Chapter 3

ISSUES IN THE EVALUATION OF INNOVATION AND TECHNOLOGY POLICY

by

Luke Georghiou
Policy Research in Engineering, Science and Technology (PREST), University of Manchester

Introduction

As policy makers struggle to improve the performance of their innovation systems, and in particular to help firms in their countries become more innovative and more able to draw upon science and technology in the enhancement of their competitiveness, it is not surprising that there is a strong desire to know what works and how to make it work better. Even more so when support measures consume scarce public funds which have to be justified to society and its political representatives as good investments which would not occur without intervention. In this environment the expectations placed upon the evaluation of innovation and technology policies are high. In this chapter a selection of the experiences in OECD countries of those who have attempted to satisfy this demand is discussed, together with the issues they raise for the practice of evaluation.

The development of evaluation in the past two decades has mirrored the evolution of technology and innovation policy, beginning in the 1980s with a pre-occupation with large-scale pre-competitive and usually collaborative RTD programmes and gradually broadening the range of instruments employed through an increasing focus on enhancing the environment for innovation and technology transfer. Probably the main trend in innovation policies during the 1990s has been away from financial support measures which assist firms to make better use of given opportunities for innovation and towards policies which create or enhance opportunities for innovation (Metcalfe *et al.*, 1997). The latter include support for networking, between firms and between firms and the science base. A strong growth area has been the provision of information to companies, to guide firms towards suitable partners or funding, to assist with decisions about the acquisition of technology or to improve management capabilities for innovation. These initiatives often make use of consultancy support. Other policies focus upon the supporting infrastructure, including incubators and science parks. Foresight programmes represent an attempt to build networks through the formation of a common vision of the future. All of these instruments and others demand evaluation but pose strong methodological challenges. To some degree this is inevitable – the level of input received by a particular firm may be quite small, while the effects of information rather than resources can be ephemeral and are often unacknowledged even when important.

In the following sections, evaluations of financial support for R&D, direct and indirect, are discussed first, following which a brief review is made of evaluations of policies which enhance innovation opportunities, including those of the activities of institutions, technology transfer mechanisms and foresight. The role of institutional settings for evaluation is considered and finally some unresolved issues and challenges are discussed in the concluding section.

Evaluation of financial support for R&D

Evaluation of collaborative R&D programmes

The development of evaluation practice during the 1980s has been well documented (Gibbons and Georghiou, 1987; Meyer-Krahmer and Montigny, 1989; Ormala, 1989; Roessner, 1989; and Callon, Larédo and Mustar, 1995). A dominant feature of technology policy during that decade was the rise of large-scale collaborative R&D programmes at national and international level. At that time a new policy instrument, these programmes in turn played a significant role in bringing evaluation to prominence, as policy makers and programme managers sought to learn how best to use this instrument.

During the 1990s evaluations of these programmes have continued, with a particularly high profile for the many types of evaluations of the EU's Framework Programmes (reviewed in Georghiou, 1995*a*), the EUREKA Initiative (Ormala *et al.*, 1993; Airaghi *et al.*, 1995; Dale and Barker, 1994) and of the US Advanced Technology Programme (Link, 1993). While these evaluations have been broadly positive about the effects of the programmes, they have been unable in most cases to deliver the type of information about return-on-investment which some policy makers desire. This is not caused by the inadequacy of the evaluators, but rather because an ROI approach usually presumes a linear/sequential model of innovation whereby the benefits of a single research project lead to or are captured by specific innovations. This is rarely the case, particularly for collaborative research.

Evaluations have demonstrated that socio-economic effects may be manifested not only through competitiveness and exploitation/market-related effects (sales of products, reduction of process costs, etc.) but also through individual and organisational learning effects (including partnerships and networking), influencing norms and standards, generation of externalities, and contributions to skills and research manpower (Georghiou and Meyer-Krahmer, 1992). These do not reduce to a single monetary variable, a view confirmed by industry (Anderson and Fears, 1996) and by specific studies (Bureau of Industry Economics, 1994*a*).

Furthermore, evaluating the relation between support for collaborative R&D and its effects requires a clear understanding of the response of the firm to that support. Two related aspects are of critical importance: additionality and the relationship of the funded project to the firm's R&D project portfolio (Buisseret *et al.*, 1995). Additionality is traditionally concerned with the difference arising from an intervention, most commonly in input terms (*i.e.* does a financial input to a firm result in that firm spending that "additional" sum on R&D?). Most directly, would the project have happened without the intervention? For example the US General Accounting Office examined whether research projects in the Advanced Technology Program (ATP) would have been funded by the private sector if they had not received ATP funding (GAO, 1996). In practice the situation is frequently more ambiguous, with the project being carried out "differently" as a result of funding (for example collaboratively rather than by a single organisation). The concept of behavioural additionality has been developed to describe this category of effect, which, it turns out, is often the most durable. On the second aspect, the methodological problem arises because the publicly funded project can be a contractual entity to satisfy the requirements of sponsoring agencies rather than the real unit of activity within the firm. The research may well have started before the funding, finish after it and encompass a substantially wider range of activity than that described in the contract. Most support schemes also stop well short of the type of development activity which brings products to the market. For these and other reasons, evaluations are increasingly having to turn away from the simple project-by-project perspective and to examine, at least for large firms, the overall strategy and portfolio.

