes similar to Magneton in other countries. Based on the interviews and the literature review, research questionnaires were constructed for the research institutes and industry given to participants in 50 projects (most projects examined in this preliminary research were performed while the programme was in its pilot phase). The survey findings, based on answers from 34 institute researchers and 20 industry participants, were summarised in a report and sent to Mr. Ilan

48. The Samuel Neaman Institute for Advanced Studies in Science and Technology is an institute for national policy research with emphasis on technology and science, industry and economics, higher education and quality of the environment. It is currently considered the largest and leading research centre in Israel on economic and social policy regarding science and technology. The Neaman Institute is an independent non0profit organisation working within the framework of the endowment made by Samuel Neaman for the advancement of technological and economic research of Israel.

Peled, the manager of the Magneton programme.[49] . Based on this preliminary research and on the results obtained by the research team at the Neaman Institute, the manager of the programme decided to order an evaluation of the Magneton programme once again in 2008 with financing from both bodies.

3. The goals of the evaluation of the Magneton programme

A number of meetings were conducted between the Neaman evaluation research team and Mr. Peled to discuss the goals of the 2008 evaluation and its scope. Based on discussions, it was decided to focus on the project's results, outcomes, and impacts. The goals discussed concern the main outcomes of research-industry co-operation in the projects performed and their influence:

- Construction of a database of all of the programme's projects to date.
- Identification of the factors influencing the success of the programme and the projects from the point of view of the characteristics of the participants, the characteristics of the projects, and the working processes of the projects' partners.
- Discussion of further issues that can influence the success of the programme. For example, would the projects have reached industry without participation in the programme? What are efficient ways to market the programme? How can these factors be identified at the stage of examination and approval of the projects?
- Formulation of conclusions and recommendations on the basis of the research findings to the manager of the programme to help decision makers establish the programme's policies regarding future projects.

The conclusions of the evaluation aim at policy makers and Magneton decision makers. It can supply them with extensive information on the efficiency and effectiveness of the programme's efforts to transfer technology from research bodies to industry. This information can then help in the management of the programme and the affiliated projects. Information on the characteristics of projects and participants with greater chances of success can help in the project selection process and increase the programme's effectiveness and success. Information on work processes and elements of the programme's management policies, which contribute to its success, can also help establish policy on links between research bodies and industry during the project.

When planning the current evaluation of the Magneton programme, the Neaman Institute's team considered other parties interested in the programme, such as: the Chief Scientist and R&D policy makers in the Ministry of Industry, Trade and Labour; the project selection team; the team of examiners from the Ministry of Industry, Trade and Labour;[50] the technology transfer offices (TTO);[51] and company managers, researchers, and project heads who participated in Magneton projects and even those whose projects were rejected.

---

49. Mr. Ilan Peled, the manager of the Magnet programme and the client of the evaluation project is a partner in the evaluation of the Magneton programme. Mr. Peled initiated the Magneton programme and took part in defining its objectives and directives. He is also responsible for implementation of the programme's policy and making policy adjustments and alterations throughout the programme.

50. The team of examiners accompany the projects from the examination phase, examine their compliance with milestones, attempt to assist in problem solving, and can authorise an additional year for projects on the basis of their advancement.

51 The TTO represents the research bodies and are responsible for preparing the agreements between them and industry, for signing the agreements, for management and for legal issues.

Those stakeholders do not take part in the present evaluation programme and no effort was made to obtain information from them. In the future, a broader scope may be considered in order to collect and exchange information with such interested parties as a way of deepening understanding of the findings and the larger picture.

4. Evaluation methodology

The Neaman Institute research team submitted to the Magneton programme manager a detailed proposal for the evaluation, which included the research goals, the indicators to be examined as part of the research, the research population, and the detailed research plan, timetable and budget.[52] The proposal was updated based on feedback from the programme manager. The project duration was set at 12 months and is being carried out according to the plan agreed between the two bodies.

In choosing the methodology for the evaluation, the research team took into consideration the fact that the Magneton programme is very new (2003-05). They also knew that the short time since the pilot phase makes it difficult to evaluate the long-term economic effects of the programme through measurement and analysis of economic parameters.

The research team decided to survey project participants because of the recent inception of the programme. The advantage of a survey is that it can obtain information on the programme and participants in its early phases. Other reasons motivating the choice of a survey as an evaluation tool were:

- It is the quickest and most readily available method of reaching all programme participants (an Internet programme was used which helps construct and distribute the research questionnaires and collect answers).
- The survey makes it possible to build a database on the projects, which can be used in the future to continue and broaden research on programme evaluation in different ways and using different methods.
- The survey supplies aggregative information, which can be analysed using statistical tools that enable examination of the connections between the different variables.
- The survey obtains both subjective information supplied by project participants on their background, the projects' characteristics, their rate of success and their opinion of the programme, as well as objective information regarding the outcomes of the programme.

In the future, it will be possible to adopt additional evaluation methods, such as economic analysis of the programme's influence and construction of an analytical model for understanding different aspects of the programme.

The evaluation process is not meant to remain an isolated event. It will include additional interested parties and broaden the evaluation to areas and aspects not included in the present phase of the project and use additional evaluation methods, which will expose other aspects of the programme.

The present evaluation focuses on the effects of participation in the Magneton programme and the characteristics of the projects (participants, fields, innovativeness, uncertainty, co-operation, etc.), which can be tied to measures of success and the project's influence.

52 The research team of the Neaman Institute is responsible for constructing and carrying out the evaluation programme, constructing the database on the Magneton programme's projects, and submitting the summary report to the project manager. The team performs the evaluation independently and is responsible for the analysis of the research results and the conclusions drawn.

The main indicators measured in the research questionnaire are:

1. Measurements of project success:

   Products resulting from, or expected to result from, the project (articles, patents, development of new product lines in enterprise, entrance of the company into new markets, etc.), direct and indirect benefits (spillovers) of the project, and the rate of success of the project according to different criteria (ability to continue to develop independently the product/technologies in the companies, realisation or expectation of realisation of the technology in the industry, broadening of employment opportunities in the company, etc.).

2. Characteristics of the projects and participants that will influence the success of the project:

- background data on programme participants: education, years of experience, function, previous experience in academic–industry co-operation, and previous participation in R&D projects;
- project characteristics: state of advancement of the knowledge prior to beginning the project, relevant infrastructure for the project such as laboratory, equipment, type of knowledge developed, rate of innovativeness, technological classification, level of uncertainty of the project;
- characteristics at the start of the project: type of connection between partners prior to the project, expectations and calculations regarding entering the project;
- characteristics of the co-operation: the pattern of the co-operation throughout the project, location of the work performed, factors contributing to co-operation, continuation of co-operation after the end of the project, etc.

The indicators are measured on a scale of 1 to 5, where 1 indicates low contribution or influence, and 5 indicates high contribution or influence.

The research questionnaire is mostly composed of multiple-choice questions. Two different questionnaires were prepared, one for researchers and one for heads of industry projects. Many of the questions were included in both questionnaires in order to examine the partners' different approaches to certain subjects. Short versions of both questionnaires were also prepared (in which the general questions were omitted) for participants who took part in more than one Magneton project. These short questionnaires were to be filled in for each additional project in which they participated. The research questionnaires were sent to the programme manager for feedback and subsequently modified on the basis of his remarks.

Most of the questions were subjective. The objective questions concerned the products resulting from the project, subsequent decisions, etc. Respondents were asked to choose what they considered the most suitable responses on the basis of a number of alternatives, or to rate the strength of the factors or characteristics presented. Opinio software was used to prepare and conduct the surveys and to follow up respondents in real time, to analyse and present data and outcomes, and to export data into statistical software for detailed analysis. After collection the data were transferred to statistical software.

A database based on the data gathered will now be compiled. Before analysing the data, their internal reliability will be examined by using Cronbach's α. The research findings will be presented using descriptive and inferential statistics, which will include the testing of the research assumptions by relevant statistical examinations. The descriptive analysis will present the distribution of replies in chosen variables through analysis of frequencies, charts, single dimensional tables and bi-dimensional tables (cross tabulations). The purpose of the inferential analysis is to identify the connections between key variables and to find differences between the two research populations (academic and industrial). The inferential analysis uses tests and models such as t-tests, linear and logistic regression models, statistical correlations, chi-square tests, as well as a-parametrical tests and models such as Mann-Whitney U, Kruskal-Wallis H test and the Spearman correlation coefficient.

The research team in the Neaman Institute and the manager of the Magnet programme were in contact and co-operated closely during the performance of the evaluation project. At the time of writing, the main phases in the research for evaluation of the Magneton programme could be summarised as:

- In 2004, the Neaman Institute carried out preliminary research for the evaluation of the Magneton Programme and the results were presented to the programme manger. This preliminary research formed the basis for the construction of the present evaluation programme.

- Several meetings were held between the research team from the Neaman Institute and the Magneton programme manager in order to agree on the targets of the evaluation. On that basis, the Neaman Institute formulated an offer for an evaluation programme, the scope of the evaluation, expectations, research population, research tools, budget milestones, etc. The programme was authorised by the manager of Magnet.

- 13 interviews were conducted with participants in Magneton projects who had finished their activities (six interviews with academic researchers and seven industry project managers) to learn whether changes had occurred in the programme after the preliminary research had been performed and in order to test the clarity of the questions in the research questionnaire and their suitability to the issues to be covered in the current project.

- A database of primary data on participants in the projects of the Magneton programme was constructed.

- The research questionnaire was updated and adjusted to the present evaluation goals, based on the interviews and feedback from the manager of the programme.

- The questionnaire was uploaded to Internet software and the e-mail addresses of the research population were updated in order to be able to conduct the survey.

- The research questionnaire was sent to the participants, followed by mail and phone reminders to ensure a high percentage of returns. The research team called all the participants who did not return the questionnaire and asked each of them personally to do so. For some participants the Neaman Institute team filled in the questionnaire during a telephone interview with the participant. Data was subsequently analysed and entered into the database.

