THE INCREASING PROFESSIONALISATION OF THE EVALUATION OF MISSION-ORIENTED RESEARCH IN FINLAND: IMPLICATIONS FOR THE EVALUATION PROCESS

by

Terttu Luukkonen
VTT Group for Technology Studies, Finland

Introduction

Even though the breakdown into time spans might seem rather artificial, I would claim that there have been three stages to research evaluation in Finland. Research evaluation – in the sense of science and technology policy evaluation – was introduced in Finland in the early 1980s. In the early phases, it was largely a question of quality evaluations of both basic and mission-oriented research using panels of external experts.

Evaluation was introduced by the Academy of Finland, which is a system of research councils. The Academy started evaluating fields of research in 1982. The idea of evaluating research and the method used were borrowed from the practices of the Swedish Natural Science Research Council, although the method was applied in a slightly different way, i.e. to whole areas of research irrespective of whether or not they were funded by the Academy. A central feature of the evaluation procedure was the use of panels of external experts. In a small country like Finland, “external” means that the experts were researchers from other countries, mostly from outside the Nordic countries. A major evaluation criterion was the quality of research in relation to the international state of the art. To date, the Academy has carried out 23 evaluations of research fields and research institutes, and one evaluation of international research collaboration with an international institute (IIASA). A further evaluation is currently underway. Twenty-five evaluations, either completed or ongoing, is an impressive figure, although fairly modest considering that the first one was initiated 15 years ago. The number of evaluations is also much smaller than that envisaged by the Cabinet level advisory body, the Science and Technology Policy Council when, back in the early 1980s, it suggested that all research fields be subject to evaluation every ten years in a rolling system of evaluations.

The use of external – in practice, foreign – experts in quality evaluation was soon adopted by other research funding agencies in Finland. The ministries responsible for mission-oriented research, such as the Ministry of Environment and the Ministry for Agriculture and Forestry, adopted the procedure of the Academy in the late 1980s and applied it to the research funded by their respective ministries. Similarly, TEKES, the Technology Development Centre, a major governmental agency responsible for funding industrial technology programmes, evaluated its first generation of technology programmes in 1988-91.
The first evaluations by the mission-oriented agencies and ministries were also quality evaluations. TEKES introduced the criterion of relevance from a practical point of view and commissioned professional external evaluators, for example from Technopolis in the United Kingdom, to carry out the evaluation exercises.

The second wave of evaluations was related to the evaluation of research organisations, research institutes and research funding agencies. At this stage, the research funding agencies which had been commissioning evaluations were themselves subjected to outside evaluation. Thus the Academy of Finland (Blume et al., 1993), TEKES (Guillaume and Zegveld, 1995) and the Technical Research Centre of Finland (VTT) (Tomner and Zegveld, 1993), were evaluated by outside experts. Large public research institutes such as the National Public Health Institute (1995) and the Institute for Occupational Health were also evaluated in the 1990s. In these cases, evaluation concerned the extent to which these organisations had fulfilled their missions and, in the case of the Academy, the evaluation was initiated with the aim of achieving organisational reform. In all cases, evaluation was carried out by panels of external experts who obtained a wide range of background information; in the case of TEKES, the information included a survey among recipients of TEKES funding about the usefulness of that funding.2

By the mid- to late 1990s, a third phase of evaluation was emerging; evaluations have now become commonplace. All major public programmes for funding research and technology development and for the development of research infrastructure are subject to external evaluation. For example, the impacts of a major increase in public funding of research in 1997-99 will be monitored and evaluated on a real-time basis. The purpose of this increase in funds is, in particular, to strengthen the national system of innovation and to support enterprises, employment and the economy as a whole. The increase over three years corresponds to one-quarter of the total appropriations for government R&D funding in 1997. The money will be obtained through the sale of shares of government-owned enterprises. An example of an initiative on a much smaller scale is the evaluation of public support to regional centres of expertise3 carried out in 1996 (Ahola and Kortelainen, 1997). At this latest stage, different types of evaluation, as described above, are carried out simultaneously.

