

OECD CONFERENCE POLICY EVALUATION IN INNOVATION AND TECHNOLOGY

SUMMARY OF THE PROCEEDINGS

MAIN POINTS

- The DSTI organised in June a conference on Policy Evaluation Practices in Innovation and Technology in the context of the work on “best practices in innovation and technology policy”, whose aim was to bring together policy practitioners with researchers working on evaluation issues in order to present and compare the experience in policy evaluation methods and practices.
- The recent interest in evaluation is partly due to budget stringency, but also to a reassessment of the appropriate role of government vis-à-vis the market across a whole spectrum of policies, which has put a premium on “good governance”: the need for better priority-setting, streamlining and policies that deliver value for money.
- Evaluation practices have mirrored the evolution of technology and innovation policy, shifting from focusing exclusively on evaluating scientific research, to an increasing focus on programs enhancing the environment for innovation and technology transfer. But further work is needed to identify the economic impacts of “softer” policies (networking, learning effects, etc.).
- Whatever the type of policy being evaluated (financial support to industrial R&D, large technology programmes, diffusion-oriented policies), a combination of different approaches (quantitative and qualitative) is needed in order to cover all aspects of the evaluation process; different approaches are complementary, not mutually exclusive.
- Identifying “additionality” as a result of government policy and the externalities or spillovers of specific government policies is crucial in clarifying objectives, laying down criteria for evaluation and assessing effectiveness. The existence of positive private returns to firms as a result of government programs is a precondition for success; but for policy to be fully justified the net social benefits of a government program must be positive.
- In addition to its function for program justification and as a guide to resource allocation, an important purpose of evaluation is program improvement, where evaluative information helps to improve program quality, responsiveness and effectiveness relative to resources and needs. This makes it important to understand the mechanism of the program under evaluation through both qualitative and quantitative information..
- In terms of institutions, evaluations must build on the strengths and variety of the national systems of innovation in different countries 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.
- In terms of the coverage of evaluations, there is a welcome shift in evaluation practices towards a more strategic (“systemic”) approach as opposed to an exclusive focus on program-specific or institution-

specific evaluation. More effort is necessary to address the question of how to attain given policy objectives through different instruments.

- Future OECD work in this area will look for best practices for designing program evaluation (articulation of objectives, criteria for assessing success/failure, institutions), in terms of the “tools” used in evaluation (international comparability and use of micro-level data sets, social cost-benefit analyses, but also of user surveys and case-studies), and address broader issues, devising “best practices” that take into account new understanding of innovation processes (“systemic failure” to complement market and government failure) while accounting for country specificity.

Background

1. The Directorate for Science, Technology and Industry organised on 26 and 27 June an international Conference on Policy Evaluation Practices in Innovation and Technology in the context of the work on “best practices in innovation and technology policy”. The aim of the conference was to bring together policy practitioners from OECD Member countries with researchers working on evaluation issues in order to present and compare the experience in policy evaluation methods and practices with a view of bringing out what constitutes “best practices” in this area. This paper summarises the papers presented in the conference and the discussions that took place. The conference was structured around 4 sessions covering different areas in innovation and technology policy: financial support to industrial R&D; large technology programs; technology diffusion policies; and strategic issues. There were in addition two discussion panels, addressing issues relating to the techniques used in evaluation and to the institutions involved. It concluded with a policy roundtable, whose aim was to discuss the feedback of evaluation results into policy design. The actual program of the conference is provided at the end of the paper.

Introductory remarks and issues paper

2. In her introductory remarks, Deputy Secretary-General *Joanna Shelton* stressed the fact that across OECD countries, the evaluation of government programs or of whole sets of government policies has emerged as a major item on the policy agenda. This is partly because of the current budget stringency, but is also due to a reassessment of the appropriate role of government vis-à-vis the market across a whole spectrum of policies, which has put a premium on “good governance”: the need for better priority-setting, streamlining and policies that deliver value for money. The role of the OECD in this area is to promote “best practices” across Member countries, taking due account of the specificities in institutions and structures. By “best practices”, it is understood that there are too many differences between countries in their science and technology systems, in their institutional structures, even in their understanding of the appropriate role for government to be able to make “one size fits all” policy recommendations. Nevertheless, the rich diversity and experience of OECD countries in evaluation holds valuable lessons. In this context, the aim of the conference is to promote such “institutional learning”: to gather information on which methods, procedures and institutional settings are most likely to yield adequate evaluations of past or existing policies and a guide to future policy-making.

3. *Thomas Andersson*, Deputy Director of the DSTI, placed the conference in the wider context of the OECD project on Technology, Productivity and Job Creation. Through this project, the OECD has become firmly involved in identifying “best practice” in innovation and technology policy. He noted that the preconditions for technical progress are influenced by a range of institutions belonging either fully or partly in the public sphere, and by incentive structures which governments interfere with in different ways. There is evidence that evaluation practices in OECD countries are becoming increasingly frequent and strategic in nature, going beyond the evaluation of individual programmes in an attempt to address some of the “systemic” features of innovation criteria. But, while practically every OECD country now evaluates at least some of its innovation and technology policies, the reality often falls short of the rhetoric about accountability, efficiency and consideration of economy-wide effects. In particular, the feedback of evaluation results back into policy design is perhaps the weakest link in the chain.

4. *David Shand* of the OECD Public Management service gave a brief review of PUMA work on evaluation, noting that evaluation of policies and programs of great interest in many OECD Directorates. PUMA is interested in performance measurement, but especially in performance management: how the evaluation process fits into decision-making, and what is the impact of evaluations. Their current work involves a survey of management techniques, and will culminate in the development of best-practices

guidelines on issues such as how to get the right sort of demand for evaluations, how to institutionalise them without bureaucratising the process, and how to ensure high quality without ending up with evaluations that are too technical.

5. In the opening paper to the conference, *Luke Georghiou* (PREST, United Kingdom) surveyed OECD countries experience with evaluation, together with the issues that it raises. He suggested that the development of evaluation since 1980 has mirrored the evolution of technology and innovation policy, beginning with a preoccupation with large scale pre-competitive, usually collaborative, R&D programs and broadening the range of instruments employed through an increasing focus on enhancing the environment for innovation and technology transfer. The evaluation of collaborative R&D programs has shown that it is often difficult to deliver the type of information about return-on-investment which some policymakers desire. Such evaluations have in fact demonstrated that socio-economic effects may be manifested not only through competitiveness and exploitation/market-related effects but also through individual and organisational learning effects which are hard to quantify. The trend in this area is now towards a portfolio approach, rather than focusing on individual projects; towards greater use of performance indicators; and towards a convergence between the activities of ex-post evaluation and continuous monitoring.

6. The evaluation of opportunity-enhancing innovation policies involves evaluating institutions concerned with implementing technology and innovation policy as well as technology transfer measures more specifically. Reflecting concerns about the role of public sector research institutions, and in particular the parallel desires to make them more commercially oriented and to improve their linkages with innovating firms, evaluations have involved attempts to ensure the continuing legitimacy of “research operators”, systematic examinations of the value of informal contacts between laboratories and industry, and analyses of the performance characteristics of firms that take advantage of diffusion-oriented measures such as technology centres or “extension” programs.

