



This contribution was prepared as a background paper for the meeting of the PUMA (Public Management Committee and Public Management Service) group at the OECD Symposium on 'Government of the Future: Getting from Here to There', in Paris, 14-15 September, 1999

Evaluation as Usable Knowledge for Public Management Reforms

JEAN-CLAUDE THOENIG

Centre National de la Recherche Scientifique, France

Do public sector reforms use evaluation? What good practices has experience brought to light? Does evaluation have a future? In this article, evaluation shall be defined as an instrument or means for improving the capacity to learn about conducting successful change and defining achievable outcomes in the fields of public efficiency and effectiveness. Whilst many forms of evaluation exist, in general it may be characterized as an activity that is devoted to the production and analysis of rigorous and relevant information about relationships between, on the one side, public acts and non-acts and, on the other, outcomes and impacts.

KEYWORDS: advice to decision makers; good practices; internal evaluation; learning; public sector reforms

A Seeming Paradox

It is hardly imaginable that reforms of administrative and public sector management would be developed and implemented blindly, thoughtlessly and impulsively, solely by order of the hierarchical authorities. However, this having been the case, a widespread demand exists on the part of practitioners, for there are significant deficiencies in the monitoring of changes introduced in the public sector. There is every reason to believe that evaluation is destined to play a major role in meeting these expectations, at least in part, since it provides relatively rigorous tools and a largely rational approach – at least on paper – for producing information and advice on a specific public policy.

However, a careful examination of the facts shows that, in most countries, evaluation has thus far only been used in a relatively limited and occasionally

sporadic way that has often proved disappointing. What is more, there is reason to believe that the reluctance to use evaluation more widely is not necessarily or primarily due to ignorance or unwillingness. As a result, it sometimes happens that governments publicly affirm the need for evaluation, but fail to practise what they preach. One example of this tendency is provided by the White Paper on administrative and staff policy published in Norway, which stressed the importance of evaluating the specific problems and reforms in these fields. But no initiative was launched subsequently to assess the ten years of reforms carried out thus far. Conversely and more surprisingly, one can encounter a relative scepticism about evaluation, particularly among well-informed and experienced practitioners of public management reform, some of whom have even spoken of an 'allergy' to evaluation that seems to prevail in their own governments. Such an attitude may derive from a variety of causes, such as the reluctance of policymakers to feel challenged by experts who are not 'hands on', the arrogance of evaluators in defining the goals and the content of reforms, the fact that evaluations may take up too much time, the idea that reforms are also a political tool of government, etc. The demand from policymakers seems to remain rather flat. Experience suggests that demand can be a problem: it occurs where ministers fear they will be embarrassed by the information generated by evaluators, wanting to hear only the good news and not what has gone wrong, or where the public and the media may use it as an opportunity to 'beat up' on governments (Kessler et al., 1998).

But this paradox is more apparent than real, and it would be wrong to throw the baby out with the bath water. Admittedly, from a quantitative standpoint, there is no reason to conclude that there is an irresistible trend towards a generalized use of evaluation. In this respect, in a number of countries, public management reform may seem to be lagging behind or to be relatively out of step with other government policy fields such as R&D, health care, social welfare, education or social assistance. At the same time, practice has shown that evaluation does occur and is used to provide governments with information.

Without trying to draw up an exhaustive list, we can briefly review a number of recent examples of public management reforms that were coupled with genuine evaluation work or practices. For example, in 1995 the Finnish government launched a concerted and detailed programme aimed at identifying and assessing the results produced by a series of major reforms implemented since the mid 1980s. This programme included groups of senior civil servants involved in the reforms and teams of experts, including foreign experts (Pollitt et al., 1997; see also Ministry of Finance, 1997).

France coupled its administrative modernization policy with a series of ongoing evaluations as early as the late 1980s. Ministries such as the Ministère de l'Équipement undertook an intensive and participatory evaluation effort, in particular with regard to the management of the objectives and resources assigned to their external services (Trosa, 1992).

In 1995 in New Zealand, the State Service Commission and the Treasury commissioned a US academic to make an independent assessment of the changes produced by the wide-ranging reorganization of the public sector in structural, organizational, strategic, financial and budgetary, and management control terms

(Schick, 1996). More recently, this same country produced a 'state of the State' report. Lastly, a central agency was given responsibility for the systematic collection of data on the overall performance of the public system and the individual performance of each of its agencies.