Diminishing returns and the rise of monitoring approaches

There is some evidence of diminishing returns in the evaluation of collaborative research programmes. As the policy instrument is repeated and replicated, the chance of any surprising lessons being encountered during an evaluation has diminished, though new circumstances always bring new challenges. Notwithstanding, there is some evidence of this type of evaluation becoming routinised, as policy makers seek economies. This has been manifested in a growing emphasis upon the use of performance indicators and a convergence between the activities of evaluation and monitoring.

The European Commission has increased its monitoring efforts for the Framework Programme while cutting back the time and resources devoted to programme evaluations. At the programme level this is manifested in the current approach which combines continuous "external monitoring" with periodic "external five-year assessments" (CEC, 1996). These take place both at the level of specific programmes and for the Framework Programme as a whole. External monitoring consists primarily of the collection of data for performance indicators, which are subsequently analysed and reported upon annually by a small group of independent experts. The five-year assessments take a more strategic view, with the aim of assessing relevance, efficiency and effectiveness (Davignon *et al.*, 1997). One interesting aspect is that they are intended to assess all of the work within their remit which has taken place in the preceding five years, that is the outputs of the previous programme and the progress to date of the current programme. While a good solution to the constant tension in evaluation between giving time for results to emerge and delivering the evaluation in time for the next cycle of decision making, this approach is dependent upon continuity between successive generations of programmes. If the Fifth Framework Programme turns out to represent a new point of departure, mapping past and present activities to each other could be challenging for those concerned.

Further evidence of the trend towards continuous evaluation within the Commission comes from the development of a project-level toolkit approach known as COMEVAL (Common Methodology for the Evaluation of RTD Results). This will be applied to each project with the aim of collecting systematic and easily aggregable information about project outputs and effects. Currently under test prior to planned implementation in 1998, COMEVAL provides a standard terminology for the collection of data on project aims and outputs, together with a standard set of criteria against which to judge project achievements.

The EUREKA initiative has gone further down this path, with the establishment in 1996 of a scheme for "continuous and systematic evaluation" which aims to routinely collect data about the effects of its projects, following a recommendation of an earlier evaluation. All participants receive a postal questionnaire on completion of their project, and then for those indicating the achievement or expectation of market effects, further short "market impact" questionnaires are sent one, three and five years afterwards. Even with this level of automation, it was found that interpretation of such results was far from straightforward, and in the second year of operation an expert panel was commissioned to oversee the process and to advise on the production of the annual report. It was also clear that a questionnaire approach alone does not capture the full complexities and richness of socio-economic effects as described above. In consequence, the questionnaires are being supplemented with face-to-face interviews.

The EUREKA evaluation activity recognises that project effects may take a considerable time to become evident. An even longer-term perspective has been achieved by a study for the Norwegian Ministry of Industry which has examined the long-term outcome of a cohort started in 1984-89, previously evaluated in 1992 and reassessed in 1996. The aim was to compare expectations with outcomes in terms of sales generated, productivity gains, networking benefits and competence contributions. One conclusion was that some 12-15 years were necessary before outcomes became clear, with consequent implications for monitoring practice.

The desire to apply performance indicators goes beyond collaborative research, being symptomatic of the move towards "new public management" with its general emphasis upon accountability. This is evident in the United States, where the Government Performance and Results Act, passed by Congress in 1993, is requiring all agencies, including those which support research, to set quantitative performance targets and report annually upon their progress towards them (Cozzens, 1995).

Return on investment

The issue of return on investment was discussed in the context of collaborative R&D above. Undoubtedly this issue will remain on the agenda for evaluators, if only because there will always be a demand for it from those responsible for the allocation of resources across the whole of public expenditure. This demand exists at two levels, one being an assurance that there is a positive return and hence that expenditure on R&D should be maintained. By and large this demand can be satisfied by academic studies taking a generally broad view (Mansfield, 1991; Martin *et al.*, 1996), although challenges to this viewpoint have begun to emerge (Kealey, 1996), whose arguments for a withdrawal of public funding for science find a receptive ear in some political quarters even if they are judged to be flawed by the science policy community. To the extent that *ex-post* evaluation is carried out in industry it mainly serves this function, that is to provide evidence to the main board of the company that R&D is yielding benefits (EIRMA, 1995).