7. Georghiou concluded by discussing some general issues that cut across evaluation experiences. He noted that the precise institutional framework for evaluation is of less importance than its functionality, and stressed the importance 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. He also pointed to a number of tensions in the practice of evaluation. One is between formalised monitoring approaches on the one hand and the findings that suggest that the individual project is not an appropriate unit of analysis on the other. Another is in the understanding of what evaluation can deliver, exemplified by differences of view on whether cost-benefit approaches or returns on investment can be calculated with sufficient accuracy to justify their use in the selection of policy instruments. A third relates to a strict interpretation of additionality, as opposed to one focusing on behavioural additionality. Finally, he stressed the general lack of cross-cutting evaluations, comparing the relative effectiveness of policies. If the emphasis in innovation policy is now upon its systemic characteristics, there would appear to be a need for the findings of evaluations to be similarly marshalled in order to understand the interaction between policies and the net effects of intervention.

Session 1. Evaluation of financial support to industrial R&D

8. *Ralph Lattimore* (Productivity Commission, Australia) examined the impacts and policy lessons from fiscal incentives to R&D in Australia, one of the most generous such measure among OECD countries. His paper developed a cost-benefit framework for evaluating the tax concession, examining the most important facets which determine whether the concession is welfare enhancing, and drawing lessons from the design and functioning of the incentive. The social cost-benefit framework developed combines qualitative and quantitative indicators of the impact of the concession and of the private and estimated

social benefits from the induced R&D. In addition to estimating the R&D induced by the tax concession, the methodology involves incorporating estimates for the spillovers benefits of the scheme, as well as costs such as the marginal excess burden of taxation, leakages to foreigners and compliance costs.

9. On the basis of this methodology, the evaluation of the R&D tax incentive suggests that the program generates overall welfare benefits. At the same time, the cost-benefit analysis points to enduring weaknesses, and in particular the problem of relatively low inducement of additional R&D, and the cost of transfers to foreign companies. The paper also discusses some options and introduces some design criteria which may help develop better R&D support programs. Targeting the source of the problem is one such design criterion. From an economic perspective, the rationale for an R&D incentive are spillovers associated with induced activity; yet many R&D tax incentives do not target induced R&D nor spillovers. Other design criteria are selection issues (does the program use eligibility criteria and incentives in a way to select the 'right' firms and projects?); inducement (does the program change the behaviour of its recipients?); avoiding strategic behaviour/distortions (does a program have undesirable effects on incentives, such as increasing the risk of moral hazard?); avoiding unforeseen government liabilities (tax, legal); and encouraging administrative efficiency, accountability, transparency and policy learning.

10. The paper by *Henri Capron* and *Bruno van Pottelsberghe* (University of Brussels, Belgium) had two objectives. The first was to review the strengths and weaknesses of existing tools for evaluating business R&D policies: peer review, matrix and systemic approaches, financial methods, technological forecasting, and quantitative indicators. The paper suggests that the choice of method should be based on prior identification of the issues to be investigated. Most of the methods are likely to be complements rather than substitutes, and each is able to provide relevant additional information in the evaluation process. Thus, to increase the credibility of evaluation results, a number of alternative methods should be used to consolidate the foundations of policy recommendations. Econometric methods are particularly appropriate for addressing specific issues regarding both the direct and indirect impact of policies. The paper offers a conceptual and econometric framework for systematic impact assessment, integrating various levels of economic analysis (micro, meso, macro) and policy (project, program, system).

11. The second objective of the paper was to provide, by reference to the suggested integrated assessment scheme, some policy implications that could emanate from quantitative evaluations of the effectiveness of R&D subsidies in the seven major industrialised countries. The evaluation procedure is implemented both at the aggregate manufacturing level and for 22 individual manufacturing industries. The empirical implementation comes within the scope of two categories of studies. First, the authors test whether R&D subsidies have a direct impact on productivity growth and find, contrary to many results in the literature, that it is difficult to distinguish between the impact of private R&D and that of public R&D on productivity growth. Secondly, they investigate the stimulation effect of these subsidies on private R&D investments and find that the impact varies across countries and industries. One of the main findings is the impact of policy persistence: the more a subsidisation rate is unstable, the less an increase in R&D subsidies is likely to stimulate private R&D investors.

12. The main purpose of the paper by *Arild Hervik* (Norwegian School of Management) was to review the experience from a modular, comprehensive evaluation system in Norway as applied to three generations (1978-82, 1984-89, 1990-95) of User Oriented Research programs (UOR). The UOR programs involve support for research granted directly to companies, rather than through R&D institutes, and with a focus on developing a more efficient innovation and technology infrastructure where collaborative research is important. The evaluations were based on telephone interviews with more than 1100 R&D projects. They focused on assessing economic impacts through a cost-benefit framework, but without expecting to come up with a single figure to show rate of return. Their most important conclusions were that while there was great uncertainty regarding financial yield and profit, and despite low additionality, the UOR programs were judged to have had considerable long-run impacts (including

external effects) due to firms' investments in development of knowledge, R&D co-operation and creation of networks. Finally, based on the experience of these evaluations, the author suggests that there is a need to develop more competitive allocation procedures to improve quality of proposals at both program and project level, and to improve funding by a better adaptation of programs to target groups and suppliers of R&D-services.

13. The experience with the evaluation of the R&D allowance scheme in the Netherlands was reviewed by *R. Dorsman* (Ministry of Economic Affairs, Netherlands). The scheme gives a reduction of the wage tax and social insurance contributions to be paid by the employer. Against the background of previous evaluations, the paper outlined the methodology of a pilot evaluation whose aim is a more systematic approach to the issue of measuring effectiveness. This involves in-depth interviews, surveys and econometric techniques based on micro-level databases covering recipients and non-recipients of the R&D allowance. The proposed evaluation scheme would address the question of the allowance's impact on the scale of the R&D effort, followed by an examination of the impact on the output of R&D (by looking at indicators such as patents), and finally on economic output (sales, profit, jobs). The participation of SMEs in the scheme, the extent of contact research, the impact on self-employed persons (which can also use the scheme), and the effects of rejection of an application for the R&D allowance would also be addressed.

14. The discussant of the papers, *Brownyn Hall* (UC Berkeley, United States), suggested that in assessing R&D fiscal incentives, the focus should be on measuring the social welfare contributions of a program. This implies the need to focus on the measurement of spillovers, which is the welfare increase due to having the program. Intermediate quantities such as the additionality induced at the firm level are only interesting to the extent that it is not easy to directly measure the social return to R&D spending. In terms of handling the question of risk when evaluating R&D support policies, she cautioned that the distribution of returns from R&D is extremely skewed at the project level; this makes it difficult to use standard techniques to look at the overall risk profile of programs. Finally, 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. In this respect, a number of the papers in the session did not seem to pay attention to the characteristics of firms that did not receive government incentives.