The Neaman Institute team analysed survey results and wrote a summary report. It was presented to the Magneton programme manager and then published respecting the anonymity of the participants.

CHAPTER 8. NORWAY: EVALUATION OF RESEARCH FIELDS/DISCIPLINES[53]

> This case study presents specific issues pertaining to evaluation. One of several specific uses of evaluations is to assess the state of disciplinary research. The Norwegian case study focuses on the evaluation of research at the university level. It discusses the entire scope of disciplinary review, including the institutional framework, the methods and the uses of this type of evaluation. Of note, one major use of evaluation in Norway is priority setting for both research institutions and national agencies.

1. Introduction

This case study was chosen because it represents the most systematic approach to evaluating research on a national level in Norway. It has become an important tool for the development of research in terms of organisation and for maintaining a more consistent focus on quality improvement. The evaluations of research fields/disciplines have been greatly encouraged by the Ministry of Education and Research and are mentioned in several political documents.

The evaluations focus on university research, but may also include research environments at university colleges or research institutes. They are used in priority setting in the research institutions and by the Research Council, and have produced important input for the development of new strategic measures (the Norwegian CoE scheme among others). The results are also useful to the Ministry of Education and Research and other ministries.

The following evaluations of research fields/disciplines have been undertaken: chemistry (1997); earth sciences (1998); biology, basic including biomedicine (2000); physics (2000); mathematics (2002); information and communication technology (ICT) (2002); linguistics (2002); political science (2002); medical and health (2004); pedagogy (2004); technology and engineering sciences (2004); Nordic languages and literature (2005); pharmaceutical research (2006); development research (2007); economic research (2007); and historical research (2008). The evaluations are conducted as a combination of quantitative analyses, bibliometry and other performance measures, along with qualitative peer reviews (an international expert committee). They are considered part of a continual learning process for all stakeholders involved in the development of basic research within fields/disciplines.

Over the years this approach to evaluation has been improved to make it more standardised and more efficient. Methods are continually improved along with the development of more sophisticated research indicators and improved focus of the terms of reference. Recently the Division adopted a five-year plan for the evaluation of research fields/disciplines for Science in the Research Council. The plan has been presented to the Ministry of Education and Research.

This case study describes only one of several evaluation activities undertaken by the Research Council of Norway. The overall evaluation "system" may be characterised by two points: *i)* evaluation of research quality (projects, scientific research fields/disciplines, programmes, institutions); and

53. Prepared by Gro E.M. Helgesen, Special Adviser, Research Council of Norway.

ii) evaluation of the quality of policies (instruments and schemes, economic impact, societal and other non-economic impact).

2. Environment

The research and experimental development (R&D) and evaluation environment in Norway is dynamic. This section describes the variable amount of Norwegian resources that may be used for R&D. It also describes the institutional framework of disciplinary evaluation, focusing most notably on the role of the Research Council.

Resources for R&D in Norway

Norwegian expenditure on R&D in nominal terms amounted to NOK 33.9 billion in 2006. This represented 1.57% of the gross domestic product (GDP). Norway spent NOK 6 410 per capita on R&D in 2005; the OECD average was NOK 5 770 the same year.

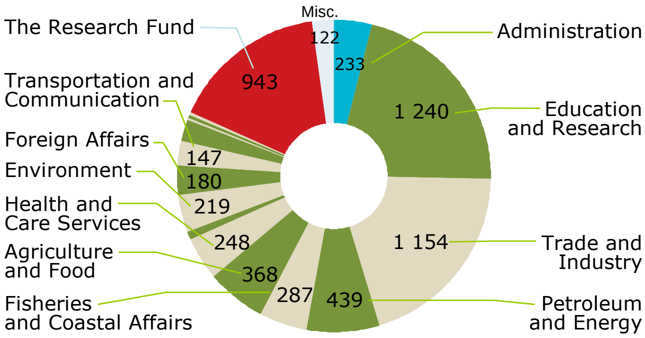
The higher education sector accounted for NOK 9.1 billion of R&D in 2005. The equivalents for the industrial sector and the institute sector were NOK 13.6 billion and NOK 6.9 billion, respectively. In 2005, industry funded NOK 13.2 billion of Norwegian R&D expenditure, while, NOK 12.9 billion by the government and NOK 3.5 billion from other sources and abroad. Approximately 30% of all public expenditure on R&D is channelled through the Research Council of Norway (Figure 8.1).

Figure 8.1. Public funding of R&D in Norway

The Research Council of Norway

The Research Council of Norway (the Council) is a public administrative body with special powers of authority and is organised under the auspices of the Ministry of Education and Research. The Research Council serves as a national strategic and executive body for research. In 2008, the Research Council's funding amounted to NOK 5 691 million (including administration) or about EUR 820 million (Figure 8.2).

Figure 8.2. Research Council funding from all ministries, NOK 56 991 million in 2008



The Council is responsible for improving the general knowledge base and for helping to meet society's research needs by promoting basic and applied research as well as innovation. The Council also promotes international research co-operation and serves as an advisory body to government authorities on matters concerning research policy. The research-oriented activities of the Council are organised into three divisions, with responsibility in the following areas:

- Academic research and disciplines (Division for Science).
- Innovation and user-initiated research (Division for Innovation).
- Strategic research in priority areas (Division for Strategic Priorities).

According to the Statutes of the Council, one of its main tasks is to "ensure the evaluation of Norwegian research activities". Among other more general evaluation tasks, the Council has performed evaluations of designated research fields/disciplines since 1996. The Division for Science has a particular responsibility for these evaluations.

Frameworks

Inspired by general evaluation trends and by experience with the evaluation of research disciplines in other countries, the Council has taken on the responsibility of assessing the quality and status of research disciplines on a national level, starting with the natural sciences. An increased focus on the role of these evaluations has been developed in co-operation with the Ministry of Research. They are referred to in research white papers, as well as in the annual budget allocation from the parliament.

Over the years the Council has developed a systematic approach to carrying out these evaluations, building on previous experience and expanding to all research fields. Recently a five-year plan was adopted by the Division for Science, within the Council, and presented to the ministry. National research statistics serve as a basis for collecting information on the target groups for specific evaluations. Research publication indicators are also available as databases. The Council's website provides information on the individual evaluations and reports.

3. Goals/strategy

Focusing on the evaluation of disciplines is a worthwhile endeavour. It can critically assess research quality in relation to both national and international standards. This section describes the purpose, scope, and role of disciplinary evaluations in Norway.

Purpose

The act of evaluating research in certain fields or disciplines may be summarised as a way of exploring the "health" of Norwegian research. The aim is to critically assess research quality in a given research field or discipline in relation to international standards. Evaluations assess a field or discipline's strengths and weaknesses, available infrastructure and organisational conditions, and training and recruitment procedures. The purpose is to:

- contribute to the dynamic development and improved quality of Norwegian research;
- provide insights on the strengths and weaknesses of the field in question for continued long-term development;
- expand and strengthen the knowledge base for the strategic and advisory tasks of the Research Council and other science actors.

The evaluations are not carried out for budgetary purposes, but to improve the quality and developmental potential of the field in question.

Scope

The evaluations mainly focus on fundamental or academic research carried out in universities. Sometimes research at "university colleges" and research institutes is included. This depends on the university colleges' or institutes' relative strength in contributing to the general development of the particular research field. The evaluations are carried out on a research group or institute level.

The research institutions are involved in the planning process to some extent and in the delimitation of the research being evaluated. Communicating the multiple goals to the researchers and institutions involved is important.

Role

The financing of public universities and university colleges in Norway is based on three components: basic funding (around 60%); a teaching component (around 25%) based on teaching credits/exams and exchange students; a research component (around 15%), only part of which is performance-based. Research institutes usually have less basic public funding and obtain most of their income from contract research. Starting in 2009, parts of the basic public funding for research institutes will be based on performance measures.

There is an ongoing debate on the financial structure of research funding, in particular on how to improve transparency in the allocation of basic resources and on how to separate teaching and research allocations on the basis of improved indicators and evaluations. As an increasing share of university budgets is allocated according to performance measures these evaluations may come to have an even broader utilisation, and even be used for the allocation of funds in the future.

The Ministry of Research and Higher Education has actively encouraged these evaluations. The evaluations are increasingly referred to in general policy debates on a national level because of the acceptance of these evaluations in both national strategy development and institutional development. Furthermore, the evaluations of research fields/disciplines are considered to be an essential part of the quality assurance of research in the higher education sector.

4. Planning

This section focuses on various periods in the evaluation processes, including: planning, implementation, and utilisation.

Budgets and resource allocation

While the Council takes on planning responsibility, the Research Board for Science, within the Council, sanctions crucial aspects of the process such as the scope of the evaluation, terms of reference, committee members, as well as the timeline for the whole process. The budget for each evaluation varies with the size of the field or discipline to be evaluated. In addition to the Council's human resources costs, the budget covers "mapping" the research population to be included, commissioning bibliometric studies, fees and travel expenses for a peer review committee and secretary, and travel expenses for the researchers involved. Another important cost element is the time spent by those evaluated in preparing the self-evaluations and presentations that constitute the basis for the evaluation.

Usually at least two years must be allowed from the start of planning until the report is presented.

Designing the evaluation

As mentioned, these evaluations are conducted as a combination of quantitative analyses, bibliometry and other performance measures, along with qualitative peer reviews by an international expert committee. This means that the design of each evaluation is more or less given. A handbook for the administrative planning process was developed to assist the officers involved, but differences in research cultures may still influence the approach. The collection of general background information is often commissioned to institutions working with research statistics and indicators. Most often the Institute for Scientific Information (ISI) database is used for bibliometric analyses.[54] Inputs also include a recently developed national system for field-specific publication indicators at universities and university colleges.

