OECD Environmental Performance Review Programme

Accountability (f)or Learning?

MARKKU LEHTONEN
University of Versailles Saint-Quentin-en-Yvelines, France

The growing interest in evaluation as a new form of environmental governance stems from two developments: the movement towards New Public Management, and the search for new policy instruments for managing complexity, uncertainty and plurality of values in the pursuit of sustainable development. The former holds increasing accountability as the main purpose of evaluation, while the latter stresses the importance of learning. These approaches are often considered as mutually exclusive, but recent literature has underlined their complementary roles in policy making. This article examines to what extent the OECD Environmental Performance Review (EPR) programme has succeeded in combining the objectives of learning and accountability within a single evaluation framework. The EPRs have been relatively successful in avoiding the negative side-effects often associated with traditional performance measurement, but have usually failed to generate broad debate. Focusing on enhancing accountability through social learning would probably contribute to overcoming this problem.

KEYWORDS: accountability; environmental performance; evaluation use; learning; peer review

Introduction

The growing interest in evaluation as a new form of governance in environmental policies can be traced back to two main developments in the field of public policy. First, there are the trends towards decentralization and 'internationalization' of public administration; citizens' demands for greater transparency of policy making; the increasing influence of experts in policy and administration; the increasing use of project management even in the basic tasks of administration; the declining legitimacy of public administrations, entailing the public's loss of faith in the government's ability to spend their taxes wisely; and the pressures to reduce public spending. All these trends have strengthened the pressures upon the public sector and non-profit organizations to demonstrate *accountability* for an appropriate conduct of policies to external sponsors and other stakeholders. This has been most clearly manifested in the rise of performance measurement

and the call for policies to be based on evidence of their success, in the framework of the so-called New Public Management. Performance measurement for accountability typically focuses on merely monitoring the results and outcomes of policies, without analysing the causal links between policies and outcomes (e.g. Feller, 2002: 438; Greene, 1999: 162; Pitarelli and Monnier, 2000; Thoenig, 2000; Uusikylä, 1999; van der Knaap, 2000).

Second, the emergence of the concept of sustainable development as the overarching principle for environmental policies, with its emphasis on complexity, interconnectedness, interdisciplinarity, pervasive uncertainty, potential irreversibilities and plurality of values and points of view, has led to a search for new policy instruments, better suited to the management of new kinds of problems. The policy approach based exclusively on 'command-and-control' instruments, which has been relatively successful in reducing the 'simple' point-source pollution problems in most developed countries, has been progressively substituted by a new approach, in which any particular problem is addressed through applying a wide range of instruments – regulatory, economic, and information-based ones – as 'policy-packages'. Stakeholders become closely integrated into the design of policy instruments, and the authorities increasingly adopt a 'facilitating' role instead of acting as mere regulators (e.g. Hajer, 1995; Joas, 2001; Mol and Spaargaren, 2000; Sairinen, 2000; Schubert et al., 2000; Weale, 1992). From this point of view, the wider use of evaluations and the development of indicators can be seen as instruments enhancing the 'reflexivity' of modernization (Giddens, 1990) and 'deliberative democracy' (e.g. Dryzek, 2000; O'Connor et al., 2001/2002) through inclusive, participatory policy making, which should ultimately contribute to sustainability through *learning* (e.g. Baron, 1999; Siebenhühner, 2001; van der Knaap, 1995). This would be achieved through in-depth evaluations that would primarily aim to explain the reasons behind the policy outcomes, so as to improve policies.

For both of these perspectives, the raison d'être of evaluations is their utilization in public policy making. Indeed, without the promise of utilization, there would hardly be a reason for policy makers to conduct and sponsors to finance evaluations. However, most empirical studies have shown that the direct, instrumental use of evaluation results in decision-making is rather an exception than a rule, often limited to the technical, operational level, while various indirect uses, often seen in terms of 'enlightenment', are much more common (Albaek, 1989/1990: 10–12; Hanberger, 2001: 58; Lampinen, 1992: 30–7; Leviton and Hughes, 1981; Muller and Surel, 1998: 110; Pollit, 1998; Weiss, 1980, 1987, 1998, 1999). At least in its simplest version the accountability perspective can be seen as a reincarnation of the traditional, rationalist-technocratic model of policy making, which expects evaluation results to be used by policy makers in a direct, instrumental manner (e.g. Davies, 1999; Greene, 1999).