Why did research evaluations start and why were they fairly widely adopted? Originally, the Academy of Finland implemented the evaluations to provide arguments for obtaining more money for basic research by showing the uses to which funding was put and demonstrating that the research carried out was of fairly high quality. The initiative originated in the late 1970s when, after the recession of the early 1970s, there was a decrease in the funding of the Academy and the universities. In addition, most evaluations by the Academy were commissioned for specific reasons, such as a perceived need to develop a particular field or to assess major investments in equipment.

With regard to the mission-oriented agencies,4 one of the reasons behind the implementation of evaluations was the need to improve quality control; another, a wish to re-orient research activities. Since the mid-1980s, the Cabinet-level Science and Technology Policy Council has paid particular attention to measures to improve the quality, effectiveness and relevance of research financed and/or carried out by the mission-oriented agencies. Evaluation has provided these bodies with a new management tool. The Science and Technology Policy Council of Finland has systematically endorsed evaluations and recommended that governmental research funding agencies commission them to assess the effectiveness and efficiency of the research activities funded. The deep recession of the early 1990s caused tight budgets in all fields of public administration and highlighted the importance of accountability in public funding of R&D. Government appropriations for R&D decreased slightly in 1994-96 after a long period of growth (see figure).5 However, total national R&D expenditure has grown continuously throughout the
15-year period. Evaluations have been part of the overall strategy to build up a national research system, in terms of both quality and quantity.

The professionalisation of evaluation

The professionalisation of evaluation implies that professional evaluators or evaluating institutions carry out the evaluations in addition to, or as a substitute for, peer review/expert panels. The latter have traditionally been used to evaluate research quality in science. Even though peers are the best experts for this task, they are not professionals in evaluation. In professional evaluation, panels may also be used to gain a better understanding of the subject matter of the area to be evaluated and/or to provide the evaluation with more credibility and prestige. There has been a trend towards more professionalisation in evaluation, both in Finland and, in particular, in other countries. The considerations about professionalisation and its impact on the evaluation process presented below draw on examples from Finland and also from other contexts.

Along with the professionalisation of evaluation, evaluation criteria have shifted from those pertaining to quality of research towards a stronger emphasis on strategic aspects and factors pertaining to the relevance of the project from a practical point of view. The type of information used has shifted away from qualitative assessments by experts to more systematic, often quantitative, data gathering and to the application of standard social science methods for the exercise. Evaluations have become more professional, in particular where they concern mission-oriented research.

A group of professionals, consultants and evaluating institutions engaged in evaluation activities has emerged. This has occurred in different ways in different countries. In Finland, it has happened on a modest scale. Evaluations use both domestic and foreign professionals.
**Independence of evaluation**

Finland has aimed at independent and external evaluation by commissioning outside experts to carry out the evaluations. Often this has meant the use of experts from abroad either as members of expert panels or as evaluation consultants or a combination of both. Whether the experts and professionals come from abroad or from Finland, external evaluation implies that those commissioned to carry out the evaluation do not have a vested interest in it. In principle, professionalisation does not to any great extent affect the independence of the evaluation.

However, in both professional evaluation and evaluation by expert panels, there remain many possibilities for vested interests to appear – or for the institutions being evaluated to protect their interests. When using expert panels, standard practice in Finland has been to prepare the evaluation process well in advance so that all the background material has been collected before the experts visit Finland for the first and, usually, only time. There are thus few opportunities to provide them with new background information which they might find useful or to arrange interviews with a new group of people. Background information is important for formulating the questions which will be asked; similarly, the composition of the group of people who are to be interviewed strongly influences the impression the experts can gain of the institution being evaluated. While the evaluation panel has little influence on these strategic matters, the organisers of the evaluation, who are often from the institution being evaluated, effectively limit the scope of evaluation and exercise undue influence on the results. In order to compensate for this, new procedures have been tested, for example inviting the foreign chair of the evaluation group to be actively involved in the design of the evaluation and the collection of the background material (Evaluation of the National Public Health Institute, 1995). In principle, professional evaluators have greater opportunities to negotiate the evaluation mandate and specify the methods to be used.

