

Chapter 1

POLICY EVALUATION IN INNOVATION AND TECHNOLOGY: AN OVERVIEW

by

George Papaconstantinou
Economic Analysis and Statistics Division, OECD, Paris
and
Wolfgang Polt
Science and Technology Policy Division, OECD, Paris

Background and context of the conference

The conference on “Policy Evaluation in Innovation and Technology” on which this volume is based was organised in the context of the OECD work on “best practices in innovation and technology policy”. Its aim was to bring together policy practitioners from OECD Member countries with researchers working on evaluation issues in order to present and compare different experiences in policy evaluation methods and practices. Its focus in particular was on examining the quantitative and qualitative tools used in evaluations, on the one hand, and the institutional set-up within which evaluations take place, on the other.

Identifying policy “best practices” in innovation and technology policy is a challenging task, made all the more difficult by the large variety of initiatives put in place in OECD countries. They cover a wide spectrum, ranging from the direct support of basic research to more indirect measures aimed at improving the capacity of firms to innovate and to use new technologies. Ultimately, such policies contribute to higher productivity and growth, and to the creation of more and better jobs. Given the growing importance of “knowledge-based” economic activities, it is all the more crucial to identify how the maximum leverage of such policy initiatives can be obtained.

Evaluation issues are central to formulating policy “best practices”. There is in effect an increased interest in OECD countries in the issue of evaluation of government programmes and policies. This can be partly traced to budget stringency and the need to better allocate what are increasingly scarce public resources. More fundamentally, however, the focus on evaluation is emblematic of a broader reassessment and examination of the appropriate role of government and of market mechanisms across a number of policy areas. Accountability, transparency and the desire to minimise distortions arising from government policies while maximising their leverage effect are all driving this trend towards more evaluations. At the same time, new developments in technology and innovation policy with an increased emphasis on technology diffusion and adoption, organisational change and innovative behaviour have raised new methodological challenges for the evaluation of these policies.

Rather than summarising the papers presented, this introductory chapter to the Proceedings will discuss briefly some of the key issues that were raised in the conference and that cut across different

sessions and discussion panels. They relate to the scope and coverage of evaluations, developments in evaluation methods and practices, rationale and criteria used in evaluations, the diverse quantitative and qualitative tools used, and the role of institutions and country specificity.

Scope and coverage of evaluations

Broadly speaking, evaluation refers to a process that seeks to determine as systematically and objectively as possible the relevance, efficiency and effect of an activity in terms of its objectives, including the analysis of the implementation and administrative management of such activities. The scope and methods of evaluation differ according to the questions to be addressed and the character of the policy measure. Thus, they can be retrospective (*ex-post*), current or prospective (*ex-ante*) evaluations, producing information that can be used in the assessment of past policies, the monitoring of ongoing initiatives or the forward planning of innovation and technology policies.

Evaluations can also have different purposes. In effect, part of the difference in opinions expressed in some of the contributions to this conference and in the various discussions can be traced to different perspectives on what evaluations aim to achieve. Policy makers and economic analysts stressed the role of evaluation in examining the justification of a programme, analysing its economic effects through its impact on the incentives of firms and individuals, and thus providing information to guide resource allocation as well as more strategic decision processes which involve selection of instruments (*e.g.* using tax-based measures *vs.* grants in order to support industrial R&D), or the thrust and direction of technology policies in general. Alternatively, many professional evaluators as well as policy makers involved in hands-on running of programmes stressed the role of evaluation in improving the conduct, quality, responsiveness and effectiveness of a programme, thus raising its leverage effect. Clearly, these objectives are complementary; achieving them, however, often calls for different evaluation tools and institutions carrying out the evaluations.

Different types of policies also ask for different evaluation methods. Within the broad area of innovation and technology diffusion policies, the assessment of large-scale mission-oriented programmes or of fiscal incentives for industrial R&D are more amenable to the use of cost-benefit analysis. Newer programmes like those fostering pre-competitive R&D collaboration ask partly for other methods, not yet fully developed. The evaluation of diffusion-oriented programmes needs more elaborate micro-level econometric analysis, and hence high-quality data and databases, than others. Yet other programmes focusing on the “soft side” of the innovation process (such as awareness, information and consulting programmes) demand the use of intensive case studies and user-surveys. Thus, the development and application of evaluation methods reflect also the stage of development of technology and innovation policy in general.