15. The ensuing discussion focused on whether evaluations should deal exclusively with measuring the social returns of government support. Many participants stressed that improving budget allocation and program management also necessitates looking into the process and performance of different policy instruments. Given the high variance of returns in different R&D projects, detailed case-studies are important to see what works and what does not. It was however acknowledged that ultimately what is important is the economic benefit that additional R&D produces, not the amount of additional R&D generated in itself. On the issue of diminishing returns to evaluation, it was suggested that finite resources for evaluations should be concentrated on what is (methodologically) new or significant from a budget perspective. It was also noted that while politically less attractive, long-run evaluations (even after programs have been discontinued) often produced unexpected insights useful for future policy. Finally, participants cautioned against simplistic evaluations of the employment impacts of R&D programs, which tend to look only at gross effects, and ignore displacement and macroeconomic adjustments.

Session 2. Evaluation of large technology programs

16. The paper by *Patrick Cohendet* (BETA, France) focused on the evaluation of industrial effects from European space programs, as an archetype of technological programs with a highly specific end product. He suggested that for the evaluation of the "social" effects of these programs relating to the use of the final product realised (e.g. effects of meteorological satellites on improved quality of weather

forecasts and hence on agriculture etc.), cost-benefit analysis seems suited. In contrast, the estimation of the “industrial effects”, both direct and indirect, are best captured by structured, well-defined interviews. Direct industrial effects arise out of contracts performed within the set framework of the program (with designers, constructors, suppliers of services and end-operators). Indirect industrial effects go beyond the scope of the contract objectives and cover benefits such as creation of knowledge, transfer of technology, building up of competencies, improvements of quality, acquisition of new processes, etc., that contracting bodies derive from their participation to space programs and are able to deploy elsewhere.

17. Following this distinction, the paper presented attempts to estimate the quantitative and qualitative indirect industrial effects/spin-offs from space programs to the contractors of the European Space Agency (ESA). Based on first-hand data, obtained by direct interviews with the managers of the ESA contractors, the focus was on evaluating technological and commercial effects, as well as the effects on work organisation and skill formation. The method used also addressed the issue of managing spin-offs, in the sense of looking closely at their mechanisms and the factors determining their success or failure. Basically, this work confirmed the existence of positive spin-offs, but the methodology applied did not make it possible to judge the long-term effects of ESA programs on the whole economy. The paper suggested in effect that it is important to develop an analytical framework able to take into account the variety of spin-off phenomena and the complexity of the channels by which they have an impact on economic activity, considering them as part of the factors shaping economic development instead of treating them as isolated phenomena.

18. In a joint paper with John Scott, *Albert Link* (University of North Carolina, US) looked into the evaluation procedures established at the National Institute for Standards and Technology (NIST) in the United States and outlined key characteristics of a best practice evaluation program based on the examples of the evaluation of the eight research laboratories of NIST, and the Advanced Technology Program (ATP) run by NIST. The paper suggested that a best practice evaluation program, as applied to a technology-based institution, has a number of functions. It must inform internal management of the importance of evaluation both as a means to enhance managerial effectiveness and also to document the value of the institution. The institution must initiate a commitment to evaluation by articulating an evaluation strategy; it must set forth guidelines regarding how to implement the evaluation strategy; and the institution must establish standard metrics to interpret the application of evaluation guidelines.

19. The paper reported that NIST has adopted two measures as evaluation metrics: an internal rate of return, and a benefit-to-cost ratio. Benefits are quantified in terms of costs avoided, and their estimation involves an implicit government-to-market return, since program beneficiaries are asked the counterfactual question of what the private company would have had to invest in the absence of the NIST support to reach the same level of technological capability. This yields a useable cost-benefit measure, but without capturing the wider benefits of the research supported. Evaluation efforts have not succeeded yet in establishing the collection of micro-level data which would be a prerequisite for more elaborate social cost-benefit evaluation techniques. Overall, the authors suggest that both the evaluation program of its eight research laboratories and the various evaluation activities around the ATP are considered successful, insofar as having established a general acceptance of evaluation within the organisation, a commonly accepted set of evaluation guidelines and a standard metric to be applied in measuring social returns of the activities of the laboratories and the joint-ventures within ATP.

20. *Mariko Sakakibara* (UCLA, United States) applied a combination of a questionnaire-type analysis with econometric techniques to evaluate the effects of government sponsored R&D consortia in Japan in a sample of firms which participated in one of these programs. The encouragement and co-funding of such consortia has been a rapidly growing area in technology and innovation policy because of the apparent success of some early Japanese programs (namely the very large scale integration --VLSI--

program for the semiconductor industry). Despite the growing popularity of this instrument, there are only few empirical studies that try to produce a comprehensive evaluation of these initiatives.

21. The combination of these two evaluation techniques proved useful as a cross-check for the subjective answers from the questionnaire, allowing a better interpretation of the results of the econometric analysis. The questionnaire results suggested that the direct (and measurable) economic effects from the participations were perceived to be modest, and participants did not, in general, estimate their participation to be critical to their competitive position. Rather, the perceived benefits were of an intangible nature such as researcher training and increased awareness of R&D and technological developments in general. Econometric analysis confirmed this by pointing to an increased level of R&D spending in participating firms, an increased “research productivity” (patents per R&D investment), and higher knowledge spillovers within consortia contributing to these positive effects. However, while the cost of the participation was captured on the side of the participating private firms (they incurred additional costs, but these were outweighed by the benefits), there was no “cost-to-government” estimation, so that a social cost-benefit-analysis could not be undertaken.

22. *Terttu Luukkonen* (VTT Group of Technology Studies, Finland) summarised the experiences with evaluation in Finland and pointed to the fact that, in the course of increasing professionalisation of evaluation, the problem of vested interests becomes a factor to be taken into account. As the extensive use of external evaluators from other countries is only a partial remedy to this problem, because of the structural dependence on a few agencies commissioning evaluations, she put forward the idea of a “code of conduct for evaluation”, which would set some out some ethical principles and basic requirements for those commissioning evaluation exercises, trying to avoid the most crude examples of advocative use of evaluation for predetermined purposes.

23. The discussant to the session, *R. Lyon* (Office of Management and Budget, United States), stressed that the issues relating to the evaluation of large technology programs centre around efficiency; nevertheless, distributional issues, such as whether subsidies to certain firms lead to displacement of activity, also play a role. He suggested that in the Cohendet paper, the focus on spin-off effects of space programs, while giving interesting insights, left out the real externalities (the real rationale of government support), which occur outside the firm itself. On the Sakakibara paper, Lyon felt that the fact that firms did not attach much importance to tangible consortia outputs (such as commercialisation of a product, or increase in patent applications), was very worrying from a policy perspective, as it cast doubt on the cost-effectiveness of supporting such research consortia. On the paper by Link, Lyon suggested that the high cost/benefit ratios found in the paper beg the question of why these were not captured by private agents. If there is a comparative advantage for government in the area of standard-setting (for example in calibration services), maybe such services should be better priced and costs recovered. Lyon also suggested that the Luukkonen paper raises an issue with respect to openness of evaluations: evaluations have a sanctioning capability, but to make them acceptable they need to be complemented by positive incentives.

