

Chapter 8

The Management of Ex Post Public Policy Evaluations

by

Nick Johnstone¹
National Policies Division
OECD Environment Directorate

and

Tom Tietenberg¹
Department of Economics
Colby College, Maine

I. Introduction

The previous chapters have revealed that the 'management' of policy evaluation extends well beyond the specific technical tasks associated with undertaking the evaluation exercise. Indeed, for evaluation to be effective it must be integrated in the public decision-making process more generally. As such, in this concluding chapter the focus extends beyond the evaluation process itself, to a broader discussion of how public policy evaluations fit into public administration more generally. An effort is made to link some of these issues to the case of the ex post evaluation of tradeable permits. It draws upon the case studies in the preceding chapters, as well as the aforementioned 'policymakers' roundtable', and a review of the secondary literature.

II. Policy Evaluation as Performance Management

It is important not to see ex post policy evaluations as a technical exercise which is a substitute for political decision-making (CEC 2002). The role of policy evaluation is both narrower and broader than this would imply. It is narrower in the sense that evaluations are only one input into the decision-making process, and it is broader in the sense that evaluations should be seen to encompass more than just a technical exercise.

It has been argued that undertaking evaluations has the following three broad benefits for public authorities (OECD 1999b, CEC DG Budget 1997):

- Improving the effectiveness and economic efficiency of those public policies which are already in place;

1. The views expressed in this chapter are those of the authors, and not necessarily those of the OECD or its Member countries.

- Discouraging the introduction of future policies which are likely to be ineffective and inefficient; and,
- Increasing the accountability and transparency of public decision-making, including communication with stakeholders.²

Thus, in addition to being an efficient tool of decision-making, systematic policy evaluation is also an important element of performance management within public administration (see OECD 1999b). By encouraging policy-makers to see their role as defending the broad public interest in a manner that is subject to a transparent assessment, the systematic use of policy evaluations can result in a culture shift within public administration. Indeed, the Australian Government has argued that the process of evaluation in general can be of more far-reaching benefit than any precise estimate that arises out of a particular evaluation (see Holmes and Argy 1997).

However, incremental and instrumental uses of evaluations need to be distinguished. In some cases the results of evaluations may translate directly into policy recommendations and, hopefully policy reforms. Such cases are likely to be exceptional. Instead, the 'culture' of evaluation may make incremental processes of reform more likely (see Boyle 1993). Changes in the direction of public policy arise out of a continuous process of collection and dissemination of evidence of the desirability of change.

III. Timing of Evaluations to Encourage Policy Reform

In order to have a policy impact, the evaluation procedure can not be divorced from decision-making processes more generally (Deighton-Smith 1997). In temporal terms this means that in order to have an impact on policies, evaluations need to be linked with the policy cycle. In the absence of any window of opportunity for reform for bureaucratic or political reasons, evaluations – no matter how revealing – are unlikely to have significant influence on policy development (OECD 1999b).

In some cases, it is felt necessary to pre-programme the schedule for evaluations. For instance, in France the "Law on Interior Transport" requires that large publicly-funded infrastructure projects must be evaluated at least 5 years after they commence operations (Delache 2003). In Japan, it is recommended that evaluations are undertaken between five and ten years after the introduction of a policy (Japan MPHPT 1997, Article 7). In the United States, U.S.C. Section 7612 requires the EPA Administrator to conduct a comprehensive analysis of costs and benefits of all major provisions of the Clean Air Act (including, but not limited to, Title IV which covers the SO₂ allowance programme). An updated report has to be submitted to Congress every 24 months. Some programmes have specific pre-specified conditions that trigger an evaluation. RECLAIM, for example, required an evaluation of the programme if and when the permit price hit a particular trigger. When the condition was met, the evaluation was conducted and reform actions undertaken (Harrison 2004).

It has even been proposed to include 'sunset clauses' with the continuation of policies – particularly those with important implications for public expenditures – dependent upon the outcome of evaluations (see OECD 1997d). More generally, linking evaluations with the budget process can further encourage impacts on policy development (OECD 1999b). For instance, in Australia a ten-year reform programme has been put in place under the National Competition

2. The OECD's Public Management Service has also argued that it can be an effective means of improving co-ordination across public sector departments (OECD 1997).

Policy with three formal reviews. To access these funds (billions of dollars), State and Territory governments have to meet performance targets relating to aspects such as pricing, separation of regulation from service delivery, and the development of competitive markets (Young 2004).

Some countries have gone even further, not leaving implementation of evaluation recommendations to chance. For instance, in Japan the Ministry of Public Management has a legislative responsibility to ensure that ministries and agencies 'act upon' the results of evaluations undertaken. Thus, if an evaluation is undertaken which indicates significant flaws and efforts are not made to rectify these faults the MPHPT has a responsibility to ensure that policy changes are introduced (Japan MPHPT 2002, Article 17).

In the case of tradeable permit programmes, many workshop participants emphasised that most of these programmes evolve considerably over their lifetime. Apparently both participants and administrators experience a considerable amount of "learning by doing". Thus in many cases it is important to allow potential for adjustments in programme choices. However, the necessity to provide flexibility for evolution can conflict with the need to make the programme rules sufficiently stable that investments in pollution abatement and resource exploitation, etc. can be justified. The effects of sunk costs (both public and private) need to be considered. Thus, the trade-off between the need for discretionary power to make adjustments, and the benefits of stability of a system creates a tension that needs to be resolved on a programme-by-programme basis.

One solution is to assure that the initial programme is not so laden down with restrictions that it can not effectively evolve. While the evidence seems to suggest that most restrictions diminish over time as familiarity increases, as the wetlands credit trading example suggests (Shabman 2004), when the initial design deviates so far from an effective market, the transition may be very difficult and inefficient. Indeed, sunk costs associated with the introduction of the programme may make reform more costly than adopting a new policy altogether.

