Research Evaluation, volume 11, number 2, August 2002, pages 81–84, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Outcomes and impacts

Research evaluation in Europe: state of the art

Terttu Luukkonen

Although there is no single European way of doing research evaluation, it has been going on since the 1970s, importantly supported by the European Community. Trends include: evaluating socio-economic impacts, commenting on the importance and appropriateness of the evaluated activities, modified peer review, semi-professionalism, and tendering to groups of experts in different countries.

Terttu Luukkonen is at the Research Institute of the Finnish Economy, Lönnrotinkatu 4 B, FIN-00120 Helsinki, Finland; email: terttu.luukkonen@etla.fi

This paper was presented at the American Association for the Advancement of Science Meeting, 14–19 February 2002, in Boston, MA, USA.

THIS PAPER USES the term 'research evaluation' rather than 'outcome measurement' because it is a somewhat broader concept. Outcome measurement refers to the use of quantitative measures in evaluation and to *ex post* evaluation. The term 'evaluation' is neutral with regard to methods and can also refer to qualitative evaluation; further, it does not only refer to *ex post* evaluation of the impacts or outcomes, but can be a question of an evaluation of ongoing activities. One of the basic assumptions of this paper is that there is no inherent superiority in methods, be they quantitative or qualitative. The choice should depend on the information needs, resource constraints and other relevant factors. I will, however, concentrate on the evaluation of outcomes and impacts, since a lot of current evaluation activities in Europe concern them.

The topic of this paper is the evaluation of impacts or outcomes of research policy initiatives (organisations, supported activities) in Europe. Even though there may be distinctive European styles in evaluation, it is good to remember that Europe is heterogeneous, and generalisations easily fall into a trap of simplification and become false. *There is no single European way of doing research evaluation*. There have been attempts at classifying the European national systems in terms of their legislative systems and frameworks which influence their research evaluation cultures:

- centralised frameworks (the UK and France);
- countries where evaluation was well-established but uncoordinated across ministries and agencies (Germany and the Netherlands);
- countries with rigid legislative frameworks for science policy which left little room for evaluative culture to emerge (most of southern Europe); and
- the Nordic countries where evaluation has long

 0958-2029/02/020081-4 US$08.00

been well-established and which have had a distinctive evaluation style involving heavy use of overseas/foreign panelists (Georghiou, 2000).

These styles have, however, become more mixed, to some extent because there is increasing exchange of experiences among the European countries. Within a particular country, there may be different types of evaluation approaches. An important influence in evaluation in Europe, the EU Commission has changed its own evaluation practices over time from more in-depth evaluation to more measurement-oriented evaluation.

Research evaluation has by now a fairly long tradition in several European countries, since the 1970s and 1980s. Some of the earliest countries to start evaluation activities of this kind — that is, addressing impacts and outcomes of research initiatives, funding programmes and research organizations — were the Nordic countries. In the beginning, they paid special attention to basic research funding and addressed research quality issues. Sweden was the first country to do so, from the 1970s. Finland, Norway and Denmark followed suite in the early to mid-1980s. The UK was among the first European countries to carry out a large-scale evaluation of a big research programme, when it commissioned the first real-time evaluation of the well-known information technology programme, called the Alvey Programme for Advanced Information Technology (1984–90). Alvey evaluation concerned ongoing activities and gave immediate feedback to the running of the programme. It was important for the development of evaluation competencies in Europe. Another important evaluation which has influenced the evaluation traditions in Europe was the second evaluation of the EUREKA initiative under the French Presidency (1992-93). EUREKA is a bottom-up programme for more near-market research than the EU Framework Programme and is funded from national sources. The EUREKA evaluation developed survey techniques to assess the expected benefits of a programme for the participants (Georghiou, 1999).

An essential part of the European scene is the fact that the EU has played an important role in promoting exchange of information and experiences among national evaluation experts and the potential clients of evaluations. The European Community supported network activities around evaluation capabilities in the late 1980s and early 1990s by inviting experts from both member states and other European states to a series of meetings and by commissioning special studies on various evaluation issues. The EU has been active in evaluation also by commissioning evaluations of its own research programmes and, today, EU research programmes are systematically evaluated at regular intervals. The EU continues to support evaluation expert networks and organises conferences and workshops in evaluation.

There are evaluation meetings and networks among European experts outside the Union, too, and there are purely national networks and associations. There are thus a lot of forums where people interested in research evaluation can meet and exchange views and experiences. This has to some extent contributed to the development of evaluation cultures in these countries.

In the rest of this paper, I will draw attention to some development trends in evaluation in Europe, evident at least in a few countries, and will discuss some of their implications.