There is an obvious tension between the two perspectives on the use of evaluation and they are often seen as irreconcilable. However, most authors seem to recognize that both providing accountability and enhancing learning are essential elements in the endeavour to promote 'social betterment' through evaluations. While it is necessary to monitor the activities and strive towards

continuous improvement of policies, one also needs to know whether policies are in fact delivering the expected results. The challenge is therefore not to choose between the two, but to look for complementarity through clearly defining the roles of the two approaches. Most suggestions for achieving such complementarity rely on the idea of practising continuous, quantitative performance measurement for accountability on one hand, and conducting more qualitative, in-depth evaluations on the other (e.g. Bernstein, 1999; Biott and Cook, 2000; Blalock, 1999; Bukkems and de Groot, 2002; Davies, 1999; Greene, 1999; Perrin, 1998, 1999, 2002; Wimbush and Watson, 2000). By contrast, this article analyses the extent to which the OECD Environmental Performance Review programme has succeeded in combining the two approaches within a single evaluation framework.

The information obtained for this article is based on:

- the author's experience as a national delegate in the OECD Working Party on Environmental Performance (WPEP) since May 1996,[1] as a country expert on teams reviewing the environmental performance of Mexico (1997–8) and Russia (1998), and as an OECD consultant on the review of Sweden (2003–4);
- an examination of diverse OECD policy documents, notably those relating to the EPR programme and the organization's work in the area of sustainable development; and
- interviews with the WPEP delegates of Canada, Hungary, Japan, Portugal and Slovakia, as well as with other stakeholders involved in the reviews of the Netherlands and Portugal.

The OECD Environmental Performance Review (EPR) Programme

The Organization for Economic Co-operation and Development (OECD) has carried out systematic reviews of its member countries' environmental policies since 1992. In 2000, all the member countries had been reviewed once, and the review programme entered the so-called second cycle. The Environmental Performance Review (EPR) programme can be seen as a response to the growing emphasis on performance measurement and accountability in the OECD countries during the past decade. The principal aim of the reviews is 'to help Member countries improve their individual and collective performances in environmental management' (OECD, 1997a: 5). The primary goals of the EPR programme are to (1) *help individual governments carry out their own national evaluations* and make progress by establishing baselines, trends, policy commitments, institutional arrangements and routine capabilities; (2) promote a continuous *policy dialogue* among member countries; and (3) stimulate greater *accountability* from member countries' governments towards public opinion within developed countries and beyond. In broad terms, the two first objectives – capacity building, policy improvement and dialogue – can be classified under the term 'learning', while the third one represents accountability, unofficially recognized as the main purpose of EPRs.[2]

Usually, an individual review process lasts about a year and a half. After receiving a formal invitation by the country to carry out a review the OECD Environment Directorate sets up a review team, which always includes three or four experts from other member countries, and often members from other OECD directorates. During the approximately one-week review mission in the country, the review team holds question and answer sessions mostly with national authorities, but also with NGOs, business representatives and researchers. The draft review report is circulated for comments within the OECD secretariat before being sent to the reviewed country and delegates of all OECD countries one month before the peer review meeting. After a consultation process within the reviewed country, the environmental authorities of the reviewed country send to the secretariat their written proposals for changes in the conclusions and recommendations of the review, which are then adopted in a whole-day session of the OECD Working Party on Environmental Performance (WPEP). During the meeting, the reviewed country's delegation – usually led by a State Secretary of the Ministry of the Environment – answers questions asked by the delegates from the other countries. To encourage open and frank debate, no minutes are taken during the meeting. A few months after the meeting, the final, corrected report is published on the responsibility of the OECD Secretary General. A press conference and a public seminar are usually held in the reviewed country, with the participation of a high-level official (often the Director) of the OECD Environment Directorate, and the Minister of the Environment of the reviewed country. Practically all member countries now follow the OECD secretariat's advice to publish a report on the implementation of the review recommendations about two years after the publication of the review.