Thus, relatively early targeted evaluations can prove to be very useful. Many countries introduce tradeable permit schemes for which they have specific concerns *a priori*. No doubt the Danish administrators were worried about market power in their CO₂ trading programme (since they only had eight firms) and British administrators were concerned about environmental effectiveness (since they had some permits denominated in relative terms). As such, with limited resources countries should focus their mid-course efforts on the elements of greatest concern.

In Australia, legislation mandating an assessment of the water entitlements trading programme at the end of two years was a clear indication that interstate water trading was a new area for States to enter into, and that regulatory authorities were concerned about its effects (Young 2004). In particular, since each State defined water entitlements in a different manner, they were concerned that the system could result in serious problems. Use of this cautionary approach enabled the States and the Commission to support the process without having to make a long-term commitment. The Australian national water reform process committed States to progress towards development of an open market and this mechanism enabled them to explore the implications of moving to a Basin-wide market.

The value of periodic evaluations is clear, particularly since mid-course corrections seem more common in the area of tradeable permits. For instance, the nature of the harvesting right was changed in the New Zealand fishery (Kerr 2004), banking was added to the lead banking programme (Nussbaum 1992) and, an additional 10% reduction in swine was added to the Dutch nutrient quota system (Wossink 2004). Since evolution is so common for tradeable permit programmes, a strong case can be made for thinking of mid-course corrections as routine.

Satisfying the special information requirements of the mid-course requirements implies the need for necessarily more limited, but more focused, evaluations at a relatively early stage in programme development.

IV. Ensuring Adequate Data Collection

Ex post evaluation depends crucially on built-in mechanisms for data collection (OECD 1997a). Data collection should be closely linked with the implementation of the programme. Indeed, there should be an 'in-built system' in place which ensures that data collection necessary for ex post evaluation is defined from the outset (see OECD 1997a). Ideally, the monitoring system should be the first source of information for the ex post evaluation of public policies (CEC DG Budget 1997). In addition, there should be strong links with the data used in ex ante evaluations. In effect, data necessary for ex post evaluation should be identified at the point at which the policy is initially appraised, and certainly prior to introduction of the policy (HM Treasury 2003).

Tradeable permit programmes present specific issues for data collection requirements. For instance, for programmes where permits are distributed gratis (rather than auctioned off), both the identification of the eligible participants and the basis for their allocation (typically based upon historic use) must be established before the initial allocation can be made. Monitoring and enforcement of the operating programmes requires the routine collection of data on transfers and actual emissions or use. Other important data, such as prices, may not be routinely collected, but may be easily obtainable from other sources such as brokers.³

Three of the case studies provide rich examples of how very useful data became available once the system was up and running. New Zealand's ITQ programme (Kerr 2004)⁴, RECLAIM (Harrison 2004) and the SO₂ allowance programme (Ellerman 2004) all report on rich data sets that were able to be assembled from data gathered by the normal administration of the program.⁵ In such cases the costs of performing an evaluation, relative to other instruments where dedicated data collection exercises are required, are likely to be much less.

However, other important data may only be become available as a result of special data-collection efforts. Examples in this category could include surveys of participants to discover the sources of transactions costs or motivations for trade. For full benefit/cost analysis it could also include the monetised environmental benefits or damages. Finally some unobserved data can be

-
3. Prices were routinely collected in the New Zealand ITQ system, but not in the SO₂ allowance programme. In the SO₂ allowance programme, for example, trades can either take place at the annual auctions conducted by the Chicago Board of Trade or privately. While the former prices are routinely made available, the firms who broker the private trades have made aggregate price data available for those trades as well.
 4. For example, Kerr (2004) reports that her group was able to assemble a considerable amount of information on the New Zealand ITQ system over the period 1986–1999. Their data set includes information on the name of each fish stock, quota transactions (i.e., prices and quantities of quota leases and sales), the export prices of fish species covered by the ITQ system, quota ownership, the total allowable commercial catch (TAC) and actual catch for each fish stock, biological information on fish species, climatic variation, and interest rates.
 5. A fourth case (Pedersen 2003) had the opposite experience. Denmark had 10 years of good data available before the implementation of their programme and so it has not felt the need to move to continuous emission monitoring.

inferred from observed data. This would include marginal control costs for pollution abatement, which can be inferred from the prices revealed by transfers.⁶

Special features of a tradeable permit programme may also affect the quality of the data. Ellerman (2004) and Harrison (2004) both point out that continuous emission monitoring was required of all (the SO₂ allowance program) and some (the RECLAIM programme) of the sources. Though some initial problems were encountered in the case of RECLAIM, the current data seem highly reliable, although the means by which it is disaggregated temporally and spatially does complicate undertaking some kinds of evaluations.

An unexpected, but occasionally significant, quality impact may come from improvement in the emission inventories used in programmes which are voluntary, or baseline-and-credit schemes. Since participation in the programme can confer some significant economic benefits, sources want to validate their inventories in order to participate. This was cited as one of the major benefits of the introduction of a programme to control particulates in Santiago, Chile (Montero et al., 2002).

While the data specifically needed for administration of a tradeable permit programme (i.e. transactions) are likely to be routinely gathered, the same cannot be said for all data. McLean (2003) points out that the collection of data deemed not to be essential for running the programme – but which are likely to be useful for evaluation – are frequently the first thing cut when budgets are tightened. Inconsistent collection over time could leave important holes in any time-series data sets. How serious a problem is posed by missing data will depend upon the elements to be evaluated, and the means of evaluation.

V. Providing Access to Data for Evaluators

Access to data is also important (Smith 2002). This is one area in which there are significant variations between countries. It would appear that it is only in one OECD country (the United States) where it is possible to undertake comprehensive evaluations of policies without requesting access to official data. Indeed much of the data necessary to undertake cost-effectiveness studies of a wide variety of policies are available on different government websites.⁷ In other countries, evaluators will be required to request access to official data. This is one of the reasons why internal evaluators are preferred in some countries (see below).