In a professional evaluation, interdependencies may evolve between those commissioning the evaluation and those performing it. If the evaluators come from an institution which carries out commissioned studies, they may be dependent on the agency commissioning the evaluation for future projects and funding. The same applies to professional evaluators wishing to obtain future commissions. In such situations, despite sincere attempts at objectivity, the evaluators are not free to design an evaluation which is sufficiently independent, neither can they voice strong criticism – or they do so at the risk of losing future clients.

**Information collection**

Evaluations usually have a limited time frame. This can lead to hastily carried out background studies, *e.g.* surveys, which do not permit the experts to acquire a deeper knowledge of the sector being studied or to reflect on possible interpretations of the answers. At the other extreme are qualitative interviews. The drawback of interviews is that if they are not sufficiently structured their content may be difficult to digest into coherent conclusions within a reasonable time frame. Unstructured interviews are labour-intensive and their use may restrict the number of people who can be heard. The professional skills of the evaluators are important in managing different types of data. Even quick surveys with multiple choice questions and boxes to be ticked can be an effective tool if prepared with skill and with a good knowledge of the field under review. In professional evaluations, studies are a major source of information and the standards applied are crucially important for the quality of the evaluation.

Evaluation studies and data gathering for an evaluation should observe the basic principles of data gathering in empirical social sciences. All too often this is not the case. For example, the large number of
studies carried out in the framework of the MONITOR-SPEAR Programme of the European Communities, published in the 1990s, although excellent in their exploration of the various dimensions of programmes which have as their primary task to promote the competitiveness of European firms, could be improved in terms of response rates, sample sizes or the use of control groups.\textsuperscript{6} The problem is even more pronounced in the collection of background information for the evaluations commissioned by the European Union.

Other factors, such as the influence of the participant's position in the organisation on his assessment, have to be taken into account when designing evaluations. For example, in an evaluation based on surveys and interviews, programme participants are the major source of information. In smaller organisations, they are presumably in a good position to assess the utility of the programmes for their institution. In larger organisations, such as large companies, participants' viewpoints might differ from those of the management. Both groups should be listened to. It is often difficult to judge whether participants' statements are made with opportunistic intentions, such as a desire to obtain money in the future, or whether they have an inbuilt bias to assess their own endeavours favourably. In the ideal case, evaluations should consult a variety of people who have different interests in the activities being evaluated.

When evaluations concern the social and economic role and impacts of research programmes or research institutions, they touch upon issues on which our understanding is still fairly limited. Even the most up-to-date research literature in the research and technology policy area cannot provide clear-cut answers as to the effectiveness of policy initiatives or the circumstances under which research programmes and institutions will attain their intended impacts. Therefore, it is difficult to evaluate such issues: the evaluation project would effectively become a research project in its own right if it attempted to answer these sorts of questions.

For example, if carried out properly, an evaluation of the Technical Research Centre of Finland (VTT) – a large public sector research institute which carries out contract research for Finnish industry and has as its objective to raise the technological level of Finnish industries – is a far from simple and routine task. The activities of the VTT are intertwined with Finnish industries and companies in a multitude of ways so that a proper assessment of the effectiveness with which it performs cannot be made simply by a quick survey or a few interviews with key people in industry.

In order to study the effects of public technological programmes, both national and international, we need further developments in conceptual and empirical analysis. As mentioned above, evaluations have tight time schedules which do not allow for studies with ambitious objectives. In practice, this may perhaps lead to unsatisfactory compromises.

**Openness and publicity**

The Nordic countries have a tradition of openness in most areas of public administration. These principles pertain to the evaluation of institutions, programmes and activities funded by public money. The research evaluation movement, from its implementation in the early 1980s, adopted from the first a principle of openness and public reporting of evaluation findings. Publicity was important for both the transparency of the evaluation exercise and the legitimation of the activities to be evaluated. In some cases, this has given rise to unwanted and unexpected effects, such as negative media attention to criticised individuals, well beyond the intrinsic merits (demerits) of the cases considered. Another drawback is the reluctance of the evaluators to explicitly voice very critical opinions. This leads to
indirect and hidden criticism and increases the interpretative flexibility of the reports – which is not
usually in the interests of those commissioning the evaluation.