The second level of demand is for cost-benefit approaches which can support the distribution of resources by feeding back information on the relative yield of different areas of R&D. Leaving aside the issue of whether past returns are a good guide to the future, the level of accuracy required has prevented widespread adoption of this approach. An interesting attempt to overcome some of the difficulties has been carried out by the New Zealand Foundation for Research, Science and Technology (FRST, 1996). This pilot study aimed to integrate new growth theory, evolutionary economics and case-study-based grounded research methods to evaluate the wider benefits of meat research. While labour intensive in its approach, the methodology yielded information at the level of order of magnitude for direct economic benefits and significant information on the indirect economic benefits, including generic knowledge and technology with spillovers to other areas of R&D. From a methodological point of view, the study confirmed the assertion made earlier in this chapter that the benefits of research cannot be attributed to individual projects, but rather need to be appraised at the portfolio level, so as to allow the indivisibilities and interrelationships between research projects to be acknowledged.

Across innovation policy more generally, the UK National Audit Office and Department of Trade and Industry developed a methodology for comparing the comparative cost-effectiveness of the Department's innovation schemes in delivering their objective of stimulating industrial innovation and increasing the competitiveness of UK industry (NAO, 1995). Grounded in decision theory, the basis of this approach was to identify 14 indicators representing effects on innovation and competitiveness and five representing costs. These were a mixture of process indicators (for example, average application processing time) and output indicators (for example, number of licences acquired). Values on each indicator were transformed onto "preference scales" which rated each scheme on a scale of 100 to 0 on a relative basis, that is the best performing scheme on that indicator was assigned a score of 100 and the least performing, zero, with the remainder interpolated. The process was repeated for the indicators themselves, to yield finally weighted preference scores based on equal weights for costs and objectives. By this methodology, the Teaching Company Scheme (a technology transfer scheme based upon joint supervision of young graduates by academia and industry) emerged as the most preferred scheme, and the LINK collaborative R&D programme the least preferred. Clearly, the outcomes are sensitive to the indicators selected and the weightings assigned to them. The low score for LINK was heavily influenced by poor performance on administrative indicators.

Evaluation of fiscal incentives

Some of the most rigorous economic evaluation work has been applied to the evaluation of fiscal incentives for R&D (Griffith *et al.*,1995). Australia has been prominent in this field, with a series of evaluations from the Bureau of Industry Economics providing not only a detailed review but collectively a picture of the evolution of the effectiveness of the schemes. Hence, in 1993 the Bureau examined the government's 150 per cent R&D tax concession, introduced in 1985, using three main approaches, a before/after comparison of beneficiaries, a survey of registrants seeking subjective views as to whether additional R&D had been encouraged, and a comparison with a "control group" of non-recipients of the concession (BIE, 1993). The report recommended retention of the tax concession but suggested some modifications. The BIE went on to examine a variant of the scheme which introduced "syndication", that is allowing two or more eligible companies to form a syndicate which contracts out or undertakes R&D and allows each company to claim a proportion of such expenditure against tax (BIE, 1994*a*). In practice the scheme allowed research companies with accumulated tax losses to exchange them for R&D funds. The effects of schemes of this type are so complex that a major part of the evaluation is concerned with modelling them. While more qualified in its enthusiasm than in the previously discussed report, the general conclusion was that "the programme generates significant net benefits, so that it should be retained". It was not long before the scheme was revisited by the same evaluator (Lattimore, 1996). By this time it was concluded that the scheme had become dynamically unstable in the intervening two years, with cost-to-revenue increasing exponentially. The new conclusion was that the programme had ceased to generate net benefits for Australians, that superior alternatives existed and hence that the main recommendation should be that "the Syndicated R&D programme should be terminated", which it was. Apart from providing an example of evaluation having a major impact on policy, this sequence of evaluations demonstrates the importance of revisiting long-running schemes periodically, and not assuming that evaluation findings established in one period hold good for the duration of the programme.

Evaluation of opportunity-enhancing innovation policies

Evaluation of institutions

The move towards strategic evaluation has become increasingly apparent in the evaluation of institutions concerned with implementing technology and innovation policy. This unit of analysis is particularly important in France where regular evaluations of "research operators" are performed by the Comité National d'Évaluation de la Recherche (Larédo and Mustar, 1995). High-level reviews of the institutions responsible for delivery of technology and innovation policy are a regular feature of the evaluation landscape. Hence, in France again, the innovation agency ANVAR has been reviewed by a high-level national expert in the light of the changing environment for innovation (Chabbal, 1994), and, in Finland the Technology Development Centre (TEKES) was reviewed by two international experts supported by a national secretariat (Guillame and Zegveld, 1995). It is interesting to note that the original mission of most of the institutions discussed here was the performance of R&D, either without recourse to external customers or to satisfy a broad mission for the public purpose. In these circumstances the main historic function of evaluation was to ensure scientific quality and the time-honoured approach of peer review, normally through the use of visiting outside experts, was applied. Even though the missions of these institutions have evolved to a more interactive and supportive role, in many countries the basic format of peer review has survived even though the questions are much more complex and data-intensive to answer. It can only be assumed that the requirement for legitimacy derived from the reputation of the panellists forms the paramount consideration.