Initiatives to signal public management reforms ahead are interesting approaches that have been adopted in several countries. Such is the case in Canada where a structured and lasting approach is being carried out. An annual report is published that sets out what are considered to be the key management issues and priorities to be addressed for the coming year. The Clerk of the Privy Council Office, who reports to the Prime Minister, publishes a document which offers a vision for management reform. He and the Deputy Ministry Community have set up task forces responsible for monitoring the programme review of the reform and for investigating subjects such as service provision to citizens or management of intersectoral and interagency issues. The idea is to promote horizontal coordination with a view to improving the consistency of initiatives and their impact on the implementation of programmes in the various fields of government responsibility. This top-down approach is supplemented by a bottom-up approach, for each department is encouraged to gather its own information on the results and progress achieved in its own sector. This is aimed at adapting evaluation to the various levels within the public system. It simultaneously covers the two facets of internal efficiency and external effectiveness, and includes both structures and processes.

A fuller picture would be given if the list were not confined to formal or traditional evaluation initiatives, since some countries have practices that might be described as 'quasi-evaluations'. Without having formal status or being conducted as specific systematically designed projects, there are initiatives through which policymakers and their staff gather information and conduct reviews and assessments of various aspects of their reforms both before and after they make decisions. This more or less informal and *ad hoc* approach is found to varying degrees in many countries.

For example, in the United Kingdom reviews of various reforms have been conducted addressing specific aspects of reform programmes such as 'Next Steps', the 'Citizen's Charter' or 'Market Testing and Contracting Out'. In Ireland, the Committee for Public Management Research, which is chaired by the Department of Finance, has just conducted a partial review of the plans of the customer service section of departments and offices in a discussion paper. All things considered, although these results are limited, they are far from negligible.

Two general observations may be made at this stage. The first is that evaluation is not by nature more characteristic of a specific type of state or administrative culture, although the examples given above do suggest that it is more frequently used in countries where reforms are more comprehensive or Anglo-Saxon attitudes predominate. The second suggests that it is not sufficient to have qualified internal or external experts, reliable tools and ample information for an awareness of the importance of evaluation to spread automatically throughout the system and be incorporated into the management of reforms. Evaluation may be practised even if the public system is not permeated by an

evaluation culture. The lesson to be drawn is a relatively optimistic one: it is not essential to have experts or sophisticated systems, for individual civil servants are free to decide whether they will carry out an evaluation or not, sometimes even without realizing consciously what they are doing. In other words, any barriers to evaluation are not so much professional, technical or intellectual, but pragmatic.

Learning from Good Practices

Under what conditions can evaluation practices be used to develop and implement public sector reforms? The answer is obvious: when they are adopted and used by policymakers. But this cannot be done by applying a ready-to-use magic formula or by following a single procedural model of best practice. The examination of a number of good practices can teach some persuasive lessons.

Observation of evaluation practices makes it possible to refute the clichés or stereotypes that encourage scepticism or fatalism about the possibility of successful evaluation. For example, one often hears that by nature public organizations are not self-evaluating (Wildavsky, 1979). At another level, the problem is laid at the door of policymakers, whose way of thinking is presented as being incompatible with an interest in the practice of evaluation, with rare exceptions. And there are other supposedly sound arguments to the effect that the problem lies in the very nature of the reforms undertaken. For example, it is argued that comprehensive, authoritarian policies such as those transferring whole sectors of goods and services to the market, since they are ideologically motivated, do not encourage government to focus on the costs and benefits expected and generated. But to conclude that therefore nothing can be expected of evaluation seems just as extreme as to state that it is the inevitable wave of the future.

Usefulness for Action

The first lesson is that a pragmatic approach to the use of evaluation practices is needed. In other words, they must in each case be focused on specific needs and opportunities for action. The needs of policymakers should be kept to the fore in deciding what to evaluate. This common sense remark is not abstract in its implications. The decision to evaluate a programme is rarely made unilaterally by decision makers alone. The work of the OECD's expert group suggests two observations. Needs are often not clearly formulated beforehand by policymakers and, when they are, the senior officials involved in the reform play a significant role in taking the initiative to use evaluation practices (the role that the budget department or ministry can play in this regard should also be mentioned). In some countries the willingness of senior officials and advisers to use evaluation practices is no coincidence – being closely linked to the fact that their university and professional training has made them aware of the contributions of the social sciences and modern management – unlike staff with strictly legal and administrative training.