24. In the discussion a number of participants suggested that it is important to understand the mechanism of the program under evaluation; a reliance on quantitative techniques alone often is not enough. In terms of R&D collaboration in particular, it was suggested that results are more long term; they involve learning and this makes it hard to account for results in an economic sense. Tangible outputs are not always perceived by participants; in any case, governments should not necessarily subsidise consortia with obviously high positive outcomes. It was also suggested that evaluations have shown that small and medium-sized firms learn most out of access to large technology programs; they learn less in projects of technology diffusion because they do not have the capacity for absorption. Finally, in terms of the role of government in standard-setting, it was suggested that this role was grounded on the independence and credibility that government labs conferred. Pricing goods out of government

laboratories such as NIST closer to market was desirable in principle, but in practice it was hard to implement.

Panel #1. Evaluation techniques

25. The first intervention in the discussion panel on evaluation techniques, by **Mario Pianta** (National Research Council, Italy), focused on the use of innovation surveys for policy evaluation purposes. Important information on the nature and economic impact of innovations, and on the sources and barriers to technological change, can be drawn from the results of the surveys on innovations introduced by firms, such as those of the EU-sponsored Community Innovation Survey (CIS). In addition, some innovation surveys, such as the one for Italy, have included specific questions on the importance attached by firms to various innovation policies. The Italian survey shows that a large majority of the 7,553 firms which have introduced innovations in the period 1990-1992 declared that the existing technology policies have had little relevance for the introduction of their innovations. The analysis suggests that the focus of most public policies and funds on the largest firms in the high technology sectors appear hardly adequate for supporting the much larger body of innovating firms of smaller size and active in all industries. Based on this experience, Pianta suggested that questions on the impact of government policies are introduced in the innovation survey questionnaires to be used later in 1997 in OECD countries.

26. **Amir Piric's** (Ministry of Research, Science and Technology, New Zealand) presentation concerned methodological issues arising in the evaluation of public investment in R&D. He reviewed the strengths and weaknesses of different techniques used in ex-ante project selection and in ex-post evaluations, ranging from the qualitative (peer reviews, case studies, technological forecasting) to the quantitative (financial methods such as cost/benefit analysis or portfolio models, technometrics, optional pricing, bibliometrics, econometric methods). This was illustrated by the experience of the Foundation for Research, Science and Technology in New Zealand which has been undertaking evaluations using both qualitative and quantitative performance indicators in an attempt to measure progress towards long-term scientific, economic, environmental and social goals. He suggested that finding useful indicators of research performance is perhaps the central difficulty for all evaluation approaches, as there is often a tension between what is measurable and what is interesting.

27. The intervention by **Ron Jarmin** (Bureau of the Census, United States) concerned the use of micro-level data, based on his experience with the evaluation of the MEP (manufacturing extension program) in the US. He suggested that the main justification for the use of micro-level data lies in the fact that the impact of programs is at firm or plant level; it is therefore sensible to examine the impact of these programs on firm or plant performance. In effect, evidence shows that variation of performance is much larger across plants within industries than it is across industries. More aggregate data are subject to aggregate shocks which cannot be easily adjusted for. Micro-level data allow researchers to compare performance of client plants with that of non-clients and control the performance of clients before and after they receive services such as those through the MEP. They also allow researchers to correct for selection bias (the fact that better than average plants seek out services such as MEP). Since it is not always possible to survey non-clients, data on the characteristics of plants or firms using MEP services are "matched" with longitudinal micro-level data on non-client firms existing in the US Census Department.

28. **Ralph Lattimore** (Productivity Commission, Australia) focused on the advantages and pitfalls in using social cost-benefit analysis for the evaluation of R&D programs. He suggested that such techniques force the policy-maker to make explicit the different costs and make estimates for the benefits of policy initiatives. In calculating the net social benefit of a program, the benefit side consists of calculating margins (the private rate of return on additional activities induced by the program net of the rate of return

without the subsidy) and externalities. The cost side includes the marginal excess burden of taxation (social cost associated with raising taxes in order to finance reductions in revenue); leakage of benefits to foreigners; administrative resources in running the program; compliance costs for firms; and rent seeking on behalf of firms (resources expended in lobbying for the program). The fact that spillovers are only on induced R&D but costs are borne on the whole eligible R&D makes additionality important. If it is small, the program has a large distortionary transfer element and a small spillover; therefore except if spillovers are very high, it is unlikely to generate a net social benefit. Lattimore warned against the spurious precision that cost-benefit calculations can give, given the use of “guesstimates” for key parameters (spillovers). This suggests the necessity to calculate distributions, not point estimates, and to acknowledge uncertainty by calculating probabilities that the program under evaluation will generate a net benefit.

29. The intervention by *J-P. Meurice* (Ministry of Industry, France) focused on the experience with qualitative evaluation in France in the context of programs such as the GPI (large innovation programs). He suggested that qualitative techniques (use of surveys with closed or open questions in questionnaires and in-depth interviews) are indispensable in order to be able to validate quantitative information and to understand the mechanisms and significance of observed results from programs. In this sense, it is important to take into account in evaluations both what is expected by program users (the demand side) and what is offered (the supply side). In the evaluation of GPI, questions such as the place of the program in the strategy of a firm cannot be addressed by quantitative techniques. Qualitative information can however be re-organised so that it can be used in a statistical sense. Broadly speaking, the results of the evaluation of the GPI suggested that these have been successful technologically but less so commercially and economically.

30. One of the questions posed in the discussion was to what extent evaluations mainly pursue questions of internal efficiency of recipients rather than measuring wider social effects. In the case of Italy, it was suggested that the innovation data allow researchers to look at the overall economic effect of policies, including wider social benefits (though they cannot separately examine the effect of different policies), but not to explicitly measure spillovers. Similarly in France the evaluation of policies to help SMEs have looked at macroeconomic effects, but without explicitly isolating/estimating spillovers. Another issue raised was that comparing returns (cost-benefit ratios) of different programs gives only part of the picture; in Australia, many programs supporting agricultural research have higher returns than programs supporting manufacturing research; yet in the face of broader structural change in the economy, it may still be preferable to switch public funding away from the former and towards the latter.

Session 3. Evaluation of diffusion-oriented policies

31. *Ron Jarmin* (Bureau of the Census, United States), in a joint paper with J. Jensen, reviewed recent efforts to evaluate the effectiveness of a specific diffusion-oriented program, the Manufacturing Extension Program (MEP) of the National Institute for Standards and Technology (NIST) in the United States, and attempted to draw some general lessons for evaluation methodology. This program offers via Manufacturing Technology Centres (MTCs) technical and business assistance services to small and medium-sized firms. Its premise is that small and medium sized manufacturers have systematically less access to technological information than do their larger counterparts. The authors suggest 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 MTCs are supposed to improve SME performance; yet it is important to have an understanding of this environment in order to optimally design, provide and evaluate program services and to ensure that program objectives are not at odds with those of the client SMEs.

32. After reviewing the advantages and drawbacks of different evaluation methodologies (experimental methods, use of administrative data, case studies, client surveys, and econometric analyses of non-experimental data), the authors conclude that econometric analyses based on non-experimental data offer researchers and policy makers the best opportunity to assess the overall performance of programs like manufacturing extension. Such techniques use data from administrative sources and from surveys of MEP participants in combination with pre-existing micro-level data sets like the Longitudinal Research Database (LRD) of the US Census Bureau to look at the performance characteristics of client and non-client firms (such as growth in value-added per worker or relative plant survival rates). They explicitly recognise that programs such as the MEP are addressed to heterogeneous populations, and hence analysis of their effects has to look at the plant level. Based on their experience, the authors stress the necessity of such micro-level analysis and the need to build up data bases that allow for such analysis.