In this section, evaluations concerned with the research quality of institutions are not discussed. Instead the focus is on 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 (an object of many innovation policy measures). Of the many evaluations of this type, one focus of activity has surrounded the large multiprogramme laboratories of the US Department of Energy. Mostly originating with a defence mission, since 1980 these labs have been subject to increasing legislative mandates to undertake commercially oriented activities. A notable step was the National Competitiveness Technology Transfer Act of late 1989. In an attempt to establish a baseline from which a debate on these issues can take place, the US General Accounting Office reported on the extent to which these labs are engaged in research related to commercial product development (GAO, 1994). Based upon a survey of those responsible for existing activities, the principal conclusion was that while the potential for commercial application existed, the timescales involved in commercialisation made it too soon to say whether the potential would be realised. An example of more detailed work of this kind is a study of the impacts of technologies transferred from Sandia National Laboratories to US industry (Falcone, 1995). The method here was to select a series of case studies of transferred technologies and to apply scaled outcome indicators in the calculation of the economic and other impacts of the transfer. Cost-benefit ratios are calculated with and without the sunk costs of producing the knowledge.

The technology transfer process between laboratories and industry is relatively complex, involving, among other mechanisms, the building of networks and linkages with industry. The challenge for evaluation has been to apply some sort of metric to these interactions. An example of a systematic examination of the value of informal contacts comes from the Australian CSIRO Institute of Industrial Technologies (CSIRO, 1995). This study involved a survey of senior research scientists to establish the nature and extent of their contacts and to identify business or other users who could be contacted in the second phase of the study. These users were interviewed about their contacts with

CSIRO and the value to them of the information and advice provided. The study found strong informal contact networks in existence characterised by important information exchanges of value to both sides.

An alternative approach to examination of these linkages is to examine the relevance of laboratories through scientometric indicators. An early example of this approach looked at the performance of two Canadian public laboratories in the field of metals and alloys technology (Dalpe and Anderson, 1993). Among the approaches used was an analysis of interactions based upon papers listed in the Chemical Abstract in 1990. Three indicators of interaction were developed: co-authorship, funding acknowledgements (indicating that another institution will use the results) and data and equipment acknowledgements (describing data exchange and equipment sharing). These interactions were used to profile networks of institutions, including interactions with industry.

Paradoxically, the more commercialised these institutions become, the less is the legitimacy of government departments to evaluate them. Ultimately, if privatised, their role in evaluation becomes the same as that of firms in receipt of support from government, the focus of the external evaluation being upon the policy which they are commissioned to execute, not upon the institution.

This does not preclude the techniques of evaluation being used internally by institutions as part of the strategic management process. A good example of such a "technology transfer" is that of ADEME, the French environment and energy agency. This organisation has taken the techno-economic network approach, initially applied in an evaluative context and over a decade refined and adapted it for application as a strategic tool.

One area of some contention in the examination of large institutions concerns their indirect effects on the economy. Two French studies have reached opposite conclusions when examining space programmes. Work for the European Space Agency (Bach and Lambert, 1992) showed that ESA's contractors gained benefits including new industrial partners, organisational improvements and gains in the competences of their personnel. On the other hand, the French National Committee for Evaluation of Research (CNER) was somewhat more sceptical about the indirect benefits of the French Space Programme, arguing that its training benefits were limited and that there was a growing tendency for it to select existing rather than new technologies (Bureau, 1997). From an evaluation perspective, the most interesting question is whether these differences arose from a genuine disparity in performance of the two programmes, or whether they were a result of the different methodologies applied.

Evaluation of technology transfer measures

Diffusion-oriented policies have been in place in some countries for several years (also exemplified by some of the activities of the large institutions described above). Most of these seek to transfer technologies to industry and in particular to small and medium-sized enterprises (SMEs), generally working on the assumption that there is a market failure in the supply of relevant information. Naturally, the effectiveness of these policies has also been subject to evaluation and, with the increasing trend towards this type of measure, the experience gained is of more than usual interest. Germany was one of the first countries to apply these policies on a large scale and it has also seen some extensive evaluation work. A prominent example in more recent times has been the real-time evaluation of the innovation support programme Microsystem Technology (MST) described by Hafkesbrink and Horst (1995). The methodology involved use of a comprehensive survey to identify and classify the existing transfer activities for MST, followed by selection of four specific transfer projects for more detailed examination. Some interesting conclusions emerging from this work concern the importance of the timing of the state intervention with regard to the maturity of the technologies and markets concerned. A related issue which the evaluation has not to date been able to cast light upon is what can be termed "policy persistence"; that is, the degree to which, once corrected by intervention, the information failure is superseded by market mechanisms.