The reviews focus on three clusters of issues: environmental management (the 'traditional' environmental issues: air, water, waste, nature conservation); integration of environmental concerns into other policy areas, notably economic policies and major economic sectors; and country's performance in international environmental cooperation. Three main types of criteria are used to analyse policy success: (1) the country's own policy objectives; (2) the country's international commitments; (3) the OECD body of commonly agreed policy orientations, principles and recommendations (e.g. the OECD Environment Strategy; Council Acts; the 'polluter pays' principle, 'greening'; Council recommendation on the access of citizens to environmental information). The three main questions concerning the objectives relate to *whether the objective has been achieved*, the *level of ambition* of the objective and *cost-efficiency* in the achievement of the objective. To the extent possible, the reviews make use of various types of indicators to evaluate the progress achieved – notably indicators relying on the OECD's Pressure–State–Response model (OECD, 2001a, 2001b). Despite this clear emphasis on concrete policy outcomes, the reports are far from mere compendia of indicators, since not all results can be quantified and even those that can need to be interpreted in their context. Hence, the reviews pay attention to the trends in the achievement of policy objectives and in the policy approaches (e.g. from purely curative to preventive and integrated approaches). Finally, EPRs attempt to evaluate whether the country relies in its policies on a

broad panoply of instruments suited to the particular context (OECD, 1997a: 5–6; Pagani, 2002).

Evaluation Influence

Apart from a few exceptions (see Chelimsky, 1997, 1998; Eggers and Chelimsky, 1999) utilization in policy making is often seen as the main, if not the only, rationale for carrying out policy evaluations. With the discovery that evaluations are rarely used directly by policy makers, but that they usually influence policy making in more subtle, indirect ways, the focus shifted first towards different types of conceptual use, 'enlightenment' (Weiss, 1980, 1987, 1998, 1999), learning,[3] dialogue,[4] argumentation processes (Valovirta, 2002), and the 'process use' (Forss et al., 2002; Patton, 1998) of evaluations, emphasizing the discursive and cognitive aspects of policy making (e.g. Haas, 1992; Hajer, 1995; Litfin, 1994; Majone, 1989; Radaelli, 1995; Sabatier, 1987; Weale, 1992). Such developments have paralleled the emphasis on deliberative democracy, which has been identified also as an essential element of sustainable development (e.g. Dryzek, 2000; Söderbaum, 2000, 2001). Recently, it has been suggested that the term 'utilization' be replaced by a more neutral notion of 'evaluation influence', which would enable an analysis of not only the intended use, but also the unintended – positive and negative – consequences of evaluation (Henry and Mark, 2003; Kirkhart, 2000; Morabito, 2002). This article takes evaluation influence as a basis of analysis.

Evaluation can produce *outcomes* of three kinds. First, the effects may concern *decisions and actions* – corresponding to direct, instrumental use of evaluations, leading to 'policy change', that is, 'major or minor, more incremental modifications of policies' (Sabatier and Jenkins-Smith, 1993). Evaluation findings or recommendations can therefore be (1) adopted as such; (2) rejected after a discussion; or (3) rejected without discussion (Valovirta, 2002). Secondly, evaluation can *create new shared understandings* through the interactive process of argumentation and dialogue, called 'policy-oriented learning', i.e. 'relatively enduring alterations of thought or behavioural intentions that result from experience and/or new information and that are concerned with the attainment or revision of policy objectives' (Sabatier and Jenkins-Smith, 1999: 123). Thirdly, evaluation may *increase or decrease the legitimacy* of past and future actions, actors or organizations – an essential element of accountability. The following will focus on the two latter types of effects, which are essential from the point of view of learning and accountability.