There is little likelihood that evaluation will be adopted for its own sake, for it is not an end but a means to an end. It will be credible and acceptable only if it meets three conditions:

1. it must be sponsored by individuals and groups that are recognized as having practical experience with evaluation as well as direct access to policymakers, or even a policymaking role in the reform (both before and after its implementation);
2. it must be based on a need or concrete problem that policymakers face and from which they will derive value added at their level; and
3. it must be well-timed in relation to reform.

These windows of opportunity often prove to be essential, especially when the practice of evaluation is still uncommon within a country, but also in public systems where it is already more widespread. Poor timing can kill a reform initiative. If evaluation is out of step with the governmental agenda, it will ultimately have no impact and merely be filed away and forgotten (this is what took place in France in 1992–7). By the same token, this also means that evaluation systems must remain flexible and adaptable to changing needs and that policymakers must have easy access, if not close control.

The First Priority: to Provide Information

The second example of good practice concerns the focus and content of evaluation practices themselves. If the use of evaluation is action-oriented, this means that evaluation focuses on providing usable knowledge. More concretely, it is much more likely that evaluation will be accepted and be of use if the aspect of providing information is stressed, while being cautious about the aspect of reaching an assessment or judgement. In this regard the distinction between evaluation and quasi-evaluation, although it makes a theoretical distinction that satisfies methodological purists, is artificial and detrimental.

A usable evaluation is first and foremost one that is aimed at making available information based on empirical data, as the examination of the practical experiences of public service reform has repeatedly shown. The reason for this is clear, and lies in the very nature of the decision-making process. More specifically, two significant facts can be distinguished.

Public decision makers are much like corporate executives (Mintzberg, 1980). They give priority to practical or qualitative information obtained by speaking to individuals they trust. This is a far cry from the theoretical model that assumes that problem-solvers take the time to think the problem through by analysing and exhaustively reviewing all the information on the specific empirical situation and the quantitative merits of the alternatives available. This means that evaluation will be more credible if it is adapted to the reality of the decision-making process.

Be this as it may, policymakers engaged in action do not stop thinking. Analysis – or evaluation – is one of many inputs they use, such as ordinary knowledge, learning, interactive problem solving, etc. Consequently, analysis must compete with these other inputs (Lindblom and Cohen, 1979) and is not automatically given priority.

These two facts point to a concept of evaluation as an activity that is relatively limited in scope, focuses on clearly defined problems, employs language that policymakers can understand, readily uses the data available even if they are not

perfect, and aims at describing a state of affairs rather than analysing it. Evaluators know how to be responsive to conditions that are different in each case and to use data, not necessarily as a competing argument, but at the very least as a means of getting the policymakers' attention, by providing information before explanations, and assistance rather than judgements.

This attempt to give evaluation greater credibility assumes that evaluators do not embark lightly upon making judgements and assessments. This is particularly true when they must evaluate on-going reforms, for they run the risk not only of substituting their own judgement for that of policymakers – a technocratic deviation – but of failing to give an objective account of the situation, in particular by focusing on mistakes, dysfunctions or deficiencies, without balancing them fairly against achievements, progress and successes, which amounts to a pessimistic bias.

This selective short-sightedness of on-going evaluation as to the real impact of a policy is encountered in a number of reform initiatives. For example, the Canadian government is extremely cautious about auditors' reports, which tend to emphasize the shortcomings of a reform – in terms of value for money as well as external effectiveness – and give the impression that little progress is being made and a great deal remains to be done. In such cases evaluation is of no practical help to governments and, because of its overall assessment function, selects information that makes it difficult to design the next stages of reform, thereby becoming in a sense self-defeating.

Should evaluations make judgments? The debate remains rather open inside the professional community and the reform entrepreneurs. Some practitioners prefer that evaluations should not explicitly spend time and energy making judgments – and instead remain either exploratory or informational, being nevertheless aware that often judgments are implied even if not spelt out. But equally strong views expect evaluators to make judgments: not to do so would reduce the value of learning from evaluation – especially when ministers want to get a clear view on a situation. Learning implies judgments.