A Swiss technology transfer programme to support rapid diffusion of basic technologies in the field of Advanced Manufacturing Technologies (AMT) is built on the assumption that a two-year effort is sufficient to overcome such bottlenecks (Hollenstein, 1997). The evaluation of this programme is interesting in that it is founded upon empirical modelling of adoption decisions grounded in the literature on diffusion of technology. However, reported progress to date covers only a more descriptive analysis founded upon the Swiss Innovation Survey. The use of the innovation survey raises a point of more general methodological significance. By definition, diffusion-oriented measures seek to reach a wide audience of firms making survey work expensive and running the risk of non-response because of the much lower level of involvement compared to receipt of a project grant.

The United States has also seen public policies to promote manufacturing modernisation, operated through manufacturing technology centres, industrial extension programmes and other initiatives at both federal and state levels. Growing experience of the evaluation of these initiatives has accumulated (Shapira and Youtie, 1995). Evaluation issues which have arisen in the context of the US experience include the view that it is necessary to go beyond counts of the number of firms served in various ways and to measure the actual benefits to those firms and welfare benefits more generally. Other concerns reported include the role of comparison groups of non-customers and the danger of imposing too many information burdens upon firms and programme operators.

Two further evaluations in the technology transfer domain have focused upon the role of technology centres in supporting such activities. From Denmark, the evaluation addresses the GTS network of approved technology service institutes, bodies which carry out applied research, certification, testing, measurement and standards, and provide information, consultation and training. Evaluation issues in this context include the question of whether such bodies need to perform R&D in order to remain at the cutting edge of the technologies which they seek to transfer. These institutes are evaluated every three years by fairly conventional means; that is, preparation of background material by the institute concerned which is then reviewed by an expert panel with international membership. The panel carries out visits and interviews within the institute and with clients. Evaluation reports are strictly confidential.

The second example of evaluations of technology centres is of the Canadian Technology Outreach Program. This programme centralised funding for technology centres across Canada, providing financial assistance for activities and services which accelerate the acquisition, development and diffusion of technology and management skills in industry. Methods used included case studies of eight of the 21 centres and a survey of 200 clients of the centres. While generally positive, the evaluation recommended a redesign of the programme in which the centres should become agents or partners in the delivery of services rather than perceived as recipients of contributions (Young and Wiltshire, 1992).

Evaluation is also applied to initiatives to promote academic-industrial co-operation (which, of course, also includes the collaborative R&D programmes already discussed). A new policy approach in this sphere has been the Competence Centres Programme in Sweden (NUTEK, 1997). Co-funded by the Swedish innovation agency NUTEK, the host academic institution and several industrial partners, each centre performs R&D in a focused cross-disciplinary area with the aim of building a solid and long-lasting bridge between the academic and industrial worlds. The evaluation approach was conventional, consisting of a report by a panel of visiting international experts. In contrast to the Danish reports described above, the findings of the panel were publicly available, including specific comments on the individual centres.

Technology transfer need not be confined to scientific and engineering knowledge. Increasingly it has been recognised that innovation policy should encompass the provision of firms with wider competences including management skills. An example of this type of programme is the United Kingdom's "Managing in the 90s" programme. This aims to improve the performance of firms through better dissemination of information about best practice, using presentations, seminars, videos and written material. The second phase of the programme was evaluated in 1995 by means of discussions with academic experts on the validity of the rationale, examination of the extensive monitoring returns (exit questionnaires at all "M90s" events followed up by telephone questionnaires of a small sample of attendees) and a limited amount of original fieldwork to confirm and calibrate this data (Gohil and Roe, 1995). It was the first of these evaluation activities which was unusual, whereby the rationale of the scheme was tested against the views of six experts. Among the challenges emerging was the view that most firms do not wish to grow but that the importance of growth meant that the programmes could and should be targeted at firms likely to grow rather than at those most likely to gain in efficiency.

Evaluation of foresight

The 1990s have also seen national foresight programmes becoming an important element of technology policy. As well as constituting in some senses a form of *ex-ante* evaluation, foresight is itself a policy instrument which will require evaluation. As with the infrastructural measures described above, it will prove difficult to separate decisions driven by foresight from those which would have occurred anyway or which have been re-labelled. In Japan, some technical evaluation of foresight has been carried out, in that predictions made in the first STA Delphi survey in 1970 were assessed after a 25-year period to identify which had occurred fully or partially. In several countries, foresight activities have been followed up by surveys of the utility of the results to firms. Thus far, there have been no comprehensive evaluations of foresight activity, extending to the networking benefits and the results of promoting a foresight culture, but those involved recognise the need for learning in these spheres and some commentators have begun at least to raise the key issues (POST, 1997). In this sense the situation is analogous to that in the early days of collaborative research programmes.