Every evaluation has its own particulars when it comes to: the delimitation of research groups to be included and the choice of methods and background information. Defining the population to be included in the evaluation is an important point of departure and involves meetings with research institutions and national organisations before the Council makes a final decision. These meetings involve discussions of the evaluation model, input for the terms of references, and input on appropriate statistics and indicators for the research field in question. The dialogue with relevant research institutions has proved important for a successful implementation and smooth operation of the evaluation process.

Terms of reference

The terms of reference for each evaluation are more or less standardised. They state the objectives and organisation of the evaluations and also contain the mandate for the evaluation committee. The committees should focus on: scientific quality, relevance, collaborations (national and international), research organisation, and scientific leadership and management.

---

[54] Among other services, the ISI database offers bibliographic database services and the Journal of Citation Reports. Both are useful tools for bibliometric analyses.

Methods and material

As stated above, the evaluations combine quantitative and qualitative approaches based on self-evaluations, bibliometric analyses, SWOT (strengths, weaknesses, opportunities, threats) analyses, and interviews.

Self-evaluation forms are part of the planning process and build on previous experience. They also address certain aspects of the research in question. The Council is responsible for collecting these forms. The Council also commissions the bibliometric analyses and schedules the meetings between the committee and the evaluated.

Site visits are very time-consuming for an evaluation committee. Unless crucial aspects concern infrastructure assessments, the meetings between the committee and research institutions are set up as "hearings". This involves setting a fixed time schedule for the committee to meet with representatives from the research groups and institutions being evaluated. It usually includes a standardised presentation in the shape of SWOT analyses and designated time for dialogue.

An instrument has also been developed for peer review. Based on pre-defined criteria it consists of a five-scale ranking system, using the categories excellent, very good, good, fair and weak. The instrument is applied in most evaluations.

A website is established for each evaluation to present information about the ongoing evaluation. This information may include time schedules, terms of reference, committee members, self-evaluation forms, etc. Transparency in the processes is a major goal.

5. Implementation

A general procedure for undertaking these evaluations has been developed. The key steps are: communicating with the research environment; appointing committee members; preparing an evaluation report and follow-up.

Usually the Council begins the evaluation process with a "start-up" meeting with representatives from the institutions. The meeting is convened to discuss the process, the purposes of the evaluation and follow-up plans. General and specific information about the evaluation is also presented on the Council's website.

Considerable work is devoted to the selection of peers. Possible committee member are evaluated based on broad scientific experience, high international standing in their research fields, and a high level of scientific integrity. Evaluation experience and experience in research organisation and policy making are also desirable qualifications. There is a sharp focus on possible conflicts of interest. The broader the research field, the more challenging the task of composing a well-qualified committee of workable size. Those evaluated and relevant institutions, including foreign research councils, are consulted and asked to suggest peers. There is close contact between the Council and the committee chair before the start of the evaluation to ensure that the purpose and the method of the evaluation are clearly communicated.

The work of the evaluation committee is based on the following elements:

- the "mapping" of the research groups/institutes and a presentation of the structural framework for that particular area of research;
- a general presentation by the Council of that particular research field in the national setting;
- research groups' self-report and self-evaluation (standardised);
- bibliometric analyses;

- meeting of the committee with those evaluated;
- committee’s overall assessment (peer review in light of international research).

The meeting of the committee with those evaluated is a central part of the evaluation. The committee invites representatives from the research groups being evaluated to meet at a set time, with a prescheduled agenda, including a short presentation and time for dialogue. Those being evaluated come from all over the country to the committee’s meeting place, and the hearings last from two to five days. This has been shown to be very efficient in terms of the committee’s time. Alternatively, site visits may be conducted, depending on the number of institutions involved and the relative importance of observing the home territory (laboratories, etc). It is the committee’s responsibility to conduct the meetings and interviews. Usually the Council is represented as an observer in these hearings/site visits, but the Council’s representative never attends the committee’s discussions concerning their assessment of the evaluated.

If the committee’s work is well planned and organised, it may have a draft report at the end of the daily meetings with the evaluated. The finalising of the report may then be carried out later through Internet contact. However, some committees prefer to have a meeting or two to discuss their final report. Usually the parts of the report that describe facts about those evaluated are sent to them and to the Council to check for corrections before the report is finalised.

Presenting the evaluation report

The evaluation report is the committee’s final product. It is made public in printed and electronic versions by the Council. Often the release is combined with a seminar or a press conference. The Council distributes the report to all stakeholders, often inviting specific comments on the committee’s findings. The Board of the Research Council has a formal discussion of the findings in a meeting. In most cases, the Council will have a meeting with representatives of those evaluated and stakeholders to discuss the findings. Those evaluated are also invited to give written comments on the final report and its findings as input to the follow-up process.

6. Utilisation of reports

The primary users of the evaluation findings are the research institutions that have been evaluated and the Council. Secondary users are ministries, in particular the Ministry of Research, and other research stakeholders, including the media.

In 2004, the Ministry of Education and Research commissioned a meta-evaluation of the evaluations that had been carried out to that date. Some general findings on Norwegian research quality gave input to policy discussions and also drew a positive conclusion about the evaluations. It also gave important input to further development of the processes involved in this kind of evaluation.

Follow-up processes

The Council initiates the follow-up of the evaluation. This usually involves appointing a working group in which major stakeholders are represented. This group comments on the findings and advice of the evaluation committee. Recommendations are classified according to possible actions and responsibilities, and adapted to a realistic follow-up plan of action.

While stakeholders discuss the plan, there is no enforcement of follow-up. The Council has increasingly been allocating funds and inviting the evaluated institutions to apply for so-called strategic project funding.

The research institutions conduct their own discussions as to the consequences of the evaluation findings. The ministries, in particular the Ministry of Education and Research, assess the report internally and use it as point of departure for policy discussions. The reports are frequently referenced in White Papers and other policy documents for higher education and research.

In line with general quality control procedures, the follow-up is also reported to the Office of the Auditor General in connection with its performance audits of Norwegian universities. The media also frequently covers findings from the evaluations, and this often leads to high-profile discussions of the quality of Norwegian research. Periodically, the Council requests reports on the consequences and actual follow-up from the institutions. Institutions generally report by either written communication or as part of a workshop. Evaluations have led to the reorganisation of institutions, to new collaborative initiatives, and to new financial schemes for the Council.

Consequences of undertaking evaluations of research fields/disciplines

Experience shows that acceptance of the findings is strongly related to how well the goal of the evaluation is communicated, and to the degree of involvement of those evaluated in the planning process. The focus should be more oriented to learning and development than to control, and this must also be communicated to the evaluation committee.

The findings from the evaluations constitute an important knowledge base for the ministries and for the Council, the importance of which usually exceeds the immediate distribution of follow-up grants. This information is used in strategy discussions in many contexts, including more general budget proposals from the Council as well as institutions. It is also used in discussions on improving policies and policy measures. One example is the development of the Norwegian CoE scheme. This scheme came about as a result of several evaluations on variable research quality, research groups below optimal size, too little international collaboration, etc. Institutions also use the results actively, in terms of organisational changes and the internal allocation of funds.

The meta-evaluation of these evaluations carried out in 2004 strongly supported the use of the evaluations, and also gave important input into the way they are currently implemented. Moreover, after each evaluation, the involved Council staff is encouraged to make a short report on the process for each evaluation. The aim is to improve the planning and methods of this type of evaluation in the future.

7. Conclusion

In conclusion, the evaluation of disciplines is strongly rooted in the Norwegian evaluation system. However, some improvements could be made to strengthen the system. Quality control of planning and implementation are very important to the Council in order to legitimise these evaluations. Ensuring quality is a continual process. Both the follow-up processes and improved utilisation of results are subject to continuous monitoring and improvement. Moreover, the overall evaluation system is split, with one branch focusing on projects and the other on policies. The five-year plan for evaluating research fields/disciplines should eventually be integrated into one of these branches. The Council should merge these branches into a more general evaluation strategy.

REFERENCES

Research Council of Norway (2007), *Report on Science & Technology Indicators for Norway*.

Brofoss, K.E. (2004), "En gjennomgang av Forskningsrådets fagevalueringer", NifuStep Arbeidsnotat 7/2004 (Norwegian only).

CHAPTER 9. JAPAN: FOLLOW-UP EVALUATION AS A TOOL FOR R&D POLICY [55]

This case study from Japan deals with specifics of the evaluation process. It examines the role of follow-up evaluations. The goal of such evaluations is to ascertain the long-term impacts of programmes, policies and projects. Also of note, evaluation in Japan makes use of both peer and expert review in regular evaluations and impact assessments in follow-up evaluations. The case study demonstrates the follow-up evaluation process through the practical example of lithium batteries.

1. Introduction

Japan's economy began to stagnate during the 1990s. One of the reasons was the mismatch between conventional technology development for industry and the emerging science-based industries, particularly in the fields of information technology and biotechnology.

In an effort to modernise Japan's science and technology (S&T) system, the national Science and Technology Basic Plan was promulgated, followed shortly thereafter by the issuance of the innovation policy. The 3rd Basic Plan (for 2006-10), promotes innovation policy, itemises priority fields and indicates the required budget.

The evaluation process reviews the entire research and development (R&D) phase. It seeks to streamline the budget process and reduce redundancy. In addition, the evaluation process allows timely adjustments and advances in innovation policy. Modern S&T, however, has become so specialised that decision makers for R&D find it difficult to fully evaluate projects. This makes well-established methodologies and standard evaluation procedures necessary.

The follow-up process identifies elements that accelerate, as well as impede, the innovation process. The two case studies of follow-up evaluation described here demonstrate the innovation processes. The results section indicates essential action plans to further promote R&D.

2. The evolution of evaluation in Japan

The history of evaluation in Japan's public entities

In 1995, the government enacted the Science and Technology Basic Law with a view to making Japan an S&T-based nation. Under this law, in 1996, the government established the 1st Basic Plan for Science and Technology. Its purpose was to attain higher standards in S&T and encourage comprehensive and systematic S&T policies, while contributing to the economic development of Japan. The plan began in fiscal year 1996 and ran to 2000. It also included projections for the following ten-year period. It has been revised every five years.