Internal Evaluation

The third example of good practice is internal evaluation, which covers self review as well as external reviews commissioned by academics and private experts. Policymakers responsible for public service reforms will be likely to use evaluation and will find it that much easier to do so if evaluation practices are developed within the public system itself, in particular at the various levels that initiate, design and implement the reform.

A number of countries have commissioned private or academic experts to conduct evaluations. These kinds of evaluations tend to be less useful to governments. They remain somewhat theoretical at bottom, inasmuch as their approach focuses on aspects or themes that do not match the specific concerns of governments, the actual agenda of the reform, the pace of policy implementation and the capabilities of policymakers. Their overall evaluation may be perceived as critical or passive, for it is an *ex post* assessment made a number of months or even years after the actual events, and therefore provides few guidelines and directions for the next stages in the field. The limitations and

frustrations are seen more clearly when, as in Finland for example, internal evaluations are carried out concurrently, in particular by groups or networks of civil servants directly involved in implementing the reforms. The value added for action is comparatively greater in this case. In all fairness, it must be pointed out that an external evaluation can be a solution in exceptional cases, as when a government finds it politically expedient to have outside experts force it to accept a public sector reform agenda that it will then put into practice. This was the case with the reform of the management of EU structural funds adopted in Greece in 1994–6.

But it is necessary to define just what is meant by ‘internal’ evaluation. Some countries have administrative bodies such as the audit board, or *ad hoc* bodies such as administrative reform task forces or public policy evaluation units. This is or has been the case in France, the Netherlands and the UK. But the fact that technical and human resources are available within the public system does not necessarily guarantee that the evaluation of public sector reforms will be assigned to them or that their work will be relevant or have a significant impact in practice. It has sometimes been suggested that evaluation should foster the development of separate balancing centres of authority within the public system. This approach is, in the present case, relatively unrealistic. Ownership of evaluation is not the same as ownership of public management reform policies. Internal professional corps do little to make evaluation more useful to policymaking. There would still be the same barriers as were mentioned above, since the evaluator can easily fall into the role of a judge who pronounces a favourable or unfavourable verdict, without being held accountable.

The operational implications of these observations are clear. In many ways public management reforms are somewhat different from other policies (which may concern more visible or tangible social issues such as health, education, etc.), and are also considered highly symbolic manifestations of governments’ discretionary action. Accordingly, if evaluation is to be useful, policymakers themselves must take it into account and tailor it to their needs and wishes. The trend towards internal evaluation reflects the determination of those responsible for reform to maintain an on-going review of individual agencies and the public sector as a whole. Otherwise evaluation will become a mere bureaucratic ritual and the reports of experts will simply be filed away.

An Incremental and Opportunistic Approach

Good practices suggest that public evaluation should be adopted incrementally. As was mentioned earlier, anyone can carry out an evaluation and a quasi-evaluation can be a perfectly good way to start. There is no surer way of stifling evaluation at birth than to confine it to the ghetto of methodology. Moreover, the practice is not the exclusive prerogative of top policymakers and their immediate staff. There are ways of disseminating evaluation throughout agencies and down the chain of command, provided that it is linked to a problem that is relevant to those who are carrying out the evaluation. This broad-based evaluation is found in cases in which the reform is aimed at the internal structures and

processes of the public system and is implemented either through a participatory and decentralized approach or through on-going government monitoring of a reform programme. Examples of this kind of evaluation may be found in a specific sector or ministry – for example, in the Ministère de l'Équipement in France – or throughout the public system – as is the case in Canada for the federal government as a whole. The incremental approach also means that the information obtained and its uses for reform make sense to all concerned, from policymakers down to rank-and-file staff, and are not perceived either as a threat or a constraint, but as a resource and an opportunity.