Institutional settings and the ability to learn

It is not the intention here to embark upon a review of the variety of institutional settings in which evaluation may be performed. This topic has been recently visited for most of Europe, revealing a situation in which three main models emerged: that of centralised systems (the United Kingdom and France) in which the requirement to evaluate and guidelines for practice emanate from the centre of government; decentralised systems (the Netherlands and Germany) where evaluation was common practice but was not promoted or co-ordinated in a systematic way across departments and agencies; and a pre-evaluation stage (most of Southern Europe) where a legislative framework was normally in place but where practice had yet to develop in the majority of circumstances (Georghiou, 1995*b*). This is not to disguise important national differences – the guarantor model of France and the ROAME system of the United Kingdom bear little resemblance. Nevertheless, always remembering that evaluation practice reflects the administrative culture in which it is located, it is possible to extend these broad categories to other OECD regions. Hence, Canada falls into the centralised group and the United States into the decentralised group (notwithstanding the GPRA), reflecting the different branches of government there which perform evaluation for their own purposes. Australia gives the appearance of a centralised approach with a high level of activity but much of this is *ad hoc* rather than institutionalised. There is a distinct Nordic model in which the most common features are panels of foreign experts. In Japan, up to the present, evaluation has normally been conducted in-house with a focus upon the achievement of technological objectives. It will be interesting to see whether the opening up of some scientific institutions (for example the RIKEN laboratory) to review by international panels presages a wider recourse to this approach.

The precise institutional framework for evaluation is of less importance than its functionality. The clear requirements are the provision of a framework which:

- ◊ ensures that evaluation takes place on a programmed and properly resourced basis;
- ◊ guarantees the independence of the evaluators; and
- ◊ provides a mechanism for feedback of the results into policy making and hence for learning.

To this list many would add the requirement for activities to be programmed in the first place (thus providing a unit for evaluation) and for objectives to be clearly formulated in advance as a basis for evaluation.

Conclusions

In reaching a conclusion, it is clear that there are tensions in the practice of evaluation. One such tension is that between the trend towards 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. The advantages of systematic information collection are clear; the task now is to understand the limitations of what these approaches produce and to ensure that they are complemented by evaluation at a higher level. In one respect systematic approaches avoid a growing problem in evaluation, the issue of the response burden laid upon recipients of assistance. When the primary support measure was large grants it was reasonable to expect co-operation from the beneficiaries, but in an information-oriented system, the benefit may be so diffuse that those sampled may resent the cost of providing information (as well as finding it difficult to attribute effects to particular policies).

A second tension lies 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. The argument against their use centres upon the difficulties involved in valuing the intangible benefits of R&D in particular. Cost-benefit approaches seem more securely founded in the evaluation of some other measures (notably fiscal incentives), but even here they depend upon behavioural assumptions.

More generally, the problem of additionality remains a core issue for evaluation and one which reaches the most sensitive areas of policy, going to the core of whether there is a rationale for intervention. A focus on behavioural additionality creates scope for a wider appreciation of the effects of funding than the conventional input and output approaches.

Some aspects remain as a challenge not yet solved for evaluators. Employment effects of innovation policy provide one such instance. To all of the usual problems of attribution are added the difficulties of calculating primary and secondary effects of innovations which may themselves aim to reduce labour. The results to date on this issue are not satisfactory, although they appear to indicate relatively low effects, probably slightly positive on balance. What is clear is that the political pressure for information of this type will remain while employment is a key issue in economic and social policy.

Another challenge concerns understanding the longer-term effects of policy measures. Little is known about the persistence of measures intended to induce behavioural changes after the initial incentive is removed. Some of the findings reported in this chapter have stressed the importance of the timing of measures and others have explored the long-term outcomes of policies but more work needs to be done in this field.

Having reviewed the practice of evaluation across the domain of innovation policy, one feature which emerges quite strongly is the general lack of cross-cutting evaluations, comparing the relative effectiveness of policies. The example of the National Audit Office discussed above is one exception and there are others which look at the net effect of particular types of schemes such as Federal Technology Partnerships (Office of Technology Policy, 1996) or policy for innovation in SMEs (Christiansen *et al.*, 1996) as well as the OECD country reviews. However, 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. More broadly, innovation is affected by a wider range of policies and regulatory frameworks which also need to be brought into the scope of such evaluation.

The implication of the new, more varied trend in innovation policies is that an adaptive learning approach is the optimal one. Within this, evaluation has a clear role to play in providing not only the necessary direct feedbacks on particular measures but also information which enables successful policies to be replicated and unsuccessful ones terminated. While evaluation thus appears as an essential component in an evolutionary approach to policy, this can only be effected if an institutional framework is in place which ensures that the policy lessons are appropriately disseminated and that all elements of the system are evaluated in a balanced way. This does not demand uniformity in evaluation practice, variety here is as productive as it is in the innovation system itself.