33. In a closely related paper, *Philip Shapira* of the Georgia Institute of Technology in the United States (jointly with Jan Youtie) compared the advantages and shortcomings of different evaluation methods (customer surveys, project impact analysis, cost-benefit analysis, controlled studies) by looking at the various evaluations of the Georgia Manufacturing Extension Alliance (GMEA), i.e. the Manufacturing Extension Program at the level of the state of Georgia. They suggest that there are a variety of tensions in implementing evaluations of technology deployment programs. These include issues of reconciling evaluation designs with the varying needs of program sponsors, service providers, and customers; accounting for the differential impacts of particular kinds of services; incorporating methods that can support program learning and improvement as well as address issues of program justification; and reconciling desired information needs with resource and information availability constraints.

34. The authors suggest that in understanding these issues, it is helpful to distinguish between two different purposes for which evaluation analyses can be used. The first is program justification and rationalisation, where the aim is to analyse the effectiveness of a program and to provide information to guide resource allocation. The second purpose is program improvement, where evaluative information helps to guide attempts to improve program quality, responsiveness and effectiveness relative to resources and needs. Accordingly, different actors value different evaluation techniques differently. State level government is more interested in program performance with the purpose of improving the functioning of the program, while at the federal level the main interest is in justification of the program by proving its broader economic and societal benefits. The interest in cost-benefit analysis is thus much greater at the federal level, while client valuation surveys are the most highly estimated source of information at the state level. In general, the authors find that there seems to be a general weakness in demand for sophisticated and complex evaluation methods; at the same time, the complementary information provided by different methodologies suggests that no one single method by itself is adequate to the task of evaluating all important aspects of a complex program.

35. *Spyros Arvanitis* (Federal Institute of Technology, Switzerland), presenting a joint paper with Heinz Hollenstein, outlined a procedure by which diffusion programs might be evaluated, designed for the Swiss program for the Promotion of Advanced Manufacturing Technologies (AMT). The methodology mainly consists of estimating an adoption equation which, in addition to the main explanatory variables as proposed by the theory of technology diffusion, takes account of variables depicting policy intervention (participation in program, type and intensity of promotion) as well as interactions between policy and basic explanatory variables (e.g. firm size, specific diffusion barriers, before-promotion adoption intensity, etc.). In order to address the issue of causality between adoption and promotion, they suggest estimating, simultaneously with the adoption equation, a second equation with “promotion” as the dependent variable, and where a set of firm characteristics that do not (or at least not overwhelmingly) determine adoption are used as regressors.

36. The authors acknowledge that this type of analysis faces a number of practical difficulties. The main is data availability; although they are able to use (mainly qualitative) firm-level data collected for this specific purpose in the course of the Swiss Innovation Survey 1996, this set of data does not allow for longitudinal analysis (some variables are one-off, others are collected in retrospect and therefore of doubtful accuracy). In addition, the explanatory variables are not differentiated for the different AMT elements; nor does the methodology account sufficiently for “soft factors” relevant to the adoption of AMT (e.g. the organisational environment). As a result of such difficulties, the authors are unable at this stage to present empirical estimates from their methodological approach. For practical purposes, they foresee a “mixed evaluation design”, using case studies and in-depth-interviews alongside econometric work. Nevertheless, they recommend to fully exploit the hitherto unexploited potential of model-based econometric work in the future and to collect appropriate data bases accordingly.

37. **Erik Arnold** (Technopolis, UK), presenting a joint paper co-authored by Ken Guy, took a more sceptical stance on the usefulness of evaluations of diffusion-oriented programs using a single approach, and more particularly of cost-benefit approaches to evaluation. Based on the experience with evaluations of a large number of programs, he discussed the advantages and disadvantages of various evaluation tools. He argued for a context-dependent use of the different evaluation tools (e.g. bibliometrics is useful in the context of the evaluation of scientific performance, but not much outside this area), as well as for a choice of evaluation tools reflecting the needs and interests of the users of evaluation (e.g. summative as opposed to formative evaluations). He suggested that cost-benefit analyses normally run into the difficulties of badly capturing external effects and problems of multiple causality of outcomes, which are the most interesting part of a program evaluation. Even where cost-benefit analysis can come up with some reasonable estimates, there is the problem of comparability of measures between the evaluations of different programs, which prevents it to be used as a means to choose from a portfolio of programs with the highest rate of return.

38. Instead, the paper suggests a more system-oriented evaluation approach. Such an approach would not only use a number of different methodological tools, but would also stress more qualitative aspects in evaluation. For example, the analysis of the history of a program and the intensive use of case-studies, especially in experimental programs, provide much information about the quality of effects and interactions in diffusion-oriented programs which can only be obtained from this type of analysis. The paper suggests that single-approach evaluations might be downright misleading, and putting too much emphasis on quantitative cost-benefit techniques, while useful as a measure of cross-checking, might miss the essential qualitative effects of new initiatives.

39. Discussing the papers in the session, **John van Reenen** (Institute for Fiscal Studies, United Kingdom) suggested that it is first important to be clear about the rationale for diffusion policies. The fact that diffusion is important for productivity and growth is not in itself sufficient; policy needs to be focused on compensating for some market failure, such as improving the information flow to small firms. He suggested that there are two issues in evaluation: how to improve programs; and how to measure effectiveness and justification of policy. Case studies and client surveys are very useful for evaluating efficiency of service provision, but less good at justifying effectiveness. Social experiments using a random sample of firms are best for measuring effectiveness; but in practice econometric models are the best analytical tools, despite difficulties in controlling for external effects and selection bias. Van Reenen also noted that the use of econometrics for measuring effectiveness of programs is independent from undertaking a full cost benefit analysis. He concluded by emphasising the importance of collecting quantitative data for evaluation analyses, and particularly panel data on firms.

40. In the discussion that followed, it was suggested that identifying positive direct impacts of programs may be enough for justifying expenditures, and in that case it might not be necessary to examine further the existence of social benefits. But other participants felt that there is in fact a two-step

procedure, with the first step consisting of identifying positive private returns; if there are none there is no sense proceeding further. The existence of private returns is however just a necessary condition for programs, not a sufficient one; the lack of social returns or the cost of raising taxes to finance expenditure may still militate against government intervention. Nevertheless, it was noted that justifying policies to policy makers is different than justifying policy in a welfare economic framework. Other participants noted that in practice, identifying such social benefits in diffusion policies involves a dynamic analysis that looks at the development of new capabilities, and of learning. Finally, it was suggested that thinking about objectives in technology policy is useful because it points to different ways (instruments) to get to results. Different policies are appropriate in order to deal with a problem of information provision that they are in order to reduce market concentration which allows firms to charge high prices for products and blocks the diffusion of technologies throughout the economy.