---

55. Report prepared by Yasukuni OKUBO, Director for Technology Evaluation, Technology Evaluation and Research Division, Ministry of Economy, Trade and Industry.

In 1997, the General Guidelines for the Evaluation of Research and Development were issued to enhance the efficiency of R&D. Soon after, ministries began to evaluate R&D projects on the basis of these guidelines. This was the beginning of public evaluation in Japan. The evaluation process was paired with a reform of the budget process from an internal decision process to a co-ordinated decision process involving the intervention of external experts. To date, the guidelines have been revised twice. The evaluation system has evolved over the past ten years and will continue to do so in line with the changing state of the science, technology and innovation (STI) world. Also in 1997, the Ministry of Economy, Trade and Industry's (METI) newly created Technology Evaluation and Research Division issued the Guidelines for Technology Evaluation that established the standard items and criteria for evaluation.

In January 2001, the central government was restructured. Ten ministries and agencies were consolidated into the Cabinet Office, while 12 ministries remained separate. In April 2001, the government established independent administrative institutions (IAI) in order to separate operational sectors from the central government. In 2004, 99 national universities were reorganised into the 89 IAI national universities. Prior to the restructuring, the number of central government staff was approximately 544 000. As a result of the restructuring, staff was reduced by approximately 25% (or 136 000 individuals).

The Cabinet Office was established in January 2001 as a new agency within the Cabinet of Japan. It is responsible for handling the Cabinet's day-to-day affairs. A sitting cabinet minister has traditionally filled the position of cabinet secretary. Also in 2001, the Cabinet Office established the Council for Science and Technology Policy (CSTP), under the leadership of the prime minister. The Council serves as the headquarters for the promotion of S&T policy. The CSTP meets every month and works to oversee, formulate and co-ordinate national S&T policy. The CSTP examines its compatibility with the Basic Plan and sets priorities for each ministerial research project. This crosscutting, inter-ministry, and top-down decision maker is responsible for organising a new science and technology administration system in Japan.

Japanese S&T administration: the role of METI

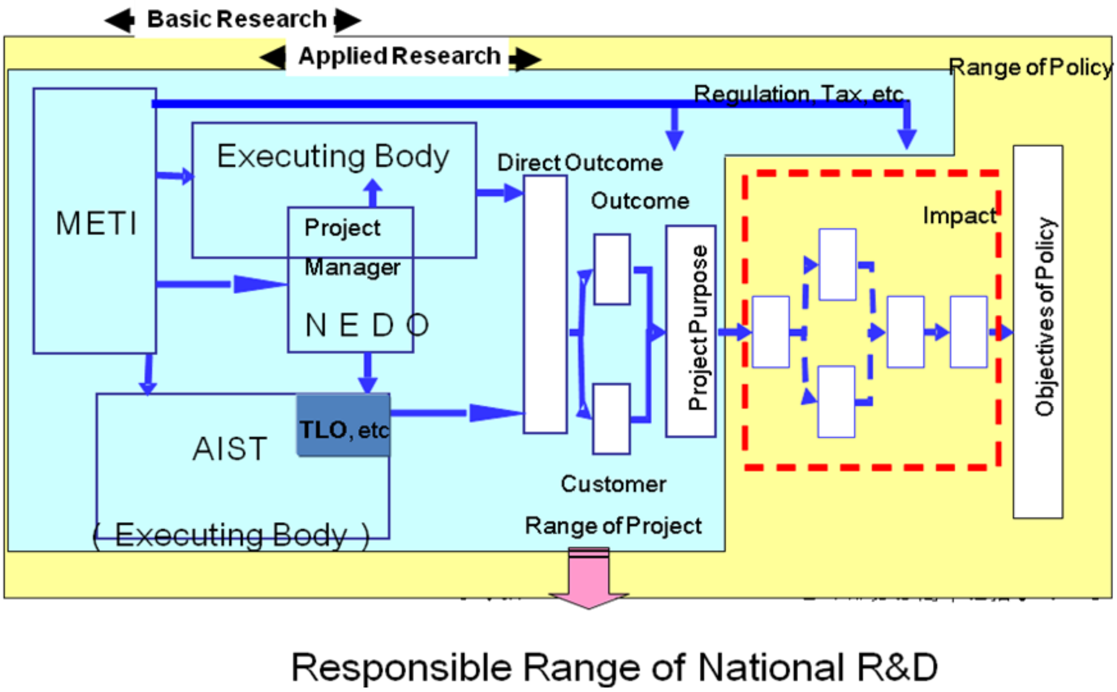
METI's mission is to secure Japan's sustainable economic growth, while ensuring stable growth of the international economy. METI seeks to achieve this goal by enhancing Japan's competitiveness, creating new markets and improving the socioeconomic infrastructure. METI covers six policy areas: *i)* manufacturing industry, economic and industrial policy; *ii)* foreign economic policy; *iii)* policy for information and communication and services; *iv)* policy for small and medium-sized enterprises and regional economy and industry; *v)* energy and environmental policy; and *vi)* nuclear and industrial safety policy. Under these policies, 34 programmes oversee approximately 120 R&D projects. These projects serve as the minimum unit of policy.

METI's R&D-related groups include: the New Energy and Industrial Technology Development Organization (NEDO), the National Institute of Advanced Industrial Science and Technology (AIST), private industry and universities. NEDO serves as a funding agency which co-operates with private bodies to conduct applied research. In some instances project managers from the executing body work directly with the NEDO office and its staff. This type of co-operation between groups has been shown to result in a strong sense of responsibility and accountability among all participants in Japan.

AIST is a research institute charged with conducting basic research; however, AIST currently extends its purview to resulting products as well. AIST has a technology licensing organisation and an innovation office that are responsible for ensuring that the research conducted at its facilities also contributes to society. METI endeavours to stimulate the international economy by promoting S&T through direct support to the private sector, NEDO and AIST. In addition, METI works with regulations, taxes, etc., to enhance the impacts to the innovation process.

METI's total budget for R&D is approximately JPY 500 billion a year. The budget is mainly allocated as non-competitive subsidies to NEDO and AIST. NEDO evaluates its own projects (which are subsidised by METI). AIST also evaluates its own performance. Excluding its non-competitive subsidies, METI mainly evaluates the competitive budget subsidies to industry. METI's Technology Evaluation and Research Division is responsible for evaluation of the R&D project and its budget and evaluates about 50 projects a year. The blue area of Figure 9.1 shows METI's range of responsibilities relating to national R&D.

Figure 9.1. Role of METI, AIST and NEDO in relation to Japan's R&D effort



Both METI and NEDO conduct follow-up evaluations, however, they differ in the range of their evaluations. NEDO's projects aim at short-term outcomes and reaching the project's goal. METI's follow-up evaluation takes into account a long-range outlook to achieve the policy objective. Hence, METI's evaluation incorporates NEDO's evaluations.

METI's evaluation process

The diffusion model of technology makes it possible to sketch the large range of outcomes from a project's outputs. In METI's current innovation policy, a key discussion point is whether and how R&D contributes to or affects various outcomes. The collection of evidence to identify and understand outcomes requires a significant effort. It has been a challenge for METI to evaluate the outcome during the project evaluation phase, the impacts on industry, and the cost-benefit ratio of R&D.

METI remits interim, *ex post* and follow-up evaluations to external committees; the *ex ante* evaluation is accomplished via self-evaluation. The two types of external evaluations assess projects in terms of output and outcome. The first consists of a peer review and the second an expert review. The peer-review evaluation team of four to five members reviews specific R&D projects. Meanwhile, the expert-review evaluation team of ten members reviews all R&D projects. Follow-up evaluation methodologies used by METI include: indicator evaluation, quantitative evaluation and interview/questionnaires. Follow-up evaluations are not carried out on all projects; rather, one or two projects are selected each year from the entire pool of projects completed five years earlier.

3. METI’s follow-up evaluation

The evaluation process mentioned above looks much like other evaluation systems described in this report. Japan, however, has gone beyond the use of evaluation for current impacts and expanded the process to account for impacts after the programme has ended. METI calls this “follow-up evaluation”.

Key role of follow-up evaluation

The first item in the follow-up evaluation process is impact analysis. Under METI’s guidelines, this analysis includes technology impact, improvement of R&D performance, economic benefit, social/life improvement and feedback to policy. These items reveal the extent of progress achieved by the project. The concept of the impact on industry is always opaque and controversial in the evaluation.

A common question of reviewers is the definition of “benefit”. It is defined in METI’s evaluation system as the cost-benefit ratio. It is a useful measure because it is quantitative and provides a numerical indicator.

The follow-up evaluation is conducted on projects concluded five years earlier. The evaluation takes a “hindsight” or “after-the-fact” viewpoint to identify long-term effects. Changes in points of view and circumstances following the project can shed further light on the earlier *ex post* evaluations and reveal the legitimacy of the follow-up actions. This is a way to learn about elements that accelerate and/or impede long-term outcomes.

METI collects and analyses external, internal and political factors affecting the flow of research via interviews of the executing bodies. Examples of external factors include advances in competing technologies and changes in the marketplace. Examples of internal factors include maturity of a technology, justification of technology, possibility of fusion with another technology, justification of strategy for industrialisation, etc.