Developments in evaluation methods and practices

Evaluation methods and practices have developed alongside the evolution of technology and innovation policy and the understanding of the innovation process. Starting from the predominant model in the post-war period, the focus was first on the assessment of the quality of scientific research, with peer review and bibliometric techniques (impact analysis and citation counts) as the main methods. Programme evaluation developed later, following the proliferation of government programmes to support industrial innovation, often through fostering collaborative research. These initiatives demanded more elaborate techniques both for the evaluation of direct and indirect socio-economic effects and for the

assessment of the conduct of the programmes. In-depth surveys of the impact of programmes were complemented by more quantitative techniques, such as econometric methods. Cost-benefit analysis, until recently reserved for the evaluation of large infrastructure or transport projects, became increasingly used in the evaluation of programmes to foster industrial research.

With the recognition of the complex, systemic nature of innovation processes, and with technology policy instruments covering a growing range of activities, evaluation techniques have had not only to develop quantitative techniques further (*e.g.* micro-level analysis for the assessment of diffusion-oriented programmes) but have had also increasingly to look at the “soft side” of innovation (*e.g.* in trying to capture networking, learning effects, etc.), where further work is still needed to identify their economic impacts. The proliferation and widening coverage of policy initiatives have led evaluation to increasingly adopt a portfolio approach, rather than focusing on individual projects; to a greater use of performance indicators; and towards a convergence between the activities of *ex-post* evaluation and continuous monitoring. This multi-faceted approach has also been made necessary by the multiplicity of actors involved in technology and innovation policies. Each of these actors (policy makers deciding on a programme, programme managers designing and conducting it, firms participating, etc.) requires different types of information, hence the need for a combination of methods shedding light on the basic rationale, the economic impact, the administrative conduct and the customer satisfaction derived from the activity.

Evaluation criteria and tools

Rationale and criteria

A crucial issue in evaluation is the criteria to be used for judging programmes and policies. The basic rationale for government initiatives to stimulate technological development in the first place is the recognition that there is a difference between the expected private rate of return and the social rate of return, with the private rate being too low to induce firms to engage in innovative activities that would be beneficial from a societal standpoint. This “market failure” rationale suggests that while “additionality” and the existence of positive private returns to firms as a result of government programmes are preconditions for success, for policy to be fully justified the net social benefits of a government programme must be positive; the programme needs not only to be effective in changing behaviour, but also efficient from a social point of view. A number of participants in the conference, in effect, suggested that the focus on additionality (the changes in behaviour and performance that would not have occurred without the programme) as a criterion for success is simply a reflection of the difficulty of accurately measuring spillovers or externalities and thus the net social benefits of programmes.

The accumulated experience of three decades of technology policies, together with recent advances in innovation theory, have shown the limits of a simple “market failure” rationale to policy. The preponderance of “government failure” has forced evaluators to be more careful in accounting for costs of programmes as well as for benefits, including those costs that are associated with the distortions to economic incentives that policy initiatives can bring about. On the other hand, the realisation that the benefits of individual programmes or policies can often be understood only in the context of their impact on a complex innovation system has given rise to the notion of “systemic failure” as a basis for policy. In terms of evaluation, this has translated into the more elaborate principle of “behavioural additionality”, which is intended to capture the many ways in which participation in a programme can change the innovative behaviour of a firm. As a number of conference participants noted, it has also forced

evaluators to recognise that identifying social benefits in diffusion policies involves a dynamic analysis that looks at the development of new capabilities, and of learning.