Thus, the fact that programme administrators may gather data, of course, does not mean that it is always made available to independent researchers or even other arms of the government responsible for evaluation. The case studies themselves provide considerable support for the view that independent evaluations are an important component of ex post evaluation and therefore it is important to provide data access if outside objective reviewers are to contribute.

6. Theory suggests that permit prices will be equal to marginal abatement costs. In markets with high transactions costs and few trades this will be a very unreliable indicator.

7. Indeed, due to the availability of data, many CBA's and CEA's have been undertaken in the United States without the regulatory authority even being aware of their existence. Such 'openness' can bring incalculable benefits in terms of improved policy design if policymakers incorporate their results in their overall assessments of a policy.

For emissions data, the SO₂ allowance programme represents the 'gold standard'. No other programme provides the amount and quality of data available to the general public.⁸ The programme also provides researchers with a high degree of access to other data such as basic heat input and permit transactions. Ellerman (2004) reports that his MIT team had no difficulty obtaining any of the data necessary for his team to do their evaluations from the Environmental Protection Agency, the Department of Energy, and the Department of Commerce web sites.

Granting access to such data without requiring permission of the authorities seems to be exceptional, but very productive. Discussions at the workshop suggested that this proclivity for providing very transparent systems may be a main reason why American environmental policies (CAFE, Green Lights, SO₂ allowances, etc.) are so much more thoroughly evaluated. The EPA can benefit (and has benefited) from these independent ex post evaluations despite having little control over them. The costs savings can be considerable. Indeed, a simple review of the academic literature may give regulatory authorities the basis they need upon which to evaluate at least some aspects of environmental policies without commissioning the work themselves.

Other programmes have made an impressive amount of data available to at least some independent researchers with mutually beneficial results. The New Zealand experience is instructive (Kerr 2004). Their analysis was done in two stages. The two cooperating research groups funded the first stage out of internal funds. During this phase they negotiated a confidentiality agreement with the Ministry of Fisheries that not only provided access to some of their confidential data, but also made available a considerable amount of indirect support through access to staff. In the second stage their work has become partially funded by the Ministry of Fisheries. This closer relationship has reportedly further facilitated data access and the researchers no longer pay even nominal sums for data.

According to Kerr (2004) the fact that the Ministry of Fisheries did not initially fund their effort made it easier for their research team to gain the confidence of both industry actors and the environmental community. Thus, as noted above, sources of funding are important irrespective of the institutional location of the evaluator. Moreover, the New Zealand experience also supports the view that if evaluations are to be credible they must not only be independent, but they must be seen to be independent. And finally, since the policy has been widely perceived to be a success and most of the players involved are quite proud of it, administrators were very supportive of research that could rigorously show its value.

A third example is provided by Wossink's (2000) study of the effect of policy uncertainty in the Dutch nutrient quota program. Some of the data used in that analysis were not publicly available, but were obtainable for this independent study from both public and private sources. Since this increases the cost to the researcher, the probability of evaluations being undertaken more widely are correspondingly less.

Sharing information about basic data facilitates independent ex post evaluation, particularly if rather detailed, disaggregated information is made available. However, making public such detailed information may compromise legitimate commercial concerns raised by market participants. One response to this is to aggregate the data in such a way that commercially sensitive information can not be traced back to individual participants. This protection comes at a cost, however, since aggregating data removes a great deal of variation that might prove analytically important in understanding the actual workings of the system.

8. McLean (2003) points out that the public can now have real-time on-line access to CEMS data from any plant in the SO₂ allowance programme. This provides a degree of public accountability that seems very conducive to good policy.

Retaining the basic data within the government and not making it available to the public would, of course, also protect privacy concerns. Although that would preclude completely independent evaluations (such as those performed on the SO₂ allowance programme), it would not prohibit independent evaluations that were approved by the government and therefore entitled to have confidential access to the data (such as those performed in New Zealand and Australia).

In general, it appears that data accessibility could be improved for most of the programmes. Although the confidentiality of proprietary information is a legitimate concern, this should not be a binding constraint for most programmes. The Danish programme, which involves only eight participants in the market, illustrates the type of very small programme that does face the trade-off between accessibility and confidentiality (Pedersen 2003).

VI. Selecting the Institutional Location of Evaluators

The central issue associated with the choice of evaluator is the decision about whether or not to choose an 'internal' or an 'external' evaluator. An internal evaluator is usually someone who is directly within the agency or ministry which is responsible for implementing the regulation, while an external regulator is usually someone who has been contracted (e.g. an academic or a management consultant) to undertake the evaluation. However, the distinction can be one of degrees and not kind. For instance, in many cases there may be a separate evaluation unit within the Ministry which is responsible for the policy to be evaluated (OECD 1999b and OECD 1997a). In ascending order of the degree to which the evaluator is described as being 'internal' to the authority whose regulation is being evaluated, the following possible classifications can be considered:

- organisation evaluated
- special evaluation unit within ministries
- central management agencies
- commissions
- management consultants
- academic or research institutions
- community and consumer groups
- national audit offices

However, institutional location is not the only determinant of the degree of 'autonomy' enjoyed by the evaluator. For instance, in the OECD Public Management Service's classification set out above national audit offices are presumably considered to be even more independent than consultants or academics since no contractual relationship is involved in the former case. However, this will depend upon the political culture within the country. No one-to-one relationship exists between the institutional location of the evaluator and the degree of autonomy which the evaluator enjoys.

There are advantages and disadvantages between choosing internal or external evaluators (OECD 1997a and Smith 2002). The use of internal evaluators runs the risk of

significant bias, sometimes not even recognised by the officials themselves. If government officials have a vested professional interest in the outcome of the evaluation (i.e. continuation of the policy), they are less likely to be critical. One economist who oversaw a review of the US Clean Air Act stated that she had to “strong-arm EPA staff to get them to conduct a sensitivity analysis that generated lower benefit estimates” (Lutter 2003). More subtly, they may have a feeling of ‘ownership’ over the project which makes it difficult for them to distance themselves from the policy in a way which allows for independent assessment (Lutter 2003, CEC DG Budget 1997). Moreover, they may not have the specialised technical capacity needed to undertake methodologically complicated evaluations. For these reasons, external evaluators are often recommended.