Undue importance should not be attached to the issue of different types of approaches or methods. Good practices do not reveal that some methods or mixes of approaches are intrinsically better than others. However, as was stated above, provided that the issue of usefulness for action remains the basic criterion, i.e. technical considerations are based on how the evaluation will be used, the following observations may be made:

- from a technical standpoint, evaluation should take into consideration the empirical data that are available or can be obtained rapidly at low cost, knowing that public systems have relatively few data on performance and effectiveness;
- evaluation should aim less at providing an overall understanding of a vast range of parameters than at producing indicators on a few well-defined aspects that can be considered as reasonably useful proxies;
- *ex post* evaluations – evaluations made once a specific reform policy has been completed and ended – are relatively rarely used for they do not really meet the needs of government action and are expensive and time-consuming. Nevertheless they may contribute to fuelling a learning perspective about the reasons why such and such outcomes and impacts were generated and the relative value of the tools and processes that were mobilized;
- *ex ante* evaluations – or evaluations made while reform programmes are being elaborated by the policymakers – are closer to quasi-evaluations or informal evaluations: they may consider prospectively the comparative outcomes of various options which could be adopted. They actually remain a relatively secondary source of information for reasons linked to time, pressure or the political visions that policymakers have in mind;
- on-going evaluations are better targeted as a managerial tool although they require relatively intensive and well-informed monitoring. That said, they are ill suited for unilateral 'top-down' or 'one-shot' policies, in particular when the reform consists of a transfer to the market sector;
- policies aimed at ensuring the year-to-year continuity of reforms of the internal processes and structures of government departments leave greater leeway for the use of more formal and technically diverse evaluations;
- evaluation practices can be developed far more easily if there are existing systems of information on costs and performance, although they do not stand in the way of developing quasi-evaluation.

Performance Management as a Learning Process

On paper, many aspects of the public sector's activity can be evaluated, such as productivity and the quality of services, the ratio of expenditure to goods produced and services provided, and the extent to which social, economic or other problems are solved and the goals of reform are achieved. But in practice in most countries, evaluation tends to emphasise what can be called the internal aspect, which concerns the performance management of staff and departments. This is the case in Ireland, Norway, Canada and New Zealand.

This general approach, which is more managerial than strategic and more oriented towards efficiency than effectiveness, is intended to close the information loop and keep policymakers informed about what is working and what is not. It is often value-for-money oriented and is designed to provide information that will be used by decision makers to improve reforms in the future.

This realistic managerial approach meets decision makers' needs. It also predominates because it is generally difficult and even arbitrary to evaluate reforms in terms of their overall external effectiveness. This is why the aims and objectives of public sector management reform are rarely set in a clear and operationally measurable way when a policy is defined or launched.

It cannot always be assumed that indicators and data on performance will be available. There are sometimes no structured information systems that measure time spent, departments' outcomes, costs and the intrinsic quality of services. It becomes difficult to say objectively what is being done or has been done by a given organization over a given time, not to speak of comparing organizations and agencies with each other. At times, the attempt to gather information clashes with existing systems, which are not designed for performance management but for checking compliance or monitoring budgets. Consequently, there is a risk that data collection will require extra time and money. The good will of agencies responsible for monitoring is a far from negligible factor in this regard. But to argue that it is pointless to undertake evaluation because adequate management data are lacking is to enter into a circular argument about cause and effect.

The advantage of stressing quasi-evaluation aimed at providing performance proxies rather than assessing external impacts and overall performance is that it makes it possible to avoid the data collection/policy evaluation dilemma, through a gradual learning process in successive stages that fosters an information culture among policymakers and agencies, without calling on outside statisticians, computer specialists and accountants. In this way evaluation has a pump-priming effect. One possible starting point could be to carry out *ad hoc* reviews using the available data, while realising the relative value of the instruments being used and the analysis made. A further stage would be to organize this process on a much larger scale. For example, a central agency could systematically gather performance data on the system as a whole and on the constituent agencies (as in New Zealand), or an administrative modernization office might include performance data on the key public management issues to be addressed in the coming year in its annual report (as is done by the Clerk of the Canadian Privy Council Office), etc.

Above all, performance evaluation must not be viewed as consisting of a comprehensive, centralized system run solely by specialists. Good practices naturally

lead to performance evaluation as a living management tool. In other words, culture and people are its core components, and production of information is merely an outcome or means to an end. The goal is not only to raise people's awareness and to disseminate a new kind of focus on performance, cost, quality and the relevance of the services provided, but also to give agencies and staff the capacity to evaluate themselves. Using this perspective, the goal of a structured approach to performance at all levels of the public system is achieved by enabling each level to produce the information it needs for its own day-to-day decision making and to conduct a self-review that will have an impact on the quality of its everyday work. Learning-by-doing evaluative performance review makes it possible to give credibility to evaluation based on factual data.