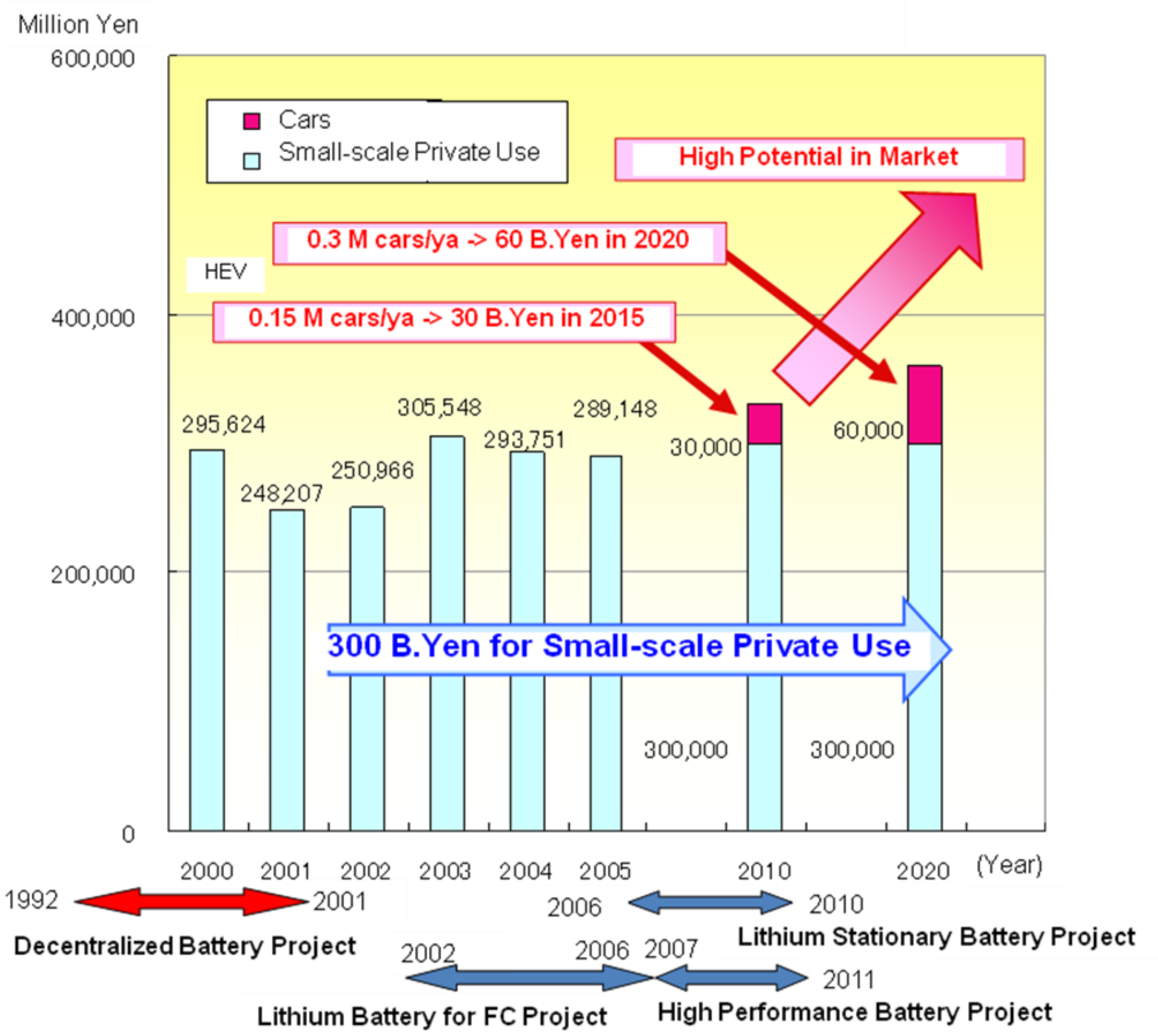
A vital role of the follow-up evaluation is to locate or identify negative elements and their linkages. Positive elements are easily identified because they advance or accelerate a project. On the other hand, negative elements inhibit the progress of technologies or practices and are usually difficult to identify. Since the innovation process may be terminated in such cases, little relevant evidence may be available. Linkages also represent a problem because missing linkages are hard to identify and can impede progress on technologies and practices. All these factors can help improve a R&D strategy and identify new R&D programmes, funding, and regulations.

METI case study of a follow-up evaluation

From 1992 through 2001, METI conducted a “decentralised battery project”. This project resulted in the launching of three new projects, one of which is an ongoing stationary battery project and the other two targeted automobiles.

Figure 9.2 examines the prospective lithium battery market. The market includes small-scale personal computers and digital cameras for which demand remains constant and stable over a span of 15 years and is worth approximately JPY 300 billion. The know-how generated from the initial project and subsequent projects was responsible for producing medium- to large-scale batteries. Currently, there is no demand for a lithium battery in automobiles, but demand should emerge within the next two years and grow over the next ten years.

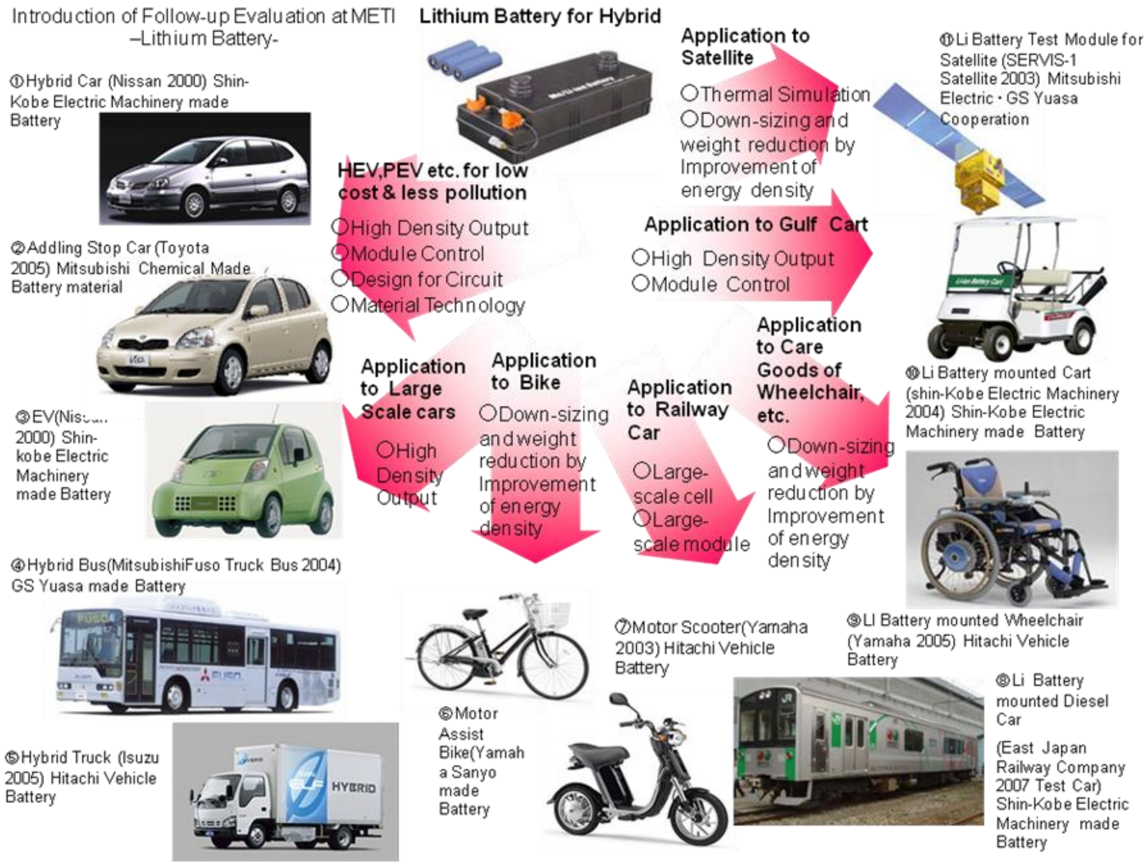
Figure 9.2. Prospective lithium battery market and subsequent national projects



The advantages of the lithium battery, compared to the lead-acid battery, include its high density storage capability, enhanced module control, design for circuits, material technology, facilitation of downsizing, reduced weight, large-scale cell and module production. The lithium battery has been applied to hybrid systems for transport, including the delivery system in hybrid automobiles, electric vehicles, hybrid buses and trucks, motor assisted bicycles, motor scooters, diesel cars, wheelchairs, golf carts, as well as a satellite test module (Figure 9.3). High-density storage helps to accelerate the development of a large-scale hybrid system.

A follow-up evaluation enabled METI to understand that the decline in the price and demand for electricity and regulation to maintain conventional use of electricity inhibited progress on the creation of new demand for battery storage systems. The stationary battery, however, has the potential to stabilise power from wind power generators. Seizing the opportunity, NEDO began a lithium stationary battery project in 2006 (scheduled to run through 2010). The project aims to develop a battery capable of generating electricity oft 40 000 kWh when combined with wind power generation. This is an example of policy intervention to assist the diffusion of technology by identifying impeding elements.

Figure 9.3. Diffusion of lithium battery for transport



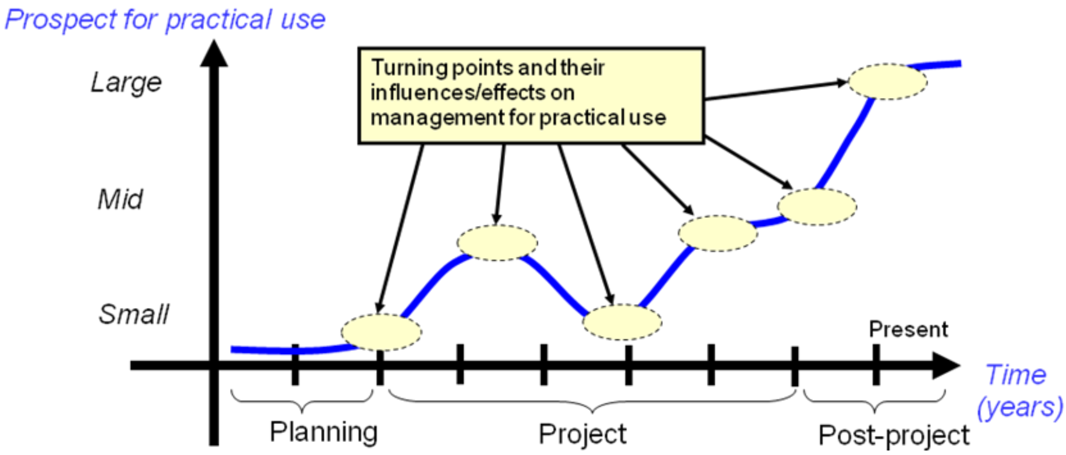
4. Follow-up evaluation at NEDO

Monitoring for follow-up evaluation

NEDO conducts follow-up monitoring to observe the direct outcome of its research. The monitoring of results ensures accountability to Japan's taxpayers, provides feedback for improving NEDO's R&D management, and contributes to the planning of future R&D strategies. NEDO monitors the activities of project participants annually, while collecting evidence related to accelerating and impeding elements. The collected evidence and the *ex post* evaluations allow NEDO to follow technology trends through post-project research activities.