On the other hand, external reviewers may be insufficiently familiar with many of the programmatic details of the policy to undertake an effective assessment, perhaps underestimating some of the costs associated with policy implementation. In addition, if data confidentiality is important they may not have access to all of the data required in order to undertake an evaluation of the programme. And finally, the use of external evaluators may restrict the potential for ‘learning by doing’ and downstream policy reform within the regulatory authority.

However, clear advantages are associated with the use of external evaluators. In the United States for example, the external ex post evaluations of the SO₂ allowance programme published by Resources for the Future and MIT have served the public, the academic community and the programme managers well. Since they approached the evaluations from different perspectives each review was helpful in its own right. These independent external evaluations were made possible by the public availability of all the key data about the programme.

These trade-offs are ‘reconciled’ in strikingly different ways in OECD Member countries. For instance, Japan stresses the value of internal understanding of the policy process, greater potential for policy reform, and ‘learning-by-doing’ in supporting the use of internal evaluators (Japan MPHPT 2002). While the European Commission generally recognises the importance of such factors, a stronger preference for the use of external evaluators is evident (CEC DG for Budgets 1997). In France, it is suggested that a combination of internal and external evaluators may be optimal (France Conseil Scientifique de de l’Evaluation 1996). In Ireland, an interesting case arose when Ministry of Finance officials were seconded to the academic institute (the Economic and Social Research Institute) undertaking an evaluation on the efficiency and effectiveness of the use of EU Structural Funds in Ireland (Boyle 1993).

Thus, a key question in ex post evaluations is identifying the most appropriate institutional location of the evaluators. This is as true of tradeable permit programmes as other public policies. The workshop participants identified many existing models. As noted above, evaluations can be conducted internally by the programme managers or by professional evaluators elsewhere in the government, or by independent evaluators either under contract to the government or acting completely on their own.

Does the institutional location of the evaluator also affect the type of evaluation undertaken? Apparently it can. Young (2004) points out that in Australia assessments prepared for governments typically tend to be broader in their focus than those under taken primarily from an academic research perspective. Furthermore, for the evaluation of the Australian water entitlements trading programme, since several States and the Commonwealth were involved, it seemed obvious to administrators and participants that this particular review would need to be independent.

In the case of RECLAIM, the initial audit was undertaken by RECLAIM staff, but it was overseen by an Advisory Committee which included regulated firms, academics, environmental

NGO's, and regulatory agencies. This provided some quality control, a point to which we return below. Moreover, since dissenting opinions were published in the report, this allows for a degree of independence which is important for the credibility of the evaluation (Smith 2002).

Discussions during the workshop revealed the existence of a difference between countries in the reception accorded to evaluations from external sources. The reception given to the outside evaluations in both New Zealand and the United States were extremely welcoming and co-operative. Yet the process was not severely tested because the evaluations were not only generally very positive, but were expected to be so. It is not clear that the reception would have been quite as welcoming or the relationship quite as co-operative if the evaluations were negative.

VII Providing Appropriate Personnel Incentives

The question of the choice of evaluator is closely linked with more general issues associated with the division of responsibilities in the evaluation process, and incentives for personnel in the implementing agency. In particular, one issue requiring further investigation is whether differences in governments' receptivity to outside evaluations is systematically related to their personnel policies. One hypothesis would suggest that bureaucracies where personnel are so vulnerable that negative evaluations could place their careers in jeopardy would be less likely to welcome external evaluations. However, these same incentives would also make it much less likely that an internal evaluation would be unbiased.

For tradeable permit programmes – which remain at a relatively immature stage of development in most OECD countries – these issues are particularly important. Since tradeable permit systems represent a significant change from legalistic direct forms of regulation to more market-oriented economic forms of regulation, they will inevitably have significant implications for personnel within regulatory authorities. In some cases, the implications for agency staffing might result in pressures for emphasising the deficiencies of reforms (Smith 2002). For instance, a strong public sector union might have trouble with programmes that fundamentally change the nature of the bureaucracy, which tradeable permit programmes seem to do.

The issue of personnel management and organisational incentives could, therefore, be a key factor in determining the nature and effectiveness of ex post evaluations. In particular, evaluations could prove to be an effective means of resolving incentive-incompatibility problems within public administrations by ensuring that the incentives of those bureaucrats who are responsible for policy implementation are compatible with the underlying objectives of the policy. In the absence of external evaluations, upper-level management are relying upon information provided by people who have a vested professional interest in making the programmes appear to look good or bad (depending upon the relative importance of the points made above).

Thus, to some extent, the choice of evaluator may be less important than the management structure in which the evaluation is undertaken. In particular, it is important to have a clear delineation of responsibilities within the management of the evaluation process. Thus, Hahn and Litan (1997) have emphasised the need to ensure that responsibilities for regulatory analysis be distinct from rule-writing and regulatory enforcement. In such cases, even if the evaluator is 'internal' there may be a sufficient degree of autonomy and independence. Thus, the trade-offs between internal and external evaluators can be reconciled through an effective division of responsibilities within the regulated authority.

However, in all cases a body with overall oversight of evaluations seems to be key. In a study conducted in the late 1990's the OECD Public Management Service found that 10 out of 17 countries surveyed had dedicated bodies responsible for regulatory oversight (OECD 1997b). In

four cases these were located in industry, economics or commerce ministries, three were in budget or public management ministries, two were in central cabinet offices, and one was in a parliamentary auditing office.