A response that works particularly well, which is suggested by the experience of Ireland, Canada and France, consists of a decentralized and participatory approach to evaluation. A performance culture can only be mastered if agencies are encouraged to collect and share information on best practice, to promote such practices and to encourage other agencies to do the same and to adopt them in turn. The sharing of all types of best practice is a major vehicle for reform. It involves participants, and makes it possible to draw lessons as to whether these practices work in different situations and structures. The lessons will have been learnt when actors have acquired the capacity to see how a technical innovation is relevant to a problem they face at their level and can provide useful new solutions for their day-to-day work.

Central policymakers and units the main mission of which is public management reform play an important role in this regard. They have a two-fold responsibility that consists, first, of ensuring the accountability of basic units through vertical reporting, as well as between departments through horizontal communication. The centre should monitor the 'big picture' performance of the public sector. Second, they must help agencies and bureaus to learn how to take initiatives and give them greater visibility and even provide incentives and support, for example through training schemes, training seminars or voluntary experiments. In practical terms, this obviously means that the centre must have the necessary professional skills at its disposal. It also means that it must be careful to strike the right balance between uniformity and diversity, combining a mix of approaches that fits the needs of each national context. Being responsive to local units (their inventiveness and diversity) goes hand in hand with developing overall coordination. The use of informal evaluations that are designed to be part of 'business as usual' often makes it possible to introduce adjustments and changes without attracting the attention of internal or external vested interest groups which might be tempted to use a formal evaluation as an opportunity to resist reform.

In summary, the examination of good practices suggests that evaluation can increase efficiency (the degree to which goals are reached relative to the available resources). It also makes it possible to generate other forms of added value that are far from negligible. It gives visibility to a judgment or a measurement in terms of adequacy (the degree to which the goals assigned are reached). It supports the administrative memory of action solutions. It provides skill variety for

decision makers at all levels. It supports an attitude of wisdom (acting with knowledge while doubting what one knows), and helps create competition for status based on managerial skill. It may to some extent impress users and outside observers.

Reasonable Optimism for the Future

The reform of the public sector is now becoming an on-going task of governments. It is very likely that the days when it sufficed to decree a reorganization of government once in a generation and then resume day-to-day routines are gone forever.

At the same time, the pressures for change are becoming increasingly global and constraining, which has put countries that have been relatively reluctant to adopt reforms in a difficult position. Two factors should be underscored in particular.

The first is the extension of forms of partnership in the governance of public affairs. The best example of this trend is the European Union, which has become a public entity that exercises responsibilities that until now have been reserved solely for countries, in particular by allocating financial resources on the basis of programmes or projects (structural funds etc.). The development of this partnership among a number of countries has involved a remarkable extension of the use of evaluation procedures.

The second factor concerns financial constraints. There is every reason to believe that these constraints, far from lessening, are in fact going to increase in the public sector and for some time to come. The European Union is again a case in point. The Maastricht and Amsterdam Treaties establish a macro-budgetary regulation mechanism that will have a considerable impact. National systems, including subnational authorities, will thus be forced not only to limit their expenditures, but also to contain their revenues drastically. These trends will reinforce a centralized approach that will rely on control by budget agencies. As a result, whether they like it or not, both national ministries and local authorities will find it necessary to step up their efforts to rationalize, beginning with their internal efficiency (costs, productivity, quality).

These developments suggest that, although further privatization of public services remains operationally possible, efforts will be focused mainly on internal reforms of the public sector, such as eliminating duplication and simplifying institutional units in various fields etc. It will be increasingly difficult and costly for a country to remain outside this generalized trend of reform. In other words, at least in OECD countries, the current disparity between a few pioneering countries and other more cautious countries can be expected to narrow. The competitiveness of public systems, which has so far been a somewhat rhetorical aspect of reform policies, is very likely to become a very real imperative that will be felt even in the day-to-day life of individual government departments. At the same time, it will become easier to compare the performance of departments and various levels of government, both at the international level and within each country. Although the reform of government departments is still a policy in which, compared with other fields such as health, R&D etc., comparability (and

thus the reference to good practices and benchmarking) has been relatively rare (as has evaluation), the situation is likely to change rapidly. It will now be more difficult to refuse to carry out some public sector reforms by arguing, as is often done, that they are based on ideological or political considerations. Reform is becoming a functional imperative in and of itself.