REFERENCES

AIRAGHI *et al.* (1995), *EUREKA Evaluation Report*, EUREKA Secretariat, Brussels.

ANDERSON, J. and R. FEARS (eds.) (1996), *Valuating and Evaluating: Assessment of the Value of R&D in Creating National and Corporate Prosperity*, SmithKline Beecham, Cambridge.

BACH, L. and G. LAMBERT (1992), "Evaluation of the Economic Effects of Large R&D Programmes: The Case of the European Space Programme', *Research Evaluation,* Vol. 2, No. 1, April 1992.

BUISSERET, T.J., H.M. CAMERON and L. GEORGHIOU (1995),"What Difference Does it Make? Additionality in the Public Support of R&D in Large Firms", *International Journal of Technology Management*, Vol. 10, Nos. 4/5/6, pp. 587-600.

BUREAU, D. (1977), "L'Évaluation des politiques et programmes portant sur l'innovation et la technologie en France", French Ministry of Economy and Finance, document submitted to the OECD.

BUREAU OF INDUSTRY ECONOMICS (1993), *Research Report 50: R&D, Innovation and Competitiveness – An Evaluation of the Research and Development Tax Concession*, Australian Government Publishing Service, Canberra.

BUREAU OF INDUSTRY ECONOMICS (1994*a*), *Research Report 60: Syndicated R&D – An Evaluation of the Syndication Program*, Australian Government Publishing Service, Canberra.

BUREAU OF INDUSTRY ECONOMICS (1994*b*), *Occasional Paper 16: Beyond the Innovator: Spillovers from Australian Industrial R&D,* Bureau of Industry Economics, Canberra.

CHABBAL, R. (1994), *The French Financial System in the Face of Innovation Investment*, Report to Minister of Enterprises and Economic Development, 7 July.

CALLON, M., P. LARÉDO and MUSTAR (1995), *La gestion stratégique de la recherche et de la technologie*, Economica, Paris.

CHRISTIANSEN, M., K. MOLLER JORGENSEN and K. SMITH (1996), *Innovation Policies for SMEs in Norway*.

COMMISSION OF THE EUROPEAN COMMUNITIES (1996), *Communication from the Commission to the Council and the European Parliament – Independent External Monitoring and Evaluation of Community Activities in the Area of Research and Technological Development*, Brussels, 22.05.1996, COM(96)220 final.

COZZENS, S. (1995), "U.S. Research Assessment: Recent Developments", *Scientometrics,* Vol. 34, No. 3, pp. 351-362.

CSIRO (1995), *Putting a Value on Informal Industry Contacts with CSIRO*, CSIRO Australia Industrial Technologies, June.

DALE, A. and K. BARKER (1994), "The Evaluation of EUREKA: A Pan-European Collaborative Evaluation of a Pan-European Collaborative Technology Programme", *Research Evaluation,* Vol. 4, No. 2, pp. 66-74.

DALPE, R. and F. ANDERSON (1993), "Evaluating the Industrial Relevance of Public R&D Laboratories" in B. Bozeman and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, Boston/Dordrect/London.

DAVIGNON, E., A. AIRAGHI, N. CHRISTODOULAKIS, J. DOOGE, R. NICHOLSON, J. KUUSI, F. MEYER-KRAHMER, A. SYROTA, J. VAN RUITEN, J. VIANA BAPTISTA and A. CALDER (1997), *5 Year Assessment of the European Community RTD Framework Programmes*, Commission of the European Communities, February.

EIRMA (1995), *Evaluation of R&D Projects*, European Industrial Research Management Association, Paris.

FALCONE, S. (1995), "Technology Transfer Impact Profiles – Interim Report No.1" prepared for Sandia National Laboratories Contract No.AM 3455, November.

FOUNDATION FOR RESEARCH, SCIENCE AND TECHNOLOGY (1996), "The Benefits of Meat Research in New Zealand: A Pilot R&D Outcome Review", Draft Report, 19 November.

GEORGHIOU, L. (1995*a*), "Assessing the Framework Programmes – A Meta-evaluation", *Evaluation,* Vol. 1(2), pp. 171-188.

GEORGHIOU, L. (1995*b*), *Research Evaluation*. For more detail see all of the other papers in this special edition of *Research Evaluation*.

GEORGHIOU, L. and F. MEYER-KRAHMER (1992), "Evaluation of Socio-economic Effects of European Community R&D Programmes in the SPEAR Network", *Research Evaluation*, Vol. 2, No. 1 pp. 5-15.

GENERAL ACCOUNTING OFFICE (GAO) (1994), *National Laboratories: Are their R&D Activities Related to Commercial Product Development?*, United States General Accounting Office, GAO/PEMD-95-2.