For example, in Canada, the Regulatory Affairs Directorate of the Treasury Board is responsible for oversight of policy evaluations (Apogee Research 1997). In the United States, the Office of Management and Budget's Office of Information and Regulatory Affairs serves this role (Hahn and Litan 1997). In Japan, the Administrative Evaluation Bureau of the Ministry of Public Management has an analogous role (Japan MPPHPT 20002).

Hard and fast rules concerning the optimal means of managing evaluations are not possible. However, it has been emphasised that two extreme positions are unlikely to be efficient (OECD 1997b), namely:

- full delegation to regulatory authorities themselves of responsibility for undertaking, overseeing, and acting upon the results of evaluations; and,
- full centralisation of responsibility for evaluation, without significant input from the agency or ministry whose regulation is being evaluated.

Not surprisingly, most OECD countries find themselves somewhere between these two extremes, although it would appear that the greater danger lies in too much decentralisation rather than too little.

VIII. Resisting Strategic Behaviour

Evaluations can upset vested interests. As such, issues of strategic behaviour and rent-seeking are endemic in the evaluation processes. Indeed, pressure on evaluators (particularly if internal) can be such that protection may be required in order to ensure that they are able to exercise their role effectively. Lutter (2003) has gone so far as to suggest the need to grant 'whistleblower' protection status for those who point out distortions which might have been applied by government officials in support of particular policies.

As noted above, those responsible for the policies that are being evaluated may present significant obstacles to effective evaluation if they have a vested professional interest in the outcome of the evaluation (i.e. continuation of the policy). Even if they are not directly responsible for the evaluation they may be unwilling to make important data and information available. Thus, resistance can often arise within the bureaucratic establishment (Morrall III 1997 and Smith 2002).

More fundamentally, regulators may see that they represent a particular constituency (often the beneficiaries of a policy) and not the community more widely (including those who bear the costs associated with the introduction of the policy). This can result in significant resistance to evaluation. Analogously, Deighton-Smith (1997) has pointed out that regulators may see themselves as representing a particular objective rather than social welfare generally.

The issue of regulatory 'capture' by those outside of the government who are affected by regulations is, of course, also important. This might manifest itself in the evaluation process itself since evaluations are never purely 'positive' exercises. The choice of criteria and methodology implies value judgements. The demand for 'better' regulation might in fact be interpreted as masking demand for 'less' regulation (see Deighton-Smith 1997).

In other cases, the beneficiaries may have significant influence. For instance, if a policy provides subsidies or other types of benefits, evaluations which might jeopardise their continuation will be seen as potentially damaging to their interests. For ex post evaluation, this may be particularly important since particular constituencies will have a stronger sense of 'rights' to benefits actually being received, rather than those being received prospectively in the case of ex ante evaluations (see Deighton-Smith 1997).

On the other hand, effective and credible evaluation can strengthen the government's hand relative to vested interests in their efforts to protect overall social welfare (see Holmes and Argy 1997). By providing reliable and analytical support for the need to undertake policy reforms, governments may be able to deflect some of the pressures from stakeholders with particular interests.

IX. Providing Adequate Quality Control

In order to be credible evaluations need to be seen to be of sufficient technical quality. More specifically, evaluations need to satisfy the following criteria (Boyle 1993):

- internal validity – i.e. are confounding factors adequately addressed, is the baseline plausible, etc.;
- reliability – i.e. is the analysis undertaken using appropriate methodologies, is the data reliable, etc.; and,
- objectivity – i.e. are the evaluation conclusions consistent with the outcomes of the empirical research.

In some cases, 'external validity' (i.e. are the results of the evaluations generalisable to other policy or sectoral contexts) may also be seen as an important attribute. In order to ensure such conditions are met, it is important to put in place mechanisms to oversee evaluation procedures. In this vein, the European Commission emphasises the importance of a steering group to oversee the management of evaluations (CEC DG for Budgets 1997). Interestingly, in the United Kingdom, Heads of Departments must sign 'Regulatory Quality Certificates' indicating that the evaluation was undertaken using satisfactory methodologies and sound data (UK Better Regulation Unit 1997).

Training and guidance can also be important means of ensuring good quality evaluations are undertaken. In Japan the Commission on Policy Evaluation and Evaluation of Individual Administrative Institutions provides guidance to ministries and agencies on undertaking evaluations (Japan MPHPT 2002). A dedicated 'Study Group on Policy Evaluation' was formed to provide specific guidance on evaluation methodologies. In Canada, the Regulatory Affairs Directorate of the Treasury Board is responsible for providing guidance on evaluation methods and management (Apogee 1997).

A large number of OECD countries provide guides for evaluators and managers of evaluations. Hopkins (1997) provides an overview of the contents of 12 such guides from seven different OECD countries. And finally, resources (human and financial) need to be available in order to ensure that evaluations are of a satisfactory quality. The EC's Communication on Evaluation estimates that budgets for evaluations will often represent in the region of 0.5% of total programme costs (CEC DG Budget 1997). However, this varies widely depending upon the nature of the evaluation and the context in which it is applied.

The final check on quality comes from having a transparent process that allows independent assessment of the evaluation. All ex post evaluations involve a host of assumptions and judgements. It is important that these be understood and evaluated by those who have the appropriate expertise and the interest. Lutter (2003) emphasises the importance of independent peer reviews of evaluations.

XI. Disseminating and Communicating Results

The need for effective dissemination is also key to ensuring that policy evaluations have an impact on the policy process. It is not enough to assume that 'good science' will lead to 'good policy'. The results of the evaluation need to be communicated to those responsible for bringing about policy change in a manner which resonates (Hahn and Litan 2002). More generally, in order to have a wide impact, the results should be disseminated widely – a point made by the Japanese Ministry of Public Management (Japan MPHPT 2002). The need for effective dissemination of evaluation results has been explicitly mentioned in the OECD's Council Recommendation on Improving the Quality of Government Regulation (1997b).