NEDO creates a follow-up chart to visualise its findings (Figure 9.4). The follow-up chart is a tool that identifies chronologically significant events in the project management flow. The main objective is to show the possible "success" or "failure" scenarios to an executing body. NEDO collects evidence by interviewing the companies involved in the project. The interviewee is a person who understands the entire project and its evaluation results.

Figure 9.4. Follow-up chart



The horizontal axis of Figure 9.4 represents time, and the vertical axis signifies the views of the interviewees on the technology's prospects to have practical use. The figure reflects the change in the practical use of the technology over time (prospect), as well as the effects on management. To date, NEDO has conducted interviews of approximately 69 companies. Of these, 46 achieved commercialisation (*i.e.* proved successful), while 23 terminated their R&D on the basis of demonstrated failure.

Based on company interviews, NEDO defines four categories of results. Of the 46 success stories, 33 are categorised as Type A, in which the executing body steadily overcame technical problems with a clear vision of social/user needs. The remaining 13 were categorised as Type B (Figure 9.5), in which the view of technical problems during the early stages had been overly optimistic but R&D breakthrough did occur during the latter half of project.

Of the 23 failures, 15 were categorised as Type C, in which technical problems were steadily overcome but soon after the project was launched, the need for and/or the cost competitiveness of the intended products decreased. The other eight were categorised as Type D (Figure 9.6), in which the executing body had been too optimistic about its capacity to solve the technical problems and failed to attain R&D breakthrough.

Figure 9.5. Follow-up charts of Type A and B

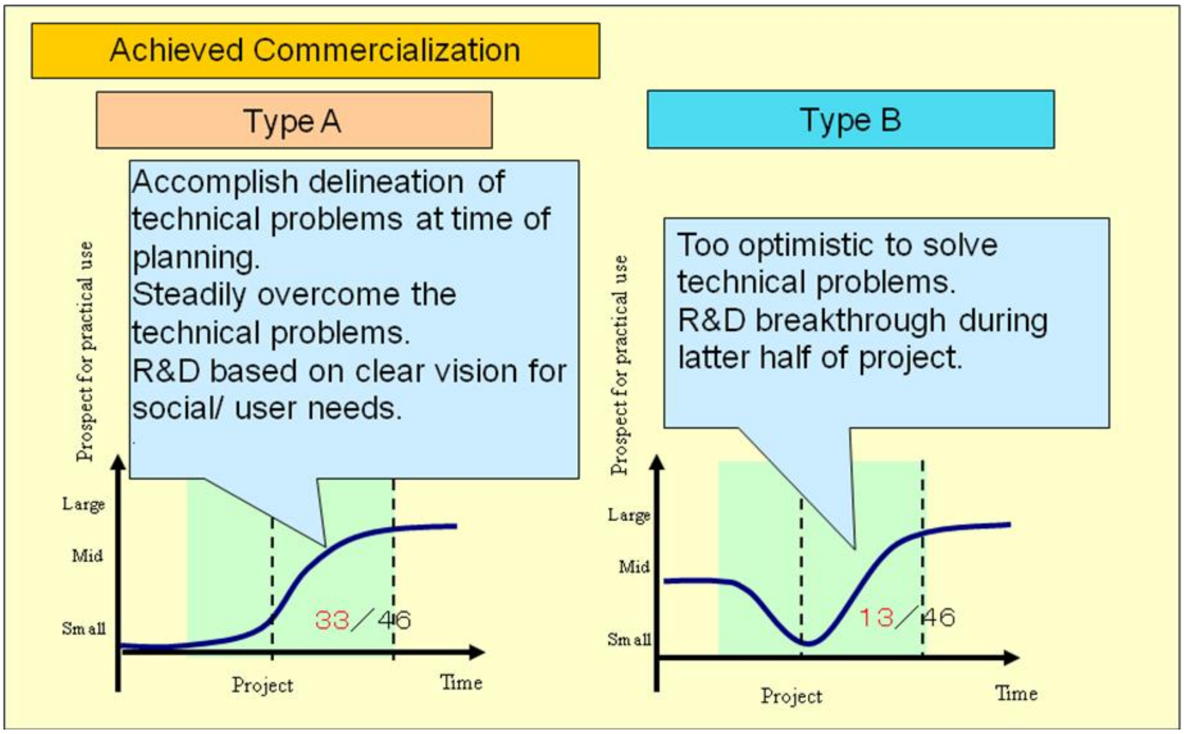
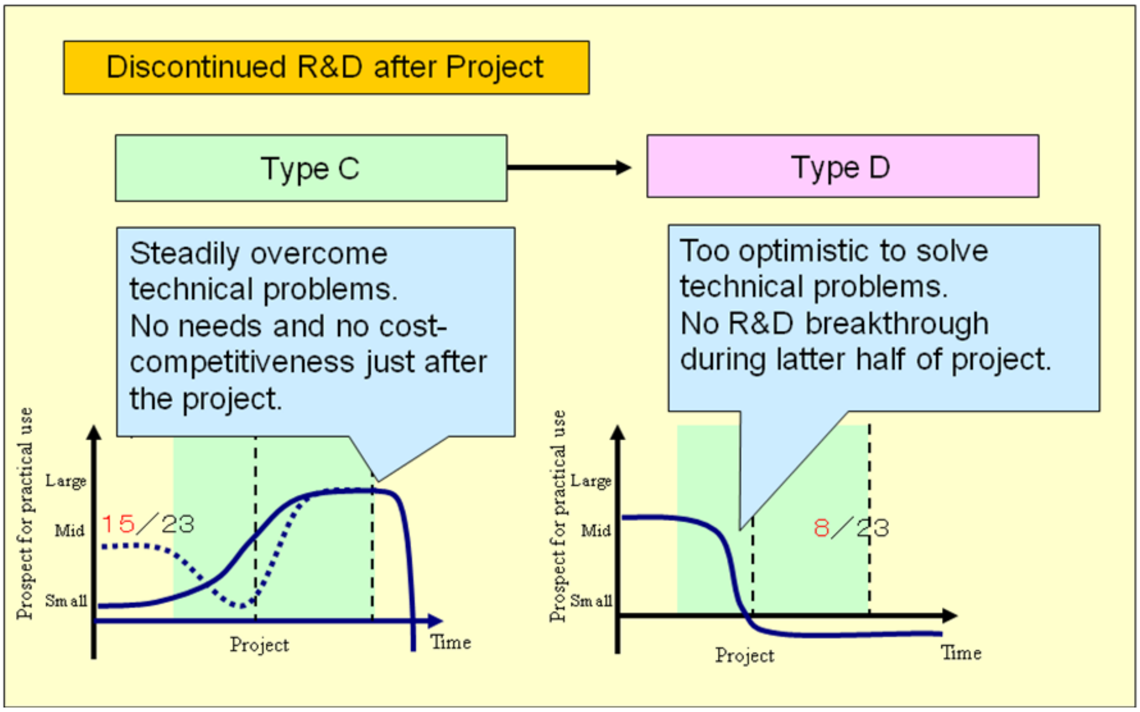


Figure 9.6. Follow-up charts of Type C and D



Advantages of follow-up monitoring

Lessons learned from follow-up monitoring have been consolidated and documented in the NEDO R&D Project Management Guideline. Results are accessible through NEDO's intranet system, which serves as a knowledge-sharing venue for companies participating in NEDO projects. NEDO also organises discussions on innovation policy with participating companies. Ultimately, participants benefit through these types of activities and evaluation exercises by understanding how the project is performing after the initial evaluation and seeing how to improve the project at a later stage. Also, while conducting follow-up monitoring, there are times when NEDO discovers that the technology developed has come to a dead end. In these cases, NEDO provides financial assistance aimed at removing obstacles. NEDO submits policy proposals, including results from the follow-up monitoring to METI. An example of this is the proposal on reforms of legislation and national innovation plans.

5. Conclusions

By monitoring the diffusion of outputs, METI uses the follow-up evaluation to view R&D projects over the long term with respect to policy objectives and long-term outcomes. This type of observation reveals factors that accelerate and impede innovation, contribute to finding new R&D projects and ensuring effective funding and regulation, while also being instrumental in justifying policy actions. The follow-up evaluation works well for adjusting innovation policy to the rapidly changing STI environment. By reviewing the diffusion of outputs, R&D decision makers ensure the legitimacy of policy intervention. This enables a "policy mix" for research and innovation policies. The disadvantages of follow-up evaluation include difficulties for collecting historical information and the inability to capture intangible factors and their impact on the project.

Unlike METI, NEDO focuses on applied R&D. NEDO monitors projects in the follow-up evaluation by including the executing body and collecting direct evidence on project outcomes. The monitoring reveals a sequence of events, which suggests many turning points and factors that accelerate or impede progress. Identification of these factors leads to appropriate reforms in R&D strategy. The results are loaded on METI's open database, thus creating feedback for the participants.