GENERAL ACCOUNTING OFFICE (GAO) (1996), *Measuring Performance: The Advanced Technology Program and Private-Sector Funding*, United States General Accounting Office, GAO/RCED-96-47.

GIBBONS, M. and L. GEORGHIOU (1987), *Evaluation of Research – A Selection of Current Practices*, OECD, Paris.

GOHIL, A. and P. ROE (1995), "An Evaluation of the Managing in the 90s – Phase 2 Programme", Assessment Paper No.31, The Assessment Unit, Technology and Innovation Policy Division, Department of Trade and Industry, November.

GRIFFITH, R., D. SANDLER and J. VAN REENEN (1995), "Tax Incentives for R&D", *Fiscal Studies*, Vol. 16, No. 2., pp. 21-44.

GUILLAME, H. and W. ZEGVELD (1995), *The Technology Development Centre of Finland, TEKES – An International Evaluation*, Ministry of Trade and Industry Publications 5/1995, Finland

HAFKESBRINK, J. and H. HORST (1995) "Evaluation of Technology Transfer Initiatives in the Framework of Innovation Support Programmes – Instruments of Technology Transfer Evaluation: The Case of Microsystem Technology" in G. Becher and S. Kuhlmann (eds.), *Evaluation of Technology Policy Programmes in Germany*, Kluwer, Dordrecht, pp. 197-213.

HOLLENSTEIN, H. (1997), "The Effectiveness of the Diffusion-oriented Technology Policy of Switzerland – The Case of Advanced Manufacturing Technologies (AMT)", paper presented to the Swedish Ministry of Trade and Industry/OECD International Conference on 'Industrial Competitiveness in the Knowledge-based Economy – the New Role of Government', Stockholm, 20-21 February.

KEALEY, T. (1996), *The Economic Laws of Scientific Research*, MacMillan, London.

LARÉDO, P. and P. MUSTAR (1995), "France, The Guarantor Model and the Institutionalisation of Evaluation", *Research Evaluation,* Vol. 5, No. 1, April.

LATTIMORE, R. (1996), *Evaluation of the Syndicated R&D Program*, Department of Industry, Science and Tourism, Australian Government Publishing Service, Canberra, July.

LINK, A.N. (1993), "Evaluating the Advanced Technology Programme: A Preliminary Assessment of Economic Impacts", *International Journal of Technology Management,* Vol. 8, Nos. 6/7/8, pp. 726-739.

MANSFIELD, E. (1991), "Academic Research and Industrial Innovation", *Research Policy, 20*, pp. 1-12.

MARTIN, B. *et al.* (1996), *The Relationship Between Publicly Funded Basic Research and Economic Performance*, Report prepared for HM Treasury, July.

METCALFE, J.S., L. GEORGHIOU and A. JAMES (1997), "Adaptation and Learning in European Technology Policy", paper presented to the 1997 Annual Meeting of the American Association for the Advancement of Science, Seattle, 13-18 February.

MEYER-KRAHMER, F. and P. MONTIGNY (1989), "Evaluations of Innovation Programmes in Selected European Countries", *Research Policy,* Vol. 18, No. 6.

NATIONAL AUDIT OFFICE (1995), Report by the Comptroller and Auditor General, *The Department of Trade and Industry's Support for Innovation,* HMSO HC715 Session 1994-95, London, 4 August.

NUTEK (1997), *The NUTEK Competence Centre Programme – First International Evaluation Group (12 Centres),* March.

OFFICE OF TECHNOLOGY POLICY (1996), *Effective Partnering – a Report to Congress on Federal Technology Partnerships,* by Richard Brody, US Department of Commerce, Office of Technology Policy, April.

ORMALA, E. (1989), "Nordic Experiences of the Evaluation of Technical Research and Development", *Research Policy*, Vol. 18, No. 6.

ORMALA, *et al.* (1993), *Evaluation of EUREKA Industrial and Economic Effects,* EUREKA Secretariat, Brussels.

POST (1997), *Science Shaping the Future? – Technology Foresight and its Impacts*, Parliamentary Office of Science and Technology, London, June.

ROESSNER, D. (1989),“Evaluating Gvernment Inovation Pogrammes: Lessons from the U.S. Experience”, *Research Policy*, Vol. 18, No. 6.

SHAPIRA, P. and J. YOUTIE (eds.) (1995), *Evaluating Industrial Modernization – Methods and Results in the Evaluation of Industrial Modernization Programs*, Georgia Institute of Technology, School of Public Policy and Economic Development.

YOUNG AND WILTSHIRE, MANAGEMENT CONSULTANTS (1992), *An Evaluation of the Technology Outreach Program of Industry, Science and Technology Canada*, prepared for Industry, Science and Technology Canada.