Learning from Evaluations

Various definitions of learning in the context of evaluation have been proposed. The following definition is here taken as a basis:

> . . . the learning processes of an individual or collective entity are characterised by a complexification of its mental models resulting from the acquisition, creation or an internal transfer of knowledge and are manifested by an increasing capacity of the entity to respond to perturbations in its environment. (Baron, 1999: 321)[5]

Different classifications of learning from evaluations have been suggested, usually distinguishing between relatively simple learning effects corresponding to direct use of evaluation results in decision-making and more profound changes in thinking patterns or even paradigms and interpretative frameworks. Van der Knaap (1995) has classified the effects as 'cybernetic control', 'cognitive development' and 'social learning'. *Cybernetic control* represents direct use by managers and policy makers, while *cognitive development* is close to the enlightenment use, evaluations providing stimuli to a cyclical learning process. Finally, *social learning* is based upon communication, evaluations providing strong and empirically founded arguments in an ongoing political discourse. Similar categorizations have been proposed by Argyris and Schön (1996) and Marcussen (2001: 6).

Learning at the *individual level* can be seen as a necessary, yet not alone sufficient condition for collective or organizational learning (Baron, 1999: 211–16; Siebenhühner, 2001). Individuals learn as they interpret information and experiences acquired throughout the evaluation process, resulting in confirmation of existing beliefs, rejection of evaluation findings, or new, transformed beliefs and capacities (Valovirta, 2002). Participants in an *evaluation process* may also acquire new skills in evaluation methodologies or cooperation, build new professional networks, or simply become more motivated to pursue their work (Forss et al., 2002; Henry and Mark, 2003).

However, it is collective learning that here constitutes the main focus of interest, because, in the end, the main purpose of the OECD EPRs, as well as that of most policy evaluations, is to bring about change at the policy level. The major vector of change at this level is the process whereby the individual interpretations enter into a collective process of deliberation, argumentation and dialogue. Argumentation and dialogue can be seen as the medium through which learning – especially collective learning – always takes place. Indeed, Valovirta (2002: 77) considers that 'perhaps the most important value of evaluation is that it incites argumentation and can direct it towards a more reasoned debate'. Evaluative information cannot be seen as indisputable knowledge, but instead as a collection of arguments, which can be debated, accepted and disputed. In a strictly logical sense, judgements and recommendations cannot be true but only more or less convincing for particular audiences. Evaluation thus provides justification and support for the existence of a problem or a course of action (Majone, 1989). Evaluations become utilized in a social context, in which arguments are presented, supported, defended and confronted, some of the arguments arising from an evaluation, others stemming from other sources.

The argumentation process inevitably leads different parties to improve information on each other's reasoning and sharpen their own arguments. Therefore, even if an evaluation seemingly leaves no marks on policy making – in some cases evaluation may produce a deadlock rather than a synthesis or a winning coalition – it may have tangible impacts on policy making in other areas, other circumstances, etc. (Billig, 1987: 106; Valovirta, 2002: 67).

The argumentation processes described mostly perceive *dialogue as an instrument of rational exchange of information*, aiming at consensus, dispute resolution, persuasion, justification, etc. By contrast, the 'substantive' view sees *dialogue as*

open conversation, aimed to enhance the mutual understanding among participants and to define actions for the future (Abma, 1998: 436–7; Abma et al., 2001: 166; Schwandt, 2001: 268–9). The latter perspective stresses the importance of socialization – the internalization of norms as a result of interaction with other people – as a pathway to learning. Another source of learning from 'substantive' dialogue is imitation, which can be triggered by dissatisfaction or the discovery of new possibilities (Marcussen, 2001). Many of the benefits of the 'process use' of evaluation stem from such conversational processes of dialogue.

Finally*, accumulation of evaluation findings* has been emphasized by the theory-based and 'realistic' evaluation approaches as an important source of learning. Accumulation is to be achieved through theory development and improvement of the theories concerning the ways in which policies exert their influence, not by carrying out a series of studies with reliable, replicable and universal applicability (Pawson and Tilley, 1997: 116).

Accountability

In ordinary language, accountability means that somebody is responsible for something and liable to account to someone else (Tamminen et al., 1999: 33). The term therefore defines neither the object of accountability, nor the body to which the liable person or organization should be accountable. In evaluation practice, accountability has customarily become associated with the judgement of whether a programme or a policy has achieved its objectives. First, this information can benefit sponsors, policy makers and other overseeing authorities in their efforts to control whether the subordinate entities use their resources appropriately and efficiently. Second, it provides the public with an opportunity to control the use of 'their' tax money by the authorities (e.g. Vedung, 1991: 74–81).