Trends in evaluation in Europe

Today we have two to three decades of experiences of research evaluation. This has led to an acknowledgement that no single method is superior and suitable for all situations and that often we have to combine many different methods and approaches. With regard to methods, there are actually opposite trends in evaluation. There is a movement towards quantitative measurement of socio-economic impacts while there is in many places also a growing demand for the evaluation of the strategic importance of activities. The latter uses various mixes of panels and more qualitative evaluation.

Evaluating socio-economic impacts

The emphasis on evaluating the socio-economic impacts of research is related particularly to increasing expectations concerning the economic impacts of research funding, and for the sake of accountability, a wish to prove such impacts quantitatively. It has led to efforts at developing quantitative outcome and impact measures. This demand is met with great difficulties in measurement, not the least because R&D often demands long lead times before the occurrence of economic returns or various societal impacts, but also because of the difficulties in attributing impact to particular funding initiatives. There have been efforts to develop questionnaires to measure socio-economic impacts, particularly the realised versus those expected, but these techniques are not regarded as 'hard' enough. Efforts at measuring quantitatively the economic impact of public R&D programmes at the firm level have produced inconsistent results. Data collection based on self-assessments of economic returns may also have potential biases or distortions in the results. In the hope of guaranteeing support in the future, it is in the self-interest of the money recipients to assess the impact positively (Venetoklis, 2000). Econometric studies can rely on more objective data on firm-level performance, but they have difficulties in controlling for intervening factors that might influence the performance of firms and thus distort the results. Here there is a gap between the expectations by research administration or agencies commissioning evaluations and that which the evaluation community can provide for.

A trend towards performance indicators and measurement runs a risk that indicators provide a narrow and one-sided view of the evaluated activities. Sometimes, the very essence of the impact may escape such measurement efforts. Performance indicators are a good complement to, but not a replacement of more in-depth evaluation exercises.

Importance and appropriateness

As said above, evaluation activities are not only aimed at assessing the efficiency or effectiveness of a set of activities or their socio-economic impacts, but also they increasingly comment on the strategic importance and appropriateness of the evaluated activities. Evaluation is expected to contribute to strategic change in the RTD system, or in the direction of RTD or in the organisation of research performing institutions (Rip, 2000). This kind of evaluation is more difficult than more traditional evaluation, which comments on issues related to quality, efficiency and effectiveness or is of an audit type. A demand for evaluation of the strategic importance of activities reinforces the trend to use a broader expertise base in evaluation teams (discussed below). Overall, an evaluation needs to involve different methods to acquire knowledge and to combine these to draw conclusions. When the evaluation issues become more complex and the evidence to be collected more varied, there will be more demands for time needed for evaluation, too. Evaluations become forums in which important issues concerning future policy directions are debated. Putting into effect the recommendations of such strategic evaluations requires a broad consensus; hence, a broader base of representation in evaluation groups, as will be seen later on.

Modified peer review or mixed panels

A fairly usual method in particularly strategic evaluation is to use so-called modified peer review or mixed expert teams. This means that an evaluation panel has members who are experts in the substantive matters to be evaluated (peers) and experts in other areas, such as evaluation methods, research organisation, research policy issues and often members who represent senior management. These panels may include representatives of what we might call reference or stakeholder groups; that is, people who represent important interests in the research policy area. For example, in the real-time evaluation of a big Additional Funding Appropriation (1996–99) in Finland, the evaluation panel consisted of members representing a public venture foundation, the industry, the trade unions, academic institutions, and international evaluation experts. The real/hidden agenda of the evaluation was to judge whether there ought to be a continued rapid growth in public and hopefully private R&D expenditures, or a level growth. This required the formulation of a broad-based consensus on the matter. Hence, the reference group representation.

Such reference group representation is intended to increase both the credibility of the recommendations and increase the likelihood of their being accepted. The group acts as a kind of forum in which negotiations about future policy directions take place. Such panels usually draw on background studies, indicator data and other relevant information. Panels often meet with key persons for the evaluated activity and carry out interviews with them.

A variation of such panels is a situation in which, in addition to a pure peer review panel, there is a group of professional evaluation experts that conducts or commissions background studies, interviews key people, and draws the overall conclusions. With a few of my colleagues, I was myself involved in such an evaluation last autumn, together with a team of evaluation experts from Technopolis Ltd from the UK and Austria. The evaluation concerned two large technology programmes of the Finnish National Technology Agency (TEKES). In addition to us, the evaluation experts and analysts, there was a panel of peers representing the forefront research in the area. The major difficulty in this evaluation exercise was the very large number of projects funded through the programmes and consequent difficulties in making peer judgements on the strategic relevance, not to speak of the quality, of the funded activities. Usual interview techniques could reach only an extremely small sample of the target groups, which rendered generalisations quite difficult.