Panel #2. The role of institutions in policy evaluation

41. **Luc Durieux** (Evaluation Unit, DG12, European Commission) reported on the experience of the European Commission in the evaluation of research and technological development (RTD) programs. Since the early 1980s, this effort has evolved in an attempt to systematically support S&T policy formulation and decision making. The present evaluation scheme is based on two activities: for each RTD program, it involves a continuous monitoring with the assistance of experts external to the Commission and, with multi-annual intervals, a five-year assessment conducted by external experts. The five-year assessment is carried out prior to presenting a proposal for the next RTD Framework Program, so that it combines an ex-post evaluation of the previous program, a mid-term appraisal of the ongoing one, and recommendations for future activities. This timing is aimed to ensure feedback from evaluation into strategy formulation and allow the possibility to consider new trends. In terms of methodology, evaluations are based on independent expert panels, interviews, questionnaires, and core indicators. They address issues of relevance (whether the initial objectives are still valid against new S&T developments and socio-economic conditions); efficiency (whether the objectives have been pursued in a cost effective manner); and effectiveness (whether the initial objectives have been achieved, or, for longer term strategy and objectives, if progress is sufficient).

42. **Jean-Eric Aubert** (DSTI, OECD) reviewed the experience of the OECD in evaluating national policies related to research and innovation; more than 60 such reviews have been completed since the early 1960s. Mirroring the evolving understanding of the S&T system, the reviews have changed over time, from covering science policy initially, to technology policy and innovation policy more generally. In terms of methodology, the reviews have been based on the preparation of a background document of descriptive nature by the country under review exposing S&T systems and policies and then on an examiner (or evaluation) report prepared by a small group of high level foreign experts co-ordinated by the OECD Secretariat, focusing on main problems and suggesting policy recommendations. Experience has shown that often the process of the review can be more important and more useful than the tangible results of the review in the form of recommendations. To increase the influence of such evaluations, Aubert suggested that it would be necessary to make the reviews more a self-assessment in which the actors themselves discover what facilitates or, on the contrary, hinders the innovation processes.

43. The intervention of **M. Stampfer** (Federal Ministry for Science and Transport, Austria) centred on the institutional set-up for policy evaluation in Austria, in the context of a science and technology policy characterised by a high dominance of the university sector, comparatively low industry involvement and a set of fragmented planning institutions and funding organisations. He suggested that one of the most important issues is the development of procedural guarantees and the building-up of a broad evaluation culture: awareness, regularity of evaluation exercises, obligations to be evaluated, publicity of results, use of external evaluators, third party as contractor, more institutional evaluations etc.

Another is active internationalisation of the evaluation process, through involvement in EU evaluation exercises, use of informal networks, and building up of internationally competitive Austrian evaluation expert groups.

44. *Annette Birch* (Agency of Trade and Industry, Denmark) reported on the experience in Denmark with the evaluation of 17 approved technological service institutes (GTS) during the 1995-97 period. The GTS are independent and self-governing institutes receiving government funding with a view to fostering industrial development, and in particular to help small and medium-sized firms in building up skills, developing new services, and participating in codification and standardisation work. The evaluations were aimed at establishing the role of the institutes in society and industrial policy in general, and the adequacy of their activities and skills in particular. They were based on peer reviews and the evaluation reports remained confidential.

45. In the discussion, it was pointed out that one of the main functions of evaluation is to make institutions involved more responsive and flexible. Developing new methods for budget allocation such as internal bidding mechanisms for funds introduces some aspects of market mechanism, but can also destroy collaborative behaviour. This suggests a two-stage process, whereby broad priorities and separate budgets (e.g. in different technology areas/sectors) are first set, and only then are the best projects within each area chosen. Another issue raised was the widening coverage of peer-review methods, which suggests that decision-making in panels which are no longer composed of specialists alone needs to be supported by complementary studies (surveys, econometrics). Good evaluation teams also need to be multidisciplinary, covering an understanding of the technology concerned, of statistics and economics, and of the policy process. The question of country specificities and general principles in institutional arrangements was also discussed, and it was felt that while practical arrangements are country specific, basic principles/challenges in evaluation are not (quality, independence, etc.). This means that a number of different institutional “pillars” are necessary: bureaucratic, scientific, and political.

Session 4. Strategic evaluation and new trends and approaches

46. The presentations in this session described the developments in evaluation practices in countries with a relatively long-standing tradition in evaluation but different organisation of the process (e.g. in terms of centralisation/decentralisation or the respective weight given to the evaluation of scientific performance and economic outcomes). *Stefan Kuhlmann* (Fraunhofer-Institute Systems and Innovation Research, Germany) described the development of evaluation practices in Germany as reflecting the broad general trend from peer-review based evaluation of science and scientific institutions towards increased awareness and attempts to measure socio-economic impacts. He summarised the main lessons learned from a recent attempt to take stock of the accumulated evaluation experience: the need for more systematic scrutiny of underlying assumptions of economic or technological problems which make state intervention necessary; the necessity for evaluation practice to be developed in the context of the specific conditions in the relevant promotion field; the lack of horizontal evaluation studies of different but related policy initiatives; the existence of tensions between the detached (as required) evaluation activity on the one hand and program managers on the other; and a need for a more intensified user orientation of evaluation studies.

47. While recommending a more ‘pro-active’ role for evaluations with explicit policy recommendations and a greater publicity of results to be fed into public policy debate, he stressed that findings from policy research as well as from experience demonstrated that the role of the state and its ability to influence scientific and technological developments has changed. Rather than being the centre of policy planning, the state in modern societies has to redefine its role as a moderator in negotiating processes between various actors with different behaviour, logic and power, itself being just one among

these actors. This change in role has to be taken into account when formulating the task for evaluation. Rather than used just as an information tool for policy planning, analyses and assessments in the framework of an evaluation procedure should take multiple actor perspectives into consideration, as regards both methodology and content. They should be brought explicitly and visibly into the communication of negotiating systems and policy networks and if possible institutionalised, functioning as a moderation medium depicting the actors' problem perceptions and reflecting their learning processes.

48. **Phillipe Laredo** (Ecoles des Mines, France) outlined the institutional structure and evolution in the evaluation system in France (the “guarantor model”), which from an initial focus on scientific institutions, saw its scope widened to include programmes and institutions of technology and innovation policy. In terms of institutions, this shift was marked by the establishment of the ‘Comité national d’évaluation de la recherche ‘ (CNER) in 1990 supplementing the ‘Comité national d’évaluation’ (CNE) dating from 1984. While the latter is responsible for the evaluation of universities, the former deals with all other research institutions and R&D activities which receive state support from the civilian budget.

49. One aspect of the institutional set-up highlighted in the paper is long evaluation cycles of CNE and CNER which would not allow them to reach their originally set targets of periodically re-evaluating all institutions and programs. Although CNER makes regular use of external expertise, and other institutions develop in-house capacity for evaluation, it seems to have too little resources. Another major result of the experience of both the CNE and the CNER seems to be that the evaluations carried out had an impact as long as the evaluated institution could implement the recommendations on its own. On the other hand, the influence of evaluations on policy making or the rearrangement of the institutional framework was limited. This absence of a “take-up mechanism” could principally be tackled in two ways: a debate on the legislative level and an increased publicity of the evaluation reports to enable an informed, and broader, political debate. On the side of the evaluating institutions, recently a “pro-active” approach to evaluation has already been taken up, which would allow for such a use of evaluation results.