Accountability is intimately linked with *legitimacy* of policies and actors, which can be seen as an indispensable resource for any actor in that it ensures support from insiders, reduces criticism by detractors, and thus provides the basis for any political action. Legitimacy is needed also in cases in which a decision has already been made, to justify the adopted line of action (Valovirta, 2002: 67–8). Legitimacy can be internal or external. An example of the former is that 'boosting moral': participation in the evaluation process may improve the motivation of the actors by strengthening their commitment to the project and motivation to pursue their activities (Forss et al., 2002). Another example is a situation in which the mere fact that a policy is being evaluated forces the policy personnel to clarify for themselves the objectives of the policy and re-examine the appropriateness of the measures in relation to the objectives. Actors are thus stimulated to elaborate on the policy issues, which leads to change in attitudes or salience of issues at the individual level, thus strengthening the project (Forss et al., 2002). External legitimacy refers to the legitimacy of a policy, an actor or an organization in the eyes of the outsiders: a ministry or a government policy may gain more legitimacy through positive evaluation findings, and lose legitimacy in the opposite case.

Over the past decade, performance measurement, which focuses on policy outcomes and effectiveness instead of implementation processes, has become the

preferred way of providing for accountability in public sector management. Performance measurement is expected to improve public policy making mainly in the following ways (Feller, 2002: 438; Greene, 1999: 162; Martin and Sanderson, 1999: 247):

- By freeing agencies and implementing authorities from their preoccupation of complying with rigid regulations, these can redirect their energies away from following rules and toward creatively solving problems in the field.
- A focus on outcomes provides the opportunity for collective, shared deliberation about what constitutes valued outcomes from a given endeavour.
- Emphasis on outcomes can illuminate whether investments are adequate to achieve expected results, and thus reveal situations in which implementation personnel are asked to accomplish tasks with clearly inadequate tools and resources.
- Performance measurement provides a readily comprehensible way in which an organization can communicate its goals and achievements to external publics.

However, performance measurement has been criticized for engendering a number of *harmful side-effects*. These side-effects can be classified into two main groups: the technical issues related to the design of performance indicators, their validity and relevance; and the 'people issues', such as trust, ownership, organizational politics, etc. (Davies, 1999: 155–7).

The technical problems relate mostly to the question of whether the indicators measure what they were intended to measure. For instance, causality between activities and outcomes is often taken for granted by the users of performance information, even if such causality has not been demonstrated. Programme personnel may thus be held accountable for outcomes which they cannot control. 'Goal displacement' results from an exclusive focus in performance measurement on measuring quantifiable results, which may divert attention from the actual outcomes to outputs and results. Managers may, for instance, focus on 'easy' cases in order to ensure good results measured through quantitative indicators. The level of ambition of goals and targets may be reduced in order to ensure compliance. Indicators are often chosen on the basis of the availability of selected forms of data rather than on the relevance of the data for policies. Unintended consequences and the continuous changes of objectives are difficult to integrate in quantitative performance measurement. Since performance measurement does not explain the reasons behind the observed policy outcomes and outputs, it is often of little help in decision-making and resource allocation. Moreover, no matter how useful it may be to track past performance and monitor dimensions of current activities, it is of limited value for prospective decisions (Blalock, 1999: 118–19; Davies, 1999: 154–6; Feller, 2002: 439–49; Greene, 1999; Leeuw, 2002; Perrin, 1998, 2002; Taylor Fitz-Gibbon, 2002: 140).