An evaluation needs to involve different methods to acquire knowledge and to combine these to draw conclusions; when the issues become more complex, there will be more demands for time, too

Semi-professionalism

Professionalism or rather semi-professionalism in research evaluation is an important development trend in Europe. Because of increased demand for evaluations, there is a growing body of (semi) professional evaluators. The leading figures in the European evaluation scene have a research background in science and technology policy studies and a few may still operate at university departments bringing in money for their departments through evaluation commissions. In their evaluation activities, they draw on the methods and assumptions of respective research fields and have disseminated these to the whole evaluation field. The above-mentioned Alvey Programme created evaluation

competencies which were used later on, not only in the UK but also at the European level. The evaluation experts, who gained their competence in the Alvey evaluation, are among the foremost experts in the evaluation scene in Europe.

There are also evaluation professionals with different and varied backgrounds. There are specialised consultant firms in this area (such as Technopolis Ltd), while other consultant firms have expanded their activities to new areas, carrying out research-evaluation-related studies and projects. The EU and, to some extent, the OECD have promoted this trend towards professionalism by organising evaluation workshops and conferences that aim to disseminate relevant experiences. There are also initiatives for R&D evaluation summer schools and courses (Twente University).

I have referred to semi-professionalism rather than full professionalism because, in spite of the extensive experience of many of the groups active in the evaluation area, only a few work on a full-time basis. Evaluation is a sideline activity for the rest. Those involved in evaluation also have varied competencies.

Tendering in Europe

A related development trend in national evaluation exercises is an increasing practice to send calls for tender to groups of evaluation experts located in different European countries. These groups compete with each other to obtain evaluation contracts. They often also collaborate on proposals, and sometimes the calls for tender specifically ask for joint groups consisting of evaluation experts from more than one country — from the country where the evaluation is going to take place, and from one or two other countries. This means that there are transnational evaluation teams working in this field (such as my example above). Such transnational experience may be needed because there is not enough competitive expertise in evaluation in the home country. Transnational teams are expected to bring a wider experience and knowledge base and more independence for the evaluation.

Evaluation expert teams that are fully foreign are quite common, too, particularly in the Nordic countries where foreign peer review panels have been used for a long time. Foreign evaluation groups have the advantage that they look at the state of things without the prejudices of those inbred in their own system. However, sometimes the cultural distance, which may bring objectivity, can be a hindrance in understanding the institutional set-up and division of labour in a country and can produce irrelevant recommendations, which will not lead to impacts. 'They don't know or understand the circumstances' is a frequent complaint in such a situation.

Working in a transnational evaluation team brings additional communication/travel costs because, in spite of email and other advanced communication means, people need to meet in order to be able to debate matters. The data collection through interviews and other personal contacts also requires extra costs. Sometimes, the tight time schedules and restricted budgets of evaluations confine the factual communication in a transnational evaluation team to a minimum, and transnational collaboration turns out to be just window dressing: the foreign team members are there to add to the credibility of the exercise.

Conclusions

In addition to earlier functions — such as advancing:

- accountability and transparency in the use of public money
- legitimation of the funded activities
- organisational learning
- improvement of the efficiency and effectiveness of activities
- improvement of the quality of the funded activities —

with the increasing importance of strategic evaluation, evaluation has additionally become a forum for policy debates.

The latter development has potentially a caveat: evaluation may lose some of its credibility if taking stands in strategic issues takes precedence over evidence-based judgements, which is the traditional role of evaluation. At the same time, increasing emphasis on quantitative outcome measurement may give a one-sided and misconceived view of the evaluated activities. There is a need to improve the quality of evaluation by drawing on many types of evidence, capable of illuminating both more and less easily measurable impacts and outcomes. There are also suggestions for developing the evaluation community to full professionalism through certification or quality control initiatives by professional bodies. However, such professional bodies do not yet exist in the research evaluation area.

References

L Georghiou (1999), 'Socio-economic effects of collaborative R&D: European experiences', *Journal of Technology Transfer*, 24(1), April, pages 69–79.

L Georghiou (2000), 'Evaluation of research and innovation policy in Europe: new policies, new frameworks?', in P Shapira and S Kuhlmann (editors), *Learning from Science and Technology Policy Evaluation, Proceedings from the 2000 US-European Workshop* (Georgia Institute of Technology, Fraunhofer Institute Systems and Innovation Research).

A Rip (2000), 'Societal challenges for R&D evaluation', in P Shapira and S Kuhlmann (editors), *Learning from Science and Technology Policy Evaluation, Proceedings from the 2000 US-European Workshop* (Georgia Institute of Technology, Fraunhofer Institute Systems and Innovation Research).

T Venetoklis (2000), 'Methods applied in evaluating business subsidy programs: a survey', *VATT Discussion Papers*, 236, (Helsinki, Government Institute for Economic Research).