Publicity does not only act as a sanction: researchers or research groups confident of their
performance have in some instances sought to be evaluated in order to use the evaluation to support their
status and to obtain further funds.

The fact that evaluations have become part of the everyday management routine of programmes and
research institutions has decreased public and media interest in them. This is not wholly to be regretted as
too much media attention is not beneficial to the communities being evaluated. In professional evaluation,
there may be other, less desirable, threats to publicity and openness: professional evaluators may face
pressure not to publish their findings if the evaluation is critical. Panels of prestigious experts face fewer
problems in this respect.

The use made of evaluations

The use made of evaluations by the (so-called) decision makers and by the researchers themselves
was analysed in the first wave of science policy evaluation in Finland and in the corresponding period in
the other Nordic countries (Luukkonen and Ståhle, 1990; Luukkonen, 1995). I will not report the
findings in detail, but I would like, however, to highlight that the everyday notion of the use of
information – that is, instrumental use – is not the only way to utilise evaluation exercises. Instrumental
use means that the information or, in our case, evaluation results, will be used to solve a concrete policy or
decision-making problem. Evaluation findings may be used in various, less direct, ways, such as to
provide new insights into problems at hand, to formulate questions not previously asked or to throw light
on the scope of problems. A further category of use – the use of evaluation findings to advocate a
previously adopted standpoint or as ammunition to support a stand – is often morally reproved as a mis-
use of evaluation findings. It is, however, more fruitful to acknowledge that any type of information can
be used to promote various goals if there are people who have strong commitments to advance these goals.
The same applies to evaluation findings: many evaluations have special purposes and hidden agendas
which may be related to an advocative use, and they are used in such a way.

An example of this last category follows. The evaluation of the Academy of Finland was aimed at
organisational change. There had been an initiative to merge the seven Research Councils into three, in
the hope of increasing the effectiveness of the Research Councils and in response to management
problems with an increasing number of research projects in interdisciplinary areas. The initiative had
been blocked in Parliament for reasons irrelevant to the major initiative. After the 1991 Parliamentary
elections, the organisational reform was given a new impetus by the Ministry of Education who
commissioned a group of foreign experts to carry out an evaluation of the Academy of Finland in 1992.
Not surprisingly, in their report, the group of experts came to a conclusion similar to that of the original
reform initiative. In the end, the number of Research Councils was reduced but, after successful lobbying
by the medical scientists, who managed to retain their own Research Council, the number of the Research
Councils was eventually reduced to four. Why did the experts come to the same conclusion as those
advocating the organisational reform? Obviously, they were largely dependent on the information
provided by the major players in science policy who were, by and large, in favour of the reform. Another
factor is that merging separate Research Councils into one large Council or into a smaller number of
larger units seems to be the trend in this decade – at least in the Nordic countries, which have a tradition
of following each other’s examples.
The consequences of the evaluation of the VTT were similar to those of the Academy; that is, they led to organisational change. However, the situation was quite different. The evaluation (1992-93) was not initiated with any particular aim in mind, but rather as a response to a vague dissatisfaction with the ability of the institute to respond to the needs of industry – and also because evaluation was becoming increasingly popular as a management tool. The initiative for the evaluation came from the Ministry of Trade and Industry. VTT is a large contract research institute with approximately 2,500 employees. It operates under the Ministry of Trade and Industry, which set up a two-man evaluation committee (Sigvard Tomner and Walter Zegveld). The two wise men recommended, among other things, a radical reorganisation of the institute: merging 34 research laboratories into ten units for reasons of efficiency, synergy and strategic management. The question of potential organisational change had been under discussion within the institute, but it was not actively on the agenda. It is, therefore, somewhat surprising that after the evaluation the recommendation for radical organisational reform was promptly put into effect. The director general of the institute became an active “champion” for change. The starting point was more open-ended than in the example of the Academy. The outcome may be related to many factors; among others, the need for the management of the institute to assume the initiative after the Ministry-led evaluation process.