50. **George Teather** (National Research Council of Canada) and **Robert McDonald** (Industry Canada), analysing the developments of evaluation in the last 20 years in Canada, shared the diagnosis by Laredo that past evaluations had their effects mainly in a localised form, that is directly in the institutions evaluated, but less so on a broader policy level. This was partly due to the fact that evaluation in Canada takes place in a very decentralised manner. In Canadian evaluation practice, an evaluation framework is typically established at the start of a program. The framework identifies its performance expectations and specific, detailed requirements for ongoing performance monitoring (to serve management decision-making) and future evaluation. This includes performance measures, data collection, analysis and reporting, and evaluation approaches. In the evaluation study systematic research methods are used and, typically, three main issues are examined: program relevance, success, and cost-effectiveness.

51. Latest initiatives try to improve the evaluation process with an eye to its strategic impact on science and technology policy. A prominent example is the “Industry Portfolio Science and Technology Action Plan”. The organisations of the so-called “Industry portfolio” (institutions like Industry Canada, Statistics Canada, Copyright board, Research councils, Canadian Space agency etc.), building on a 1996 government white paper “Science and Technology for the New Century - A Federal Strategy”, have sought to develop guidelines covering all the different objectives pursued by the heterogeneous organisations represented: general advancement of scientific knowledge, improvement of innovative behaviour in firms, general support to the innovation system by providing public goods, and improvement of mechanisms for the management and governance of science and technology policy. While general agreement among the participants has been reached about the general evaluation framework, the implementation of the concrete performance measures will be the next challenge. The ultimate goal is to enable program managers to base their decisions on up-to-date information as well as to allow the federal government to oversee all its S&T initiatives combined.

52. **Luke Georghiou** (PREST, United Kingdom), the discussant in this session, framed his remarks around the question of whether best practices exist independently from the national institutional framework. He suggested that the papers presented make two important points: that history matters, in the sense that current evaluation practices are the outgrowth of a specific historical development; and that evaluation is primarily a social process, as it involves interaction of individuals, organisational beliefs, practices, and routines. In the case of France, he noted that while in principle the “guarantors” system has

a strategic element built-in, in practice it misses the systemic interactions in the innovation system. In Germany, he suggested that the main challenge is whether moderation in a decentralised structure is a desired goal or just an accommodation of the system; and in Canada, which has done a lot to systematise evaluation, he wondered on the benefits of working via a “portfolio” framework as against the constraints that might impose. He suggested that a broader question is whether evaluation is fated to be an instrument always for incremental change (how to improve a program), rather than an instrument for radical change (should we have a program at all). In order to achieve the more radical perspective evaluation may have to be embedded in a wider strategic intelligence system, with a linkage to foresight and technology assessment.

53. The discussion focused on what constitutes strategic evaluation. To some participants, the question of strategy is one of positioning of policy as a whole; the rest is about operational effectiveness. Certain participants felt that evaluation should be limited to learning from the functioning of programs and institutions; broader issues should not be addressed in an evaluation framework. To them, professional evaluation communities cannot take the role of public debate; they can simply provide information on what each programme does and this learning input can be put into debates. To others, strategic issues relate to the interaction of different programs, and how it is possible to evaluate the effectiveness of alternative instruments that can be used in order to achieve a given policy goal (such as grants vs. tax-based incentives for promoting R&D). Evaluation operates in different levels, from the project/program level to the wider policy level, and policy evaluation consists of examining the role of government in ensuring that the various pieces of the puzzle which constitute the innovation process fit together.

Policy roundtable: the feedback of evaluation results into policy design

54. Opening the policy roundtable, *John Ryan* (Department of Industry, Science and Tourism, Australia) suggested that in thinking about evaluation, both supply and demand issues need to be addressed. On the demand side, it should be borne in mind that evaluations often derive from budget constraints, and sometimes are simply used in order to stop a program; there are multiple clients, each with its own agendas. On the supply side, experts do not come to evaluations without any a priori beliefs; they express viewpoints based on their own mental models of what constitutes appropriate government action.

55. As a client of evaluations, *Dominique Bureau* (Ministère de l'économie, France) stressed that the interest in methodology issues stems from the fact that it is very difficult to use simple rate of return approaches to summarise the impacts of innovation programs. He noted that the results of evaluations can be either of a quantitative or a qualitative nature, even if the evaluation itself is quantitative (such as calculating a rate of return); evaluations in other words are useful for improving programs. He thought that the material assembled in the conference was reassuring, as it showed that there are many good practices which allow rigorous reasoning. The development of econometric techniques was useful, but more needed to be done in understanding the impact of diffusion programs. Finally, he cautioned that evaluation illuminates choices, but does not substitute for decisions.

56. Speaking from the perspective of the European Commission's experience with evaluation, *Jean Gabolde* (DG12, European Commission) suggested that the evolution over time from an ad hoc to a more systematic evaluation approach focusing on policies, operators and projects has followed the evolution of S&T policy. Evaluation is important but should be put in context of all the parameters that define policy-making; its impacts are higher when its results are interpretable by decision makers. Demand for evaluations will increase if evaluations are shown to be useful for policy-makers; there could be in this sense a trade-off between academic and policy-relevant research. The evolutions of the Japanese science and technology system were described by *Mr. Wakimoto* (MITI, Japan), who noted that

evaluation will be an important part of the new system now being set in place. MITI in effect is setting up in July 97 an R&D evaluation Division which will draft evaluation guidelines on the evaluation of projects and institutes.

57. **Malcolm Bradbury** (DTI and Treasury, United Kingdom) noted that evaluation is part of a political and management process; it does not take place in a vacuum. Based on UK experience, he suggested that there are a number of important lessons. Publication of evaluation reports is one; once in the public domain it is more difficult to ignore results of evaluations, hence in DTI there is a presumption in favour of publication. Putting a formal obligation on those responsible for the policy to respond to the evaluation report is another; otherwise policy divisions only pick out points they like out of evaluations. Another issue concerns contrasting and comparing the results of evaluations; there is a danger of evaluating smaller and smaller programs. There is instead a necessity to look at a whole block of evaluations to see what lessons emerge, something the UK system is less good at.

58. In the US, **R. Lyon** (Office of Management and Budget, United States) suggested that while evaluations are widely used in the policy process (there are many examples where good analytical work can make a difference), they are not and could not be decisive, since policy involves trade-offs and values. They are also not decisive because often evaluations are not good enough, partly because of uncertainty in the impact of many programs. In effect, many policy decisions are based on intuition and first principles. There is a need for more and better evaluations, even though there may be diminishing returns in evaluating certain type of programs. One of the most difficult issues is comparing the impact of different spending programs (i.e. using different policy tools to get to a given objective); work on this in the US is only now beginning, and benefit-cost analysis is an important tool in this respect. Lyon also thought that in the US the institutional framework for evaluations was probably adequate; the Government Performance and Results Act (GPRA) focuses extensively on performance measures; but they are not a panacea and in reality there is an arbitration between analytic discoveries and political decisions.