From the perspective of learning, the structural or 'people issues' are of most interest. Accountability-based performance systems are said to rely on an outdated top–down command-and-control approach to management, which has

been proven ineffective, and tends to lead to organizational paralysis, 'tunnel vision' and suboptimization, diverting funds from policy development to monitoring and control. The control-based approach tends to discourage responsibility, as well as engender resistance and risk aversion instead of fostering innovation, creativity and motivating people to do their best. Paradoxically, this may result in a reduced focus on outcomes, as emphasis is laid on justification instead of improvement. To meet the performance objectives and prove compliance, entities may be tempted to dissimulate and distort data, or even lie and cheat. Crucial from the perspective of sustainable development is the alleged incompatibility of performance measures with deliberative democracy, stemming from their authoritative, closed way of representing the quality of programmes and policies, which ignores the plurality of values and points of view. The use of administrative rhetoric associated with performance management is considered incompatible with the institutional structures in areas with a non-administrative tradition (e.g. science and technology policy). Finally, performance measurement tends to legitimize and reinforce the prevailing power structures. For instance, the use of citation measures as an indicator in science tends to reinforce the existing perceptions of a hierarchy of scientific journals.

Peer Review as a Mechanism for Learning and Accountability

Within international organizations, 'peer review' mechanisms can be seen to fulfil largely the same functions as performance measurement at the level of national governments. In this context, peer review can be defined as 'the systematic examination and assessment of the performance of a State by other States, with the ultimate goal of helping the reviewed State improve its policy making, adopt best practices, and comply with established standards and principles' (Pagani, 2002: 4). The examination should be *non-adversarial* and rely on mutual *trust* among the states involved, as well as on their shared *confidence* in the process. With these elements in place, peer review is assumed to create a system of mutual *accountability* (Pagani, 2002: 5–6).

The key concept and mechanism through which peer reviews are expected to exert their influence is '*peer pressure*' – a mechanism of 'soft persuasion' or coercion. Public scrutiny, dialogue with peer countries, comparisons and in some cases even ranking of countries exert pressure on domestic public opinion, national administrations and policy makers. Media involvement and the subsequent public scrutiny are essential elements in enhancing the effectiveness of peer pressure (Pagani, 2002). A country seldom wants to be seen in an unfavourable light among its peers, and therefore peer pressure may be a powerful tool in promoting compliance, notably in the 'laggard' states (Beyeler, 2002; Marcussen, 2001: 8; Strang and Chang, 1993).

In the light of this discussion, peer reviews can be seen as a mechanism attempting to combine the functions of learning and accountability within a single evaluation framework. The following brief analysis will look at the ways in which the OECD Environmental Performance Review programme enhances

both learning and accountability, identify the synergies and contradictions between the functions of learning and accountability, and point out some of the main problems and limitations of the OECD approach.

Learning from EPRs

Since most of the information used by the OECD review team is provided by the participating actors themselves, notably by the Ministry of the Environment, the reviews do not reveal significant new findings and therefore do not considerably change existing beliefs held by individual actors. None of the persons interviewed could identify any such 'novelties'. It is more common that review conclusions and recommendations serve to confirm existing beliefs or, sometimes, are rejected.

On the contrary, *participation in the evaluation process* was identified as a major source of benefits from EPRs. For the national delegates, participation in the EPR process – both as national delegates in the peer review meetings and as country experts on the review teams – can represent an important learning opportunity. However, rather than citing specifically EPRs, the persons interviewed referred to the work carried out by the OECD in general as a source of new 'reference frameworks'. Participation in OECD work helps the delegates build professional networks, develop a common language, identify issues of common concern, start using the same kind of causal reasoning when discussing these issues, learn from 'best practices' and, at a deeper level, even develop a common frame of reference and a common worldview.

While it is difficult to point out concrete situations in which the OECD language and ways of thinking would have been transmitted to the reviewed country, the interviews seem to corroborate Marcussen's (2001) finding that, in a longer perspective, such 'socialization' and 'imitation' processes have, indeed, taken place. The EPRs are one of the many elements promoting the adoption of the 'building blocks' of the OECD environmental policy doctrine, such as a more integrative policy approach, greater use of economic instruments in environmental policies, and, in particular, a results-oriented policy style. Here, one should not underestimate the importance of the informal discussions at the margins of the peer review meetings – over lunch, during coffee breaks or receptions organized by the reviewed country following the formal meeting – as occasions of 'dialogue as open conversation' (Abma, 1998: 