Consequently, the overall situation appears to be shifting towards a relative extension of evaluation practices: comparability, the need for more on-going fine tuning, large-scale partnerships and macro-budgetary regulation will all play a role. There is no reason to believe that, as regards the good practices identified above, there is a substantial change in the profile of evaluation itself. It will remain more oriented towards internal efficiency than towards external effectiveness, more internalised than externalised, informal although more systematic, and will be accompanied by cultural and behavioural changes while being increasingly linked to solving the concrete problems of policymakers etc. Evaluation is and will remain a valuable tool for anyone willing to take advantage of it.

The lesson to be learned from the experience of evaluation in the late 20th century is one of both realism and modesty. Some may find it disappointing. One might have thought that the pressure for democratic accountability would lead to greater transparency, collectively shared by citizens regarding the impact of public reforms on society, or that the use of more 'scientific' guidance methods would make it possible to implement rigorous comprehensive data systems more rapidly. But there are also reassuring aspects. The guidance of reforms remains broadly under the realistic control of governments and the use of scientific methods has not made policymakers any less accountable. In this regard, evaluation has not gone the way of PPBS (Planning, Programming and Budgeting System), which was a good idea in itself, but proved unusable because it was alien to policymaking practices and arrogantly ignored the judgment of public officials. On the contrary, evaluation teaches an optimistic lesson through its emphasis on the principle of usable knowledge. In this regard, it is both its focus on describing the specific circumstances of policy action and its ability to foresee cause and effect – if I choose this solution, it will probably have this impact – that make evaluation a realistic tool for action. Is it futile, naïve or arrogant to imagine that reforms of public sector management:

- are aimed at genuinely changing the 'bottom line' of the day-to-day behaviour of government employees and in the way organizations really operate;
- can choose the processes, structures and strategies that actually bring these changes about;
- can make these changes lasting;
- can limit the unwanted effects these changes may have on the efficient and equitable provision of goods and services to society?

The focus on empirical information and consequences of reform actions that modern evaluation helps to provide may offer safeguards against misdirected efforts, especially in an era in which unsuccessful reform may well prove to be as unsatisfactory and harmful as the refusal to undertake a reform at all. The optimistic lesson for the future is that more and more countries have entered a

process of cumulative learning about reform actions. The quest for efficiency, effectiveness, transparency and democratic accountability does not any longer stop after one and only one reform step. Learning means an ongoing and organic process to which evaluation can offer critical insights, rather than being treated as a one-off exercise. To some extent, and as suggested by examples given above, the analytical aspect of government has improved recently even if external, independent evaluations have lessened. The 21st century may show that analysis is part of government even in the field of public sector reforms. It takes a certain maturity, sophistication and mindset to use evaluative information well, for instance as a tool for learning that can help governments and the public to form well-based views and take informed decisions. Setting up learning bureaucracies has become a major challenge.

References

- Kessler, M. C., P. Lascoumes, M. Setbon and J. C. Thoenig (1998) *L'Evaluation des Politiques Publiques*. Paris: L'Harmattan.
- Lindblom, C. and D. Cohen (1979) *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, CT: Yale University Press.
- Ministry of Finance (1997) *Public Management Reforms: Five Country Reforms*. Helsinki: Edita.
- Mintzberg, H. (1980) *The Nature of Managerial Work*. London: Prentice Hall.
- Pollitt, C., S. Hanney, T. Packwood, S. Rothwell and S. Roberts (1997) *Trajectories and Options: An International Perspective on the Implementation of Finnish Public Management Reforms*. Helsinki: Edita.
- Schick, A. (1996) *The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change*. Wellington, NZ: State Service Commission.
- Trosa, S. (1992) 'La modernisation est-elle évaluable?', *Politiques et Management Public* 10 (Dec): 65–84.
- Wildavsky, A. (1979) *Speaking Truth to Power: The Art and Craft of Policy Analysis*. Toronto: Little Brown.

JEAN-CLAUDE THOENIG is Director of Research at the Centre National de la Recherche Scientifique (Groupe d'Analyse des Politiques Publiques) and a professor at the Institut Européen d'Administration des Affaires (INSEAD). He has chaired the Conseil Scientifique de l'Évaluation (France).
ADDRESS: 11 Boulevard du Temple, 75003 Paris, France.
[email: jean-claude.thoenig@gapp.ens-cachan.fr]