In drawing attention to the uses and impacts of evaluations, too little attention is paid to the process impacts: the fact that activities, institutions, etc., are being evaluated does have an impact on those involved in the process. It may be simply that they are forced to think of their activities over a longer time perspective than would routinely be the case, or that the evaluation provides new contacts for the people involved in the activities being evaluated. We do not have enough information on these impacts. The process effects may be useful to the evaluated institutions in helping people reorient their activities or reinforce their current plans. However, process impacts can also be negative: the evaluation process is a source of uncertainty for staff in the institution being evaluated and may be detrimental to their work motivation.

**Ethical considerations**

There are important principles which should be taken into account and which we might call a “code of conduct of evaluation”. This code includes the basic principles of competent empirical investigation and aims to produce studies carried out with skill and professional integrity. The code would set some basic requirements for those commissioning evaluation exercises, trying to avoid the most crude examples of advocative use of evaluation for predetermined purposes.

However, we have to acknowledge that evaluation plays a limited role in decision making; providing only part of the background information. Demands that evaluation recommendations be always put into practice are unrealistic and represent an unduly rationalistic view of the decision processes. This view contains an implicit faith in the rationality of the evaluation process and the superiority of the knowledge thus acquired compared with the knowledge previously held by the people concerned. Local decision makers have the responsibility for carefully judging the evaluation findings and recommendations and coming to their own conclusions. The evaluators, especially if they are outsiders, are not (at least, not always) in a position to come up with the best possible or most realistic practical solutions for perceived problems or performance failures.

Decision makers and those commissioning the evaluations will benefit from an open-minded reading of the evaluation reports. The information provided by the evaluation may provide them with new and unexpected insights; it may challenge their traditional ways of thinking and help formulate new solutions to problems. However, it is their responsibility to draw the conclusions: if the end result does not follow
the recommendations of the evaluators, this may be unsatisfactory from the evaluators’ point of view, but it is part of the local research and technology policy context.

The evaluation exercise and the evaluated activities gain in credibility if the evaluation is transparent for all the parties concerned. This includes informing both the evaluators and those being evaluated as to follow-up, potential changes in the activities evaluated, and plans concerning the monitoring of these changes.
NOTES

1. The author would like to express her gratitude to Tarmo Lemola for his useful comments and to Pirjo Niskanen and Sasu Hälkkä for their help in procuring the statistical information.

2. Parallel to the evaluation efforts described above, there has been a movement towards quantitative evaluation of university research. The quantitative evaluation movement started in the mid-1980s simultaneously with, but independently of, a movement to improve effectiveness and increase accountability in the government sector. Quantitative evaluation was from the start related to efforts to apply evaluative criteria to the allocation of university resources. Alongside quantitative evaluation, there have been qualitative evaluations of whole universities. These have been learning exercises rather than evaluations related to any decisional needs. The most recent development is a move towards qualitative evaluation, such as to define centres of excellence by qualitative criteria, and at the same time, to increase the proportion of funds allocated by performance criteria. As an indication of the importance of evaluation as a management tool for universities, the former Council for Higher Education was reformed and its name changed to Evaluation Council for Higher Education at the beginning of 1996.

3. The main purpose of the regional centres of expertise is to promote the establishment of new knowledge- and technology-based firms.

4. Mission-oriented agencies refer to the ministries and units of central administration which finance research to attain different socio-economic objectives, such as health, transport, agriculture, defence, etc. The Ministry of Education has a prime responsibility to promote the “general advancement of knowledge” through the funding of the universities and Research Councils of the Academy of Finland. The latter are excluded from the concept of mission-oriented agencies, although this chapter draws on their evaluation experiences.

5. The sharp peak and subsequent decrease in the percentage of government R&D appropriations in GDP are caused by the trends in GDP. When measured in fixed prices, the changes are small.

6. For a summary of the results, see Georghiou (1994).
REFERENCES