The widespread use of the World Wide Web has made the current possibilities for sharing information historically unprecedented. It is now possible to make vast amounts of information available to the public at very low cost. In many senses, tradeable permit schemes have been at the forefront of the effort of regulatory authorities to make use of modern technologies to disseminate information widely and inexpensively (See <http://www.epa.gov/airmarkets/trading/> for an example).

However, the existence of these technological possibilities also means that cost can not be used as an excuse for inadequate dissemination of evaluation findings. In general, the results of evaluations should be shared as widely as possible in as complete a form as possible. This should include not only the complete evaluation report with all supporting analysis, but also more synthetic documents which summarise the methodology and results. Others must be able to assess and comment on what was done in the evaluation. Since peer review has served the process of policy reform well for a long period of time, it should be empowered by the transparency of the evaluation methods and results.

The summary findings and non-technical presentation of the results should be disseminated as widely as possible for the non-specialist audience. While such reports should be intelligible to a wide audience – and particularly those affected directly by a regulation – the reports should be careful not to underplay the uncertainty and complexity of evaluation methodologies (see UK Treasury 2003).

Conceivably many different groups could be seen as entitled to have access to the information, including: programme managers, the broader bureaucracy, legislative committees or individual legislators, researchers in the academic community or independent research institutes and finally the general public. The proper management of ex post evaluations will have to not only identify who should have access to this information, but to whom the analysis should be targeted. These are not necessarily the same – even reports targeted at policy makers, may be made available to the public.

It is also worth noting that the form of the evaluation could have an impact on dissemination. Independent reviews are much more likely to be published in the journals that are regularly read and referenced by the academic community. Academics have a strong incentive to get everything they do into print. If, as generally supposed, academic journals are the most effective means for spreading the insights learned (particularly among foreign scholars), this

means that the choice of evaluation type (internal or external) may influence dissemination as well as content.

XI. Generating Demand for Evaluation

Analysis of the 'supply side' of policy evaluation is relatively well-developed. However, the demand side is perhaps more important. Those responsible for introducing, implementing and reforming public policies need to see the merits of policy evaluation. Perhaps more importantly, they need to face incentives to ensure that evaluations are undertaken and acted upon. Such 'demand' can be created through (OECD 1999b):

- sticks – such as mandatory requirements to undertake evaluations at specified points, or withdrawal of budget allocations if such evaluations are not undertaken;
- carrots – such as earmarking of funds for evaluations, or links with personal performance management systems; and,
- moral suasion – such as workshops and seminars which serve to point out the value of evaluation for efficient public administration.

The key point is that the internal demand for evaluations has to be genuine, and this must be reflected in the incentives faced by those who are responsible for implementing regulations. To some extent this runs counter to the bureaucratic culture which exists in many countries, encouraging those who are responsible for implementing particular regulations to feel 'ownership' over the programme rather than the fundamental policy objective.

If appropriate incentives are not in place, resistance to the evaluation process may arise, and the evaluation findings are unlikely to be implemented. In a sense, the incentives of the regulatory authorities need to be linked as closely as possible with meeting the social objectives of the programme writ large, and not the precise programmatic details of the instrument applied. This is, of course, a significant management challenge.

XII. Conclusions

While providing information and insights on technical issues associated with undertaking evaluations of tradeable permit schemes, the workshop also provided valuable lessons on the links between policy evaluation and policy-making generally. The main lessons in this area included:

- Policy evaluation should be seen as an integral part of performance management within public authorities, and not 'add-ons' which are discrete and isolated assessments of individual policies.
- The timing of evaluations is crucial, since this will determine the 'scope' for policy adjustments and reform. Since programmes typically evolve, mid-course evaluations can be particularly helpful in shaping that evolution. Mid-course evaluations will typically be more targeted and less formal than retrospective ex post evaluations.
- Much of the data that would be needed to begin to perform at least mid-course evaluations of these programmes are collected as a normal part of the monitoring and

enforcement process. When this is not the case, it is important to devote resources to the collection of data which is needed for evaluating a programme.

- Data access is at least as important as data collection. Wide access to the data necessary for undertaking evaluations reduces costs and increases the range of evaluations which the regulatory authorities can draw upon.
- The institutional location of the evaluator is key, with a strong case to be made for the use of external evaluators. However, it is also important to ensure that there is an appropriate division of responsibilities for the management of evaluations, irrespective whether the evaluation is conducted internally or externally.
- In a closely related vein, it is important that those within public agencies who are likely to be affected by the outcomes of policy evaluations face incentives which encourage full co-operation. Similarly, they must also be in a position to resist pressure from lobby groups who have a vested interest in the outcome of evaluations.
- Since a peer review of evaluations is so important, at a minimum complete reports on the evaluation methods and results should be transparent to the public. Where possible all the data used to perform the evaluation of the programme should also be publicly available.

In conclusion, while the specific results generated from individual evaluations are clearly important, some of the most important benefits associated with undertaking policy evaluations relate to the broader process of undertaking and managing evaluations and linking them with public sector reform more generally. Policy evaluation must be seen as an element of efficient public management in a broad sense.