Concluding remarks and future work

59. In his concluding remarks, DSTI Director **R. Nezu** started by recalling the main aims of the conference: presenting and comparing the experience in policy evaluation practices in countries; and discussing “best practices” in different policy areas and using different techniques. To achieve these aims, a number of issues were addressed: the lessons to be learned (positive and negative) from the evaluation of different types of policies/programs in innovation and technology policy; the extent to which evaluation practices (techniques, tools, institutions) are transferable across countries and policy areas or instead specific to countries, institutions and national innovation systems; and how socio-economic objectives and economy-wide effects (in particular employment) should be taken into account in evaluations

60. The papers and discussions in the conference produced a number of lessons. Chief among these is that a combination of different approaches (quantitative and qualitative) is needed in order to cover all aspects of the evaluation process; different approaches are complementary, not mutually exclusive. Trying to quantify effects is important, even if very difficult and often imprecise; it is always a good learning device. In this respect, further work is needed to identify the economic impacts of “softer” policies (networking, learning effects, etc.). Another conclusion is that identifying “additionality” as a result of government policy and what the “social returns” are is crucial in clarifying objectives and laying down criteria for evaluation. In terms of institutions, evaluations must build on the strengths and variety of the national systems of innovation in different countries 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. In terms of the coverage of evaluations, the conference showed that there a welcome shift in evaluation practices towards a more strategic (“systemic”) approach as opposed

to program-specific evaluation. Finally, another important conclusion was that more effort is necessary to address the question of how to attain given policy objectives through different instruments.

61. Against the background of these lessons and conclusions the OECD's future work in this area will have a number of dimensions. First among these is looking for best practices for designing program evaluation (issues such as articulation of objectives, criteria for assessing success/failure, institutional issues etc.), 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 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: 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.

OECD CONFERENCE ON POLICY EVALUATION IN INNOVATION AND TECHNOLOGY
Paris, 26-27 June 1997

PROGRAM

Thursday 26 June 1997

Opening addresses

Ms. Joanna Shelton, OECD Deputy Secretary-General

T. Andersson, Deputy Director, Directorate for Science, Technology and Industry, OECD

D. Shand, OECD Public Management Service (PUMA)

Introductory paper

L. Georghiou (Director, PREST, United Kingdom): "Issues in evaluation practices in OECD countries"

Session 1: Evaluation of financial support to industrial R&D (Chair: *Ms. J. Seyvet*, Deputy-Director ANVAR, France)

1. *R. Lattimore* (Productivity Commission, Australia): "R&D fiscal incentives in Australia: impact and policy lessons"
2. *H. Capron/B. van Pottelsberghe* (University of Brussels, Belgium): "Public support to business R&D: an integrated assessment scheme and some additional quantitative evidence"
3. *A. Hervik* (Norwegian School of Management, Norway): "Evaluation of user-oriented research in Norway: estimation of long run economic impacts"
4. *M. Dorsman* (Ministry of Economic Affairs, Netherlands): "Evaluation of industrial R&D support in the Netherlands: recent developments"

Discussant: *B. Hall* (UC Berkeley, United States)

Session 2. Evaluation of large technology programs (Chair: *Ms. C Stevens* (Head, Science and Technology Policy Division, DSTI, OECD)

1. *P. Cohendet* (BETA, France), "Evaluating industrial indirect effects of technology programs: the case of the European Space Agency"
2. *M. Sakakibara* (UCLA, United States), "Evaluation of government-sponsored research joint ventures in Japan"
3. *A. Link* (University of North Carolina, United States) and *J. Scott* (Dartmouth College, United States) "Evaluating technology-based public institutions: lessons from the National Institute of Standards and Technology"
4. *T. Luukkonen* (VTT Group of Technology Studies, Finland) "The increasing professionalisation of the evaluation of mission-oriented research: implications for the evaluation process"

Discussant: *R. Lyon* (Senior Economist, Office of Management and Budget, United States)

Discussion panel #1: Evaluation techniques (Chair: *D. Malkin*, Head, Economic Analysis and Statistics Division, DSTI, OECD)

M. Pianta (National Research Council, Italy): The use of innovation surveys for policy evaluation purposes

A. Piric (Ministry of Research, Science and Technology, New Zealand): Methodological issues in evaluating public investment in R&D

R. Jarmin (Bureau of the Census, United States) The use of micro-level data

R. Lattimore (Productivity Commission, Australia) Cost-benefit analysis

J-P. Meurice (Ministry of Industry, France)

Friday 27 June 1997

Session 3. Evaluation of diffusion-oriented policies (Chair : D. Deniozos (Head, European Framework Programs, Ministry of Education, Greece))

1. *R. Jarmin* (Bureau of the Census, United States) and *J. Jensen* (Carnegie Mellon University, United States): “Evaluating government technology programs: the case of manufacturing extension”
2. *P. Shapira* (Georgia Institute of Technology, US) and *J. Youtie* (Georgia Tech, US): “Evaluating technology deployment at State level: Methods, results and insights from the Georgia Manufacturing Extension Alliance”
3. *S. Arvanitis and H. Hollenstein* (Federal Institute of Technology, Switzerland) “Evaluating the promotion of advanced manufacturing technologies using micro-level survey data”
4. *E. Arnold and K. Guy* (Technopolis, UK) “Technology diffusion programs and the challenge of evaluation”

Discussant: *J. van Reenen* (Institute for Fiscal Studies, United Kingdom)

Discussion panel #2: The role of institutions in policy evaluation (Chair: *E. Ormala*, Chief Planning Officer, S&T Policy Council, Finland; Chairman, OECD TIP Group)

L. Durieux (Evaluation Unit, DG12, European Commission) on the scheme used for evaluation of European RTD Programs

J-E. Aubert (OECD) on OECD experience with evaluating innovation policies

M. Stampfer (Federal Ministry for Science and Transport, Austria) on S&T policy evaluation in Austria

A. Birch (Agency of Trade and Industry, Denmark) on the evaluation of the GTS-institutes in Denmark

Session 4. Strategic evaluation and new trends and approaches (Chair: *M. Bradbury*, Chairman of Evaluation Methodology Committee, Department of Trade and Industry, United Kingdom)

1. *S. Kuhlmann* (Fraunhofer-Institute Systems and Innovation Research, Germany), “Technology Policy Evaluation: Present State in Germany and European Outlook”
2. *P. Laredo* (Ecoles des Mines, France): “Evaluation of technology policies in France”
3. *R. McDonald* (Industry Canada) and *G. Teather* (National Research Council of Canada): “Science and Technology evaluation practices in the government of Canada”

Discussant: *L. Georghiou* (Director, PREST, United Kingdom)

Policy roundtable : The feedback of evaluation results into policy design (Chair: *J. Ryan*, First Assistant Secretary, Department of Industry, Science and Tourism, Australia; Chairman, OECD Industry Committee)

D. Bureau (Sous-directeur de la prévision, Ministère de l'économie, France)

J. Gabolde (Director, Strategy and Coordination, DGXII, European Commission)

S. Wakimoto (Director for Research Coordination and Planning, MITI)

M. Bradbury (Chairman of Evaluation Methodology Committee, Department of Trade and Industry, United Kingdom)

R. Lyon (Senior Economist, Office of Management and Budget, United States)

Closing remarks

R. Nezu, Director, Directorate for Science, Technology and Industry, OECD