Tools and methods

A large part of the conference papers and discussions was devoted to the critical presentation of different methodological tools for evaluation: cost-benefit techniques, econometric methods, case studies, in-depth surveys and peer reviews. There were clear differences between participants in the faith that they placed in conclusions based on quantitative as opposed to qualitative techniques. Nevertheless, whatever the type of policy being evaluated (financial support to industrial R&D, large technology programmes, diffusion-oriented policies), it was equally clear that a combination of approaches (quantitative and qualitative) is needed in order to cover all aspects of the evaluation process; different approaches are complementary, not mutually exclusive. In effect, the distinction is not that clear-cut: it often turns out that quantitative techniques produce mainly qualitative information. Thus, to increase the credibility of evaluation results, a number of alternative methods should be used to consolidate the foundations of policy recommendations.

Some of the most interesting recent developments in evaluation methodology concern the use of econometric techniques based on longitudinal micro-level data, where the impacts of programmes are examined by comparing the performance characteristics of firms that are clients of government initiatives (such as extension services) with those of non-client firms. The quality of results based on this approach is, however, conditional on the extent to which researchers can control for firm characteristics other than programme participation. Furthermore, this technique is only the first step in a full cost-benefit analysis: at its best it establishes the private benefits conferred to firms as a result of the programme; justification of a programme needs to account for social benefits against the total costs.

Many participants felt that evaluation schemes should be constructed around social cost-benefit frameworks which estimate the impact induced by the policy measure, its spillover benefits, as well as costs such as the marginal excess burden of taxation and compliance costs. Nevertheless, they warned against the spurious precision that cost-benefit calculations can give, and suggested that such schemes should combine qualitative and quantitative indicators of the impact of the policy measure and of the private and estimated social benefits from the induced change in behaviour. Ideally, they should be combined with the qualitative information from user-surveys, in-depth case studies and interviews to produce the variety of information needed by the different users of evaluations. Single-approach evaluations might in effect be downright misleading, and putting too much emphasis on single quantitative estimations, while useful as a measure of cross-checking, might miss the essential qualitative effects of new initiatives. Furthermore, it is clear that quantitative techniques have to be developed further, especially with regard to the challenge of capturing the economic impacts of the “soft factors” of innovation (impact of programmes on learning, co-operative and innovative behaviour).

The need for an approach combining quantitative with qualitative information is also underscored by the fact that programme management also necessitates looking into the process and performance of different policy instruments. Given the high variance of returns in different technology projects, detailed case studies are important to see what works and what does not. But, whatever method is used, the importance of having a “counterfactual” in policy evaluation exercises was stressed: evaluation requires comparing the absence of policies with the impacts in the presence of policies. Furthermore, a number of participants stressed the fact that much of the evaluation work to date ignores the competitive environment

within which client plants and firms operate, and in which the services provided by governments are supposed to improve performance; yet it is important to have an understanding of this environment in order to optimally design, provide and evaluate programme services and to ensure that programme objectives are not at odds with those of the client firms.

Institutions and country specificity

Techniques aside, evaluation is very much a social process, as it involves interaction of individuals, organisational methods, practices and routines. The institutional set-up within which programmes and policies are evaluated in effect determines the nature, quality, relevance and effectiveness of evaluation practices. Many papers and discussions in the conference treated this issue, and addressed the question of whether there is such a thing as an “optimal” institutional set-up that is transferable across countries. On this question of country specificities and general principles in institutional arrangements, it was felt that while practical arrangements and also to some extent policy issues are country-specific, basic principles/challenges in evaluation are not. From a practical point of view, what is necessary is to reconcile evaluation designs with the varying needs of programme sponsors, service providers and customers; to incorporate methods that can support programme learning and improvement as well as address issues of programme justification; to reconcile desired information needs with resource and information availability constraints; and to ensure that evaluation takes place on a programmed and properly resourced basis, guarantees the independence of the evaluators and provides a mechanism for feedback of results into policy making.

In general, evaluation practices are far from uniform across OECD countries; as the different contributions to this volume show, in the mid-1990s programme and policy evaluation is characterised by different degrees of maturity as well as by a large measure of variety in terms of the tools used, the institutions involved, and of the place of evaluation in policy making in general. Despite such differences, there are some common lessons that can be drawn. A basic lesson is that evaluations must take into account the strengths and variety of the national systems of innovation in order to develop systematic evaluation practices embedded in the policy-making process; there is no optimal institutional design for evaluations that is transferable across countries. Nevertheless, there are some general conclusions to be drawn from the comparison of country experiences. First, experiences from several countries show that evaluations should be designed together with the programme or policy to be evaluated. Only such an early preparation would secure the collection and provision of the data needed and the common acceptance of the evaluation procedures and criteria among the institutions involved.