REFERENCES

- APOGEE RESEARCH (1997) 'Regulatory Reform through Regulatory Impact Analysis' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD).
- AUSTRALIA OFFICE OF REGULATION REVIEW (1998) *A Guide to Regulation* (Canberra: ORR).
- BLIGHT, S. (2003) 'Ex Post Evaluation in Canada' personal intervention at the OECD Workshop on the "Ex Post Evaluation of Tradeable Permits", OECD, Paris, January 21-22, 2003.
- BOYLE, Richard (1993) *Making Evaluation Relevant : A Study of Policy and Programme Evaluation Practice in the Irish Public Sector* (Dublin: Institute of Public Administration).
- BRODER, Ivy E. and John F. Morrall III (1997) 'Collecting and Using Data for Regulatory Decision-Making' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- CANADA, Treasury Board of Canada (2001) 'Evaluation Policy' (Ottawa: Treasury Board)
- DEIGHTON-SMITH, REX (1997) 'Regulatory Impact Analysis: Best Practices in OECD Countries' OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- DELACHE, X. (2003) 'Ex Post Evaluation in France' personal intervention at the OECD Workshop on the "Ex Post Evaluation of Tradeable Permits", OECD, Paris, January 21-22, 2003.
- ELLERMAN, A. Denny. (2004) "The U.S. SO₂ Cap-and-Trade Program". Proceedings of the OECD Workshop on Ex Post Evaluation of Tradeable Permits: Methodological and Policy Issues. Paris: OECD (January 21-22) [Chapter 2 in this volume.]
- EUROPEAN COMMISSION, DIRECTORATE-GENERAL XIX BUDGETS (1997) *Evaluating EU Expenditure Programmes: A Guide - Ex Post and Intermediate Evaluations* (Brussels: DG Budgets)
- EUROPEAN COMMISSION, WORKING GROUP ON GUIDELINES, NETWORK OF EVALUATORS OF THE EUROPEAN COMMISSION (2000) 'Good Practice Guidelines for the Management of the Evaluation Function' (Brussels: European Commission).
- FRANCE, CONSEIL SCIENTIFIQUE DE L'EVALUATION (1996) *Petit Guide de l'Evaluation des Politiques Publiques* (Paris : La Documentation Française).
- HAHN, Robert W. and Robert E. Litan (1997) 'Improving Regulatory Accountability' (Washington D.C.: American Enterprise Institute for Public Policy Research and the Brookings Institution)

- HAHN, Robert W. and Robert E. Litan (2002) 'Recommendations for Improving Federal Regulation' (Washington D.C.: AEI-Brookings Joint Center for Regulatory Studies)
- HARRISON, David, Jr. (2004) HARRISON, David, Jr., Ph.D. (2004). "Ex Post Evaluation of the RECLAIM Emissions Trading Programme for the Los Angeles Air Basin". In OECD *Tradeable Permits: Policy Evaluation, Design and Reform*. (Paris: OECD) (Chapter 2 in the present volume).
- HARTRIDGE, O. (2003) 'The UK Emissions Trading Scheme: Progress Report' intervention at the OECD Workshop on the "Ex Post Evaluation of Tradeable Permits", OECD, Paris, January 21-22, 2003.
- HOLMES, Sue and Steven Argy (1997) 'Reviewing Existing Regulations: Australia's National Legislative Review' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- HOPKINS, D. (1997) 'Alternative Approaches to Regulatory Analysis: Designs from Seven OECD Countries' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- JAPAN, MINISTRY OF PUBLIC MANAGEMENT, HOME AFFAIRS, POSTS AND TELECOMMUNICATIONS (2002) 'The Policy Evaluation System of the Government of Japan' (www.soumu.go.jp/english/kansatu/evaluation/evaluation_01.html)
- KERR, S. (2004) "Evaluation of the Cost Effectiveness of the New Zealand Individual Transferable Quota Fisheries Market". Proceedings of the OECD Workshop on Ex Post Evaluation of Tradeable Permits: Methodological and Policy Issues. Paris: OECD (January 21-22) [Chapter 5 in this volume.]
- LUTTER, Randall (2003) 'Agencies Regulatory Analyses Should be Subject to Genuinely Independent Peer-Review' (Washington D.C.: AEI-Brookings Joint Center for Regulatory Studies)
- MCLEAN, B. (2003) 'Ex Post Evaluation in the United States' intervention at the OECD Workshop on the "Ex Post Evaluation of Tradeable Permits", OECD, Paris, January 21-22, 2003.
- MONTERO, J. P., J. M. Sanchez, et al. (2002). "A Market-Based Environmental Policy Experiment in Chile." *Journal of Law and Economics*, Volume 45, No. 1, Part 1, pp. 267-287.
- MORRALL, John F. (1997) 'An Assessment of the US Regulatory Impact Analysis Program' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- NUSSBAUM, B. D. (1992). "Phasing Down Lead in Gasoline in the U.S.: Mandates, Incentives, Trading and Banking." *Climate Change: Designing a Tradeable Permit System*. T. Jones and J. Corfee-Morlot. Paris, Organization for Economic Co-operation and Development Publication: 21-34.
- OECD (1995) 'Recommendation of the Council of the OECD on Improving the Quality of Government Regulation' (OCDE/GD(95)95).
- OECD (1997a) *Evaluating Economic Instruments for Environmental Policy* (Paris: OECD).
- OECD (1997b) *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)

- OECD (1997c) *Reforming Environmental Regulation in OECD Countries* (Paris: OECD).
- OECD (1997d) *The OECD Report on Regulatory Reform* (Paris: OECD).
- OECD (1999a) *Implementing Domestic Tradeable Permits for Environmental Protection* (Paris: OECD).
- OECD (1999b) 'Improving Evaluation Practices: Best Practice Guidelines for Evaluation and Background Paper' (PUMA/PAC(99)1).
- PEDERSEN, S. L. (2003) 'Experience Gained with CO₂ Cap-and-Trade in Denmark' intervention at the OECD Workshop on the "Ex Post Evaluation of Tradeable Permits", OECD, Paris, January 21-22, 2003.
- SHABMAN, Leonard. (2004) "Compensation for the Impacts of Wetland Fill: The US Experience with Credit Sales". Proceedings of the OECD Workshop on Ex Post Evaluation of Tradeable Permits: Methodological and Policy Issues. Paris: OECD (January 21-22) [Chapter 7 in this volume.]
- SMITH, Stephen (2002) 'Ex Post Evaluations of Tradeable Permit Programmes' in OECD *Implementing Domestic Tradeable Permits: Recent Developments and Future Challenges* (Paris: OECD).
- UNITED KINGDOM, HM TREASURY (1997) *'The Green Book': Appraisal and Evaluation in Central Government: Treasury Guidance* (London: The Stationery Office).
- UNITED KINGDOM, HM TREASURY (2003) *'The Green Book': Appraisal and Evaluation in Central Government* (London: The Stationery Office).
- UNITED KINGDOM, THE BETTER REGULATION UNIT (1997) 'Regulatory Compliance Cost Assessment: UK Experience' in OECD *Regulatory Impact Analysis: Best Practices in OECD Countries* (Paris: OECD)
- WOSSINK, Ada. (2004) "The Dutch Nutrient Quota System: Past Experience and Lessons for the Future". Proceedings of the OECD Workshop on Ex Post Evaluation of Tradeable Permits: Methodological and Policy Issues. Paris: OECD (January 21-22) [Chapter 4 in this volume.]
- YOUNG, Michael D. (2004) "Learning from the Market: Ex Post Water Entitlement and Allocation Trading Assessment Experience in Australia". Proceedings of the OECD Workshop on Ex Post Evaluation of Tradeable Permits: Methodological and Policy Issues. Paris: OECD (January 21-22) [Chapter 6 in this volume.]