The disadvantages to follow-up monitoring include the lack of timely data. Since the follow-up review is conducted on projects completed five years earlier, obtaining historical information is cumbersome. Another disadvantage with the follow-up monitoring process is that it does not capture intangible factors that may have affected the R&D project. Some of these disadvantages can be overcome. For instance, the monitoring after the end of project conducted by NEDO is a good way to refresh the data. Moreover, the next step for METI would be to co-operate with NEDO for better follow-up evaluation.

REFERENCES

Arimoto, T. (2007), “Innovation Policy and Evaluation – Changing the national innovation system of Japan”, International Symposium on Innovation Policy and Evaluation, 19-20 November.

Nakamura, O. (2007), “Strategic Policy and R&I Management by METI: ‘Eco-innovation’ and ‘Strategic Technology Roadmap’”, International Symposium on Innovation Policy and Evaluation, 19-20 November.

Nakamura, O., M. Sawada, S. Kosaka, M. Koyanagi, I. Matsunaga, K. Mizuno and N. Kobayashi (2008), “Strategic evaluation of research and development in Japan’s public research institutes”, In C.L.S. Coryn and Scriven (eds.), *Reforming the Evaluation of Research, New Directions for Evaluation*, 118, pp. 25-36.

Ogasahara, K., O. Nakamura, K. Inahashi, C. Miyokawa, Y. Kimura and Y. Okubo (2007), “Analysis of Follow-up Evaluation Results of Research and Development (R&D) Projects Applying Logic Model to Elucidate the Process of Innovation”, American Evaluation Association Annual Conference in Baltimore, November 7.

Suzuki, J. and S. Yumitori (2006), "Behavioural Additionality of Public R&D Funding in Japan”, *Government R&D Funding and Company Behaviour*, pp. 151-165, OECD, Paris.

CHAPTER 10. AUSTRIA: THE EVALUATION OF GOVERNMENT RESEARCH FUNDING IN A SYSTEMS PERSPECTIVE [56]

This case study briefly explains the state of evaluation in Austria and offers guidance about how to strengthen the current system. This is particularly useful as Austria has historically had a relatively strong S&T programme.

1. Introduction

The Austrian innovation system accompanied by favourable political and economic conditions helped Austria's income and productivity catch up with the most advanced countries by the 1970s. Its innovation system was instrumental in helping Austria to forge ahead, and today, Austria is one of the top five countries in the European Union as measured by income per capita and is ranked among the top ten industrialised countries worldwide.

Complacency is Austria's greatest danger, as strains have appeared in the Austrian science and technology (S&T) system. It will need to increase its innovation efforts, boost efficiency and promote radical changes in the innovation system. New challenges come from new global framework conditions (globalisation, EU enlargement, internationalisation of research). Austria faces intense competition from both neighbours and Asian countries. Other countries are moving into Austria's specialisation in medium technology, deriving their competitive edge by adapting technologies imported from abroad and producing them at lower costs. In addition, innovation inputs are insufficient, particularly in the highest-quality, fastest-growing sophisticated industries. Business research expenditure is highly concentrated in a small number of firms, and few firms are highly innovative. Business research financed by foreign resources is twice the EU average. In addition, the Austrian innovation system is only loosely linked to and supported by the education system. The gap between the human capital available at the highest educational level and the demand from firms is increasing. Radical change is urgent: a high-income country such as Austria has to continue to compete in sophisticated markets and products.

2. A radical strategic shift in six dimensions

A change in innovation policy will require a radical strategic shift:

- From a narrow innovation policy to a comprehensive innovation policy. A comprehensive innovation policy is linked with education policy and includes improved framework conditions (*e.g.* competition, international openness, mobility). A narrow innovation policy concentrates on the measures and institutions directly involved in S&T.

---

56. This case study is based on a report on the evaluation of government funding of research, technology development and innovation (RTDI) from a systems perspective in Austria by Karl Aiginger, Rahel Falk, Andreas Reinstaller. It is one of several reports prepared for "Reaching out to the future needs radical change: Towards a new policy for innovation, science, and technology in Austria" (see Annex 10.A1). It was commissioned by Bundesministerium für Wirtschaft, Jugend und Familie und Bundesministerium für Verkehr, Innovation und Technologie.

- From a blurred division of responsibilities between and within ministries (and other "players") to well-defined responsibilities. In a situation in which responsibilities are well defined, ministries devise strategies for their area of responsibility based on a top-level vision. These are co-ordinated and monitored by a high-level government commission and a Council of Science, Technology and Research.
- From fragmented public interventions with no overarching vision to co-ordinated and consistent interventions based on a common vision, with the vision specifying economic objectives, external and internal challenges, and market failures that call for public intervention.
- From an imitation strategy to a frontrunner strategy. In a frontrunner strategy firms and researchers strive for excellence and market dominance in niches and high-quality segments, increasing market shares in sophisticated industries and technology fields and in areas or missions of particular relevance to society. In an imitation strategy, researchers strive for market dominance by adapting foreign imports and imitating foreign innovation.
- From a multitude of narrowly defined financial programmes to a flexible, dynamic policy defining broader tasks and priorities. In a flexible policy broad technology and research fields which are important for society should be defined top-down, while clusters and centres of excellence should be defined bottom-up. The latter should be sufficiently well funded to attract international leadership. Funding for programmes must be established through tax incentives and direct support to strengthen the frontrunner strategy.
- From managing public intervention by bureaucratic procedures to modern public management techniques. When modern public management techniques are used, ministerial competence centres or outside agencies pursue S&T goals. Agencies are free to choose instruments and are monitored on the basis of pre-defined output criteria without micro-interventions.

This report summarises some 50 recommendations for enabling a shift in strategy and improving the Austrian System of Science, Research, and Innovation (ASSRI).

3. The reference point: an overarching vision

The preconditions of a new innovation policy are:

1. a shared belief that research, technology and innovation are crucial for the welfare, growth and competitiveness of the Austrian economy;
2. consensus on which policy changes are necessary to increase the effectiveness of Austria's innovation policy.

This requires developing a general "vision" as a basis for the mission and goals of the Austrian science, technology and innovation (STI) system. The vision includes the ASSRI's relation to the education system, to societal and economic goals, and the framework conditions for innovation.

The highest level of government needs to develop this "vision" as a reference point for all strategies of ministries, regions, institutions and agencies. It can serve as a blueprint for a new science, technology and innovation policy. A team of national and international experts should prepare the "vision", but it should be finalised and "owned" by the government. It should be enacted into legislation by parliament, monitored by a Council of Research, Science and Technology, and serve as the basis for the strategies of ministries, regions, institutions and agencies.

4. Co-ordination and monitoring: reformed institutions

Implementation of a new STI policy requires co-ordination of separate policy strands. It is proposed to set up a high-level co-ordination commission on research and technology composed of the ministries responsible for innovation and education and headed by the Chancellor and the Deputy Chancellor. It should meet about twice a year to monitor the implementation of the vision. The government should be accountable to a new permanent parliamentary committee for science and technology (a merger of two existing committees). The parliamentary committee should discuss the progress of the vision in the annual report of the Council of Research, Science and Technology.

5. Better governance: new role of ministries

The change in strategy calls for a new and better-defined role for the ministries in charge of innovation policy. The ministries will devise strategies based on the overall vision in their respective areas of responsibility, focusing on being a frontrunner in the innovation race and on links with other policies. They decide which parts of their strategy have to be ensured internally (*e.g.* linking the innovation and education systems; improving framework conditions), and which parts have to be delegated to agencies or institutions. Each ministry should be responsible for the implementation of well-defined parts of the new strategy. The "high level co-ordination commission on research and technology" should co-ordinate ministries' activities and define their goals and milestones.

6. Better governance: increased autonomy of agencies

The autonomy of individual agencies should be increased. This will require new governance procedures. At the administrative level, Austria needs to systematically build competency to manage the agencies and to co-ordinate the intra-ministerial processes of policy development. Processes to co-ordinate policy development activities across departments in order to avoid overlaps and conflicting assignments to the agencies should be developed. Broad tasks should be delegated to agencies and be monitored according to output goals whenever feasible. For these tasks actual goals and outcomes are specified. With their goals in mind, the agencies should develop suitable programmes (or strategies) that fit into their overall portfolio. If programmes are delegated rather than tasks, they should be much broader than at present and undergo a strict test based on need. Austria needs to move away from programmes that run forever to programmes that are extended or ended on the basis of their evaluations. To be efficient the process needs new and compatible reporting systems across ministries and within and across agencies. Micro-management and micro-intervention should be abolished. Control should be exclusively based on *ex post* assessments of outcomes.

7. Switch to a frontrunner strategy

The new STI policy should be a frontrunner strategy and should aim at helping Austrian firms to achieve and sustain economic leadership through product innovation and productivity growth in niche markets. This requires Austrian companies to obtain a winning margin in technological and market competencies over their principal competitors. This can only be achieved through progressively more ambitious research and development in the business sector and more better-qualified people.

8. Government commitment: ambitious goals for 2020

The Austrian government has set as a goal to increase research expenditures to 4% and expenditures for tertiary education institutions to 2% of its GDP by 2020 (part of the latter is included in the former). Europe trails the United States and Japan in research, and set a goal of 3% of GDP for 2010 with no chance of reaching it soon. As a high-income country, Austria should be more ambitious. The Austrian government should take the necessary steps to make sufficient financial means available in public budgets to finance a tax credit and directly support R&D, to finance university research sufficiently, and to improve the quality of education and the number of graduates from higher education institutions. Since the 2% and the 4% of GDP are only input goals, complementary output goals would be necessary to track efficiency.

9. In the face of crisis: maintaining dynamics

Keeping research expenditure dynamic is absolutely necessary. Over the past 10-15 years, Austria's research expenditures were dynamic, but they are in danger today. Private investment in research will be curtailed during the crisis, particularly by multinational firms. Empirical evidence shows that research expenditures are highly pro-cyclical, and this is very much the case in Austria. Elections and deferrals in the budgeting process for 2009 have already delayed spending by public funds and institutions. Other sources are drying up as the crisis deepens, and public money will be scarce during and after the crisis. In a frontrunner strategy, it is extremely important to maintain dynamic research expenditures. The current expenditure path for the coming year is lower than planned and necessary to reach the target of 3% of GDP in 2010.

10. A new simple tax credit: broadening the base and shifting the level

A frontrunner strategy needs a broader base. This means a higher number of firms active in research, more firms innovating regularly, more innovative business start-ups, new research departments in existing firms, a larger number of firms choosing locations in Austria for research facilities, and more firms co-operating with universities. As a driving force for broadening and shifting the level of innovation expenditures, this case study proposes using a single tax incentive, namely an extended tax credit of 12%. This new tax credit would replace all existing schemes for expenditures based on patents or certified "importance for the Austrian economy", all tax allowances, and the scheme for incremental innovation. It is more generous for firms, with the exception of those liable to income tax. It is visible and provides incentives for shifts from physical to intangible investments within firms (*e.g.* creation of research departments) or for shifting a multinational firm's research towards an Austrian location. A strong driver for more investment in R&D is necessary if the goal of 4% of GDP is to be reached in the foreseeable future, particularly since multinationals contribute strongly to R&D expenditures in Austria. Total investment in R&D must increase by approximately 8% between 2008 and 2020 (in nominal terms) to reach the target of 4% of GDP by 2020. To reach this goal it is necessary to create tax incentives that lead to continuous investment and location decisions. Equally or even more important is the knowledge base, the research capacity of universities, research labs and human capital.

11. The complementary role of direct support

A frontrunner strategy must not only have a solid base, it must also support excellence. Structural shifts, top-level research and efficiency are essential to realising the "vision". However, such changes are difficult to bring about. Direct support should therefore aim in particular at improving quality and funding high-risk projects. It should focus on firms with high innovation and knowledge intensity, technological excellence, quality in niches, high-technology start-ups, programmes of technological excellence and thematic priority. Moreover, direct support is an appropriate instrument for promoting activities in the pre-market phase whose benefits are external to the firms. Direct support is no substitute for tax incentives but complements it. Support also enables a learning process, because it provides information and a certain degree of consulting on programmes and projects. Therefore, direct support is important for firms starting or upgrading innovation ("changing track").

12. Direct support and tax credit: no trade-off

Given the goals of 4% and 2% of GDP, it is necessary to increase funds for direct support at a rate exceeding GDP growth. To increase the effectiveness of direct support this study recommends *i)* reducing the number of programmes (not the money spent) and allowing agencies more discretion in the choice of instruments; *ii)* defining output goals for agencies rather than input goals; *iii)* keeping thematic programmes from defining narrow sub-fields and allowing these programmes to develop and cluster bottom-up; *iv)* increasing and promoting quality and co-operation with universities by implementing basic and open programmes; and *v)* establishing science programmes to support thematic fields and fostering co-operation, competence centres and excellence programmes in a bottom-up manner.

There is no trade-off between direct support for business research, scientific projects and tax incentives. These drivers for increasing research efforts should grow faster than the total economy. Otherwise, Austria should probably give up its expenditure targets and its ambition to become a frontrunner.

13. Higher education: new funding rules and additional research money

The quality of universities, universities of applied sciences and non-university research institutions is a crucial determinant of a frontrunner position. Quality is related, among other things, to the financial means of higher education institutions and proper incentives. Currently, research institutions are not sufficiently funded to ensure high-quality research or teaching. It is therefore necessary to increase spending for tertiary education to the level recommended by the European Commission (2% of GDP). The current lack of tertiary graduates, especially in science and technology, is an important bottleneck for industry and academia.

To increase the efficiency of tertiary institutions, expenditure for research and teaching should be separated. Funding per student at universities should be based on the model currently applied for the universities of applied sciences, and additional research money (which should not reduce the funds for the Austrian Science Fund, FWF) should be allocated to universities on the basis of performance criteria (the criteria should also include co-operation with firms).

Money should be distributed within universities in a more competitive manner. For instance, funding should go to persons, specifically young scientists, and not to institutes. A new tenure track system based on international best practice should be envisaged and career steps in universities should depend on international experience and a competitive framework.

The budget of the Science Fund should be partly used to finance thematic programmes (if defined by the vision). Research infrastructure should be supported, *e.g.* by increasing the overhead costs covered in FWF projects from 20% to 50%.

Block grants to non-university research institutions should depend on the existence of a mission as well as defined milestones in academic research and infrastructure.

A career path from apprenticeship to the universities of applied sciences should be created, marketed and promoted by organisational instruments (modules) with appropriate financial support.

R&D co-operation between university and industry should be stimulated since radical innovations often arise from academic research and scientific discoveries.

14. Guiding principles of a new policy: non-exclusivity, learning and mobility

The new strategy should be built on the principles of openness, non-exclusivity and mobility between firms and institutions. The strategy has to be implemented top-down, while information has to be gathered bottom-up. Openness to change and obtaining knowledge from external sources should be the overarching principles. Funding and policy decisions should be less influenced by the weight of interested parties, insider knowledge and entropy.[57] The innovation system should be open to experiments (*e.g.* pilot studies). Moreover, a shift from programme-based to task-based intervention would make it easier to end programmes. Continuous assessment and external evaluations by international teams should ensure that if the economic environment changes the system also changes. Insiders and users of the current system are not overly critical of it. They complain about administrative costs, but emphasise that the system provides guidance, although their direct funding does not dramatically affect their decisions. In comparison, the new strategy allows new techniques to be learned, provides information and control, and helps with planning.

15. Regional, national, European: co-ordination and agenda setting at all levels

The Austrian research promotion policy should be redesigned and anchored in a multi-level system encompassing the EU, federal and regional levels. Deficits in co-ordination and specifically in agenda setting should be tackled. Problems associated with cross-policy incoherence should also be tackled. Demand for action exists specifically at the interface with the European level. Austrian firms and institutions get more money back from the EU than government pays, but the European Commission does not have a strategic, active co-design of STI policy.

At the federal level there should be integration with education, health and environmental policies. At the regional level, reorganisation is needed to address the problem of one-way communication from the federal level. A two-way exchange of information and combined learning as well as possible support from cross-region activities should characterise the new system.

16. Conclusion: a task beyond policy borders

Change is necessary, not because the science, research and innovation system has not worked, but because of new challenges, including the economic crisis, and the position of Austria as a high-income country. A successful innovation policy for a frontrunner has to be more comprehensive and needs to interlink with other policies. The system should react to external and internal challenges, as well as to economic and societal trends. The necessary changes are not minor. They are significant and call for attention from the top political level and an overhaul of current management and monitoring techniques. They should rely on human capital, and build on the quality of the education system. The competitiveness of firms during and after the current crisis will depend greatly on education and innovation and this emphasises the need to switch from an imitation strategy to a frontrunner position.

---

57 Defined as the system's lack of permeability to information from external sources.

ANNEX 10.A1
BASIS OF THE AUSTRIAN CASE STUDY

The table presents the nine special reports on which the Austrian case study is based.

Report	***Title***	***Authors***	***Institute***
1	Framework Conditions	Jürgen Janger Michael Böheim Nadine Grieger	OeNB and WIFO
2	Strategic Governance	Gabriele Gerhardter Markus Gruber Simon Pohn-Weidinger Gabriel Wagner	convelop
3	Governance in RTDI – Relation between Ministries and Agencies	Sabine Mayer Iris Fischl Sascha Ruhland Sonja Sheikh	KMFA
4	Tax Incentive Schemes for R&D	Rahel Falk	WIFO
5	Direct Public Funding of RTDI	Sabine Mayer Iris Fischl Sascha Ruhland Sonja Sheikh	KMFA
6	Effects of Block Grants on Research Institutes and Universities	Michael Astor Ulf Glöckner Stephan Heinrich Georg Klose Daniel Riesenberg	prognos
7	Public RTDI Funding - The Users Perspective	Sabine Mayer Sonja Sheikh Jürgen Streicher	KMFA
8	Coherence of the Instrument Mix	Rahel Falk	WIFO
9	Intervention Logic - Interaction between Institutions and Actors	Michael Astor Stephan Heinrich Georg Klose	prognos
Synthesis Report	Reaching out to the Future needs Radical Change	Karl Aiginger Rahel Falk Andreas Reinstaller	WIFO

Internal referees (WIFO)

Martin Falk, Klaus Friesenbichler, Werner Hölzl, Margit Schratzenstaller, Fabian Unterlass

External experts

David Campbell (IFF Vienna), Heinz Hollenstein (KOF-ETH Zürich), Gernot Hutschenreiter (OECD), Barbara Kehm (University of Kassel), Hannes Leo (Leoon.at), Andrea Löther (GESIS Bonn), Herwig W. Schneider (iwi Vienna), Dorothea Sturn (University of Vienna), Gunther Tichy (ÖAW, WIFO)

Research assistance

Julia Borrmann, Simon Braker, Dagmar Guttmann, Nadine Grieger, Elisabeth Neppl-Oswald, Michaela Reischitz, Karolina Trebicka, Ursula Weixlbaumer

Survey

Gerhard Schwarz (WIFO), Roald Steiner (Salzburg University of Applied Sciences)

Steering Board

Federal Ministry of Transport, Innovation and Technology (BMVIT)
Federal Ministry of Economics, Family and Youth (BMWFJ)
Federal Ministry of Finance (BMF)
Federal Ministry of Science and Research (BMWF)
Austrian Council for Research and Technology Development

The Advisory Panel

All members of the Steering Board, and

The Austrian Federal Economic Chamber (WKO)
Federation of Austrian Industries (IV)
Federal Chamber of Labour (AK)
Austrian Trade Union Federation (ÖGB)
Austria Investment Agency and Promotional Bank (AWS)
Austrian Research Promotion Agency (FFG)
Austrian Science Fund (FWF)