Secondly, as the results of evaluations are often taken up only in a “localised form” (that is, they are only implemented if the institution evaluated could implement the recommendations on its own), there is a need to secure the take-up on a higher level of policy making. Thus, it seems necessary to put a formal obligation on those responsible for policy making to react to the results of evaluations or to expose the results of evaluations to a public discussion, which would also result in a higher awareness of policy makers. A presumption in favour of publication of evaluation reports was thought to be very important in this respect in many countries; once in the public domain it is more difficult to ignore results of evaluations. Some countries have gone far in implementing such disclosure and feedback arrangements, but many still lack this form of feed-back mechanism.

Third, evaluations ought to be more “user-oriented”, that is to address the informational needs of the respective “clients” (policy makers, firms, programme administrators on various levels); hence they

should encompass an appropriate mix of methods to produce these different types of information. This would also improve the take-up of their results. There are some promising examples of this approach, notably in countries where there is a strong budgetary pressure on programmes and institutions. Finally, in almost all countries evaluations have so far been used mainly for incremental changes (*i.e.* for the improvement of programmes), but hardly to guide more fundamental shifts and re-orientations in technology and innovation policy. To empower evaluation for such a task, one would have to embed it into a wider system of information gathering and preparation, linking it to technology foresight and technology assessment exercises. In addition, the role of the evaluator would then have to change from being predominantly a “referee” to a “moderator” of this information-gathering process and a “coach” for the strategic policy decision-making process.

This raises the question of how far evaluation can go. Although welcoming an increased scope of evaluation techniques to produce a greater variety of information more reliably, policy makers during the conference warned against stretching evaluation too far. In their perspective, evaluation can help to guide informed choices, but not to substitute for a political decision-making process. While evaluations are increasingly used in the policy process (especially when results are easily interpretable by decision makers), they are not always decisive, partly because policy involves trade-offs and values, and partly because evaluations are often not conclusive, due to uncertainty in the impact of many programmes. In effect, many policy decisions are based on intuition and first principles, and often evaluation is just used to justify certain decisions after the fact. Nevertheless, it was clear that while certain programmes may have been over-evaluated, overall we have yet to reach diminishing returns. There is still a need for more and better evaluations, and especially for evaluations that go beyond individual programmes and compare the impact of different spending initiatives, examining in other words the appropriateness and efficiency of using different policy tools to achieve a given objective.

Future OECD work on evaluation: towards best practices

The rich variety of perspectives presented in the conference and the intensity of the discussions suggested that there is ample scope for mutual learning in the comparison of “good practices” between countries and from advancements in economic theory, the ultimate goal being the regular conduct of evaluations using “state-of-the-art” methods and an institutionalised feedback into policy decision making. The OECD has in its past and current work tried to further this objective, both in the context of its analysis of innovation and technology policies, but also more generally in terms of its work in many other areas of policy making.

The discussions and papers in this conference directly feed into the final report from the OECD work on “best practices in innovation and technology policy”, to be presented at the OECD Council at Ministerial level in 1998, and published thereafter. Furthermore, against the background of this conference, future work on evaluation by the Directorate for Science, Technology and Industry will have a number of dimensions. First among these is looking for best practices for designing programme evaluation, and covering issues such as articulation of objectives, criteria for assessing success/failure, and institutional issues, in collaboration with the OECD Public Management Service. Second, is looking for best practices in terms of the “tools” used in evaluation, in particular encouraging the international comparability and use of micro-level data sets, of social cost-benefit analyses, but also of user surveys and case-studies. Finally, a third dimension of future OECD work is addressing broader issues in policy evaluation in innovation and technology: devising “best practices” that take into account our new understanding of innovation processes (concepts such as “systemic failure” to complement market and government failure) but also account for country specificity.