TABLE OF CONTENTS

Chapter 1. EX POST EVALUATION OF TRADEABLE PERMITS: METHODOLOGICAL ISSUES AND LITERATURE REVIEW

By Tom Tietenberg and Nick Johnstone	9
I. Introduction.....	9
II. Criteria for the Ex Post Evaluation of Tradeable Permits	11
III. Methodological Issues in Ex Post Evaluation.....	14
IV. A Review of Ex Post Evaluations of the Impacts of Tradeable Permit System.....	20
V. Conclusions.....	34
References	36

Chapter 2. EX-POST EVALUATION OF THE RECLAIM EMISSIONS TRADING PROGRAMMES FOR THE LOS ANGELES AIR BASIN

By David Harrison, Jr., PhD	45
I. Introduction.....	45
II. Overview of RECLAIM and Initial Experience	47
III. More Recent Experience with RECLAIM	54
IV. Conclusions Regarding RECLAIM Experience.....	64
References	67

Chapter 3. THE US SO₂ CAP-AND-TRADE PROGRAMME

By A. Denny Ellerman.....	71
I. Introduction	71
II. The Political and Regulatory Context of Title IV	72
III. Institutional Location and Methodology	73
IV. Economic efficiency	75
V. Environmental Effectiveness	83
VI. Dynamic Effects	83
VII. Other Costs and Effects.....	92
VIII. Conclusions and Implications	94
References	95

Chapter 4. THE DUTCH NUTRIENT QUOTA SYSTEM: PAST EXPERIENCE AND LESSONS FOR THE FUTURE

By Ada Wossink 99

I Introduction.. 99

II. Environmental Problems In Animal Agriculture And Initial Policy Regulation.....100

III. Overview Of The Quota System101

IV. Evaluation.....108

V. Overview of Main Results.....111

VI. Conclusion: Implications For Policy Reform.....116

VII. References...118

Chapter 5. EVALUATING THE NEW ZEALAND INDIVIDUAL TRANSFERABLE QUOTA MARKET FOR FISHERIES MANAGEMENT

By Suzi Kerr121

I. Introduction121

II. Overview of the Tradeable Permit Regime.....121

III. Context of Evaluation.....123

IV. Description of Assessment Methodology125

V. Main Results126

VII. Potential Implications For Policy Reform.....131

References133

Chapter 6. LEARNING FROM THE MARKET: EX-POST WATER ACCESS ENTITLEMENT AND ALLOCATION TRADING ASSESSMENT EXPERIENCE IN AUSTRALIA

By Michael D. Young.....135

I. Background136

II. Definitional Overview137

III. The Condition of the Resource137

IV. The Trading Environment137

V. Market Assessments138

VI. Overview of the Evaluation of the Tradeable Permit Regime144

VII. Concluding Comments.....147

Appendix 6.1 Executive Summary Of “Interstate Water Trading: A Two Year Review.....147

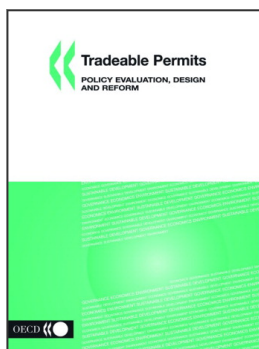
References152

**Chapter 7. COMPENSATING FOR THE IMPACTS OF WETLANDS FILL:
THE US EXPERIENCE WITH CREDIT SALES**

By Leonard Shabman	155
I. Introduction..	157
II. Wetlands Credit Sales: Origins and Current Status.....	160
III. Wetlands Credit Sales in their Regulatory Context.....	160
IV. Toward a Market-like Credits Sales Programme.....	164
V. Conclusion	166
Appendix 7A:The Approach to Programme Evaluation	167
References	169

Chapter 8. THE MANAGEMENT OF EX POST PUBLIC POLICY EVALUATIONS

By Tom Tietenberg and Nick Johnstone	173
I. Introduction.	173
II. Policy Evaluation as Performance Management.....	173
III. Timing of Evaluations to Encourage Policy Reform	174
IV. Ensuring Adequate Data Collection.....	176
V. Providing Access to Data for Evaluators	177
VI. Selecting the Institutional Location of Evaluators.....	179
VII. Providing Appropriate Personnel Incentives	181
VIII. Resisting Strategic Behaviour	182
IX. Providing Adequate Quality Control.....	183
X. Disseminating and Communicating Results	184
XII. Generating Demand for Evaluation	185
XIII. Conclusions	185
References	187



From:
Tradeable Permits
Policy Evaluation, Design and Reform

Access the complete publication at:
<https://doi.org/10.1787/9789264015036-en>

Please cite this chapter as:

Johnstone, Nick and Tom Tietenberg (2004), "The Management of Ex Post Public Policy Evaluations", in OECD, *Tradeable Permits: Policy Evaluation, Design and Reform*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264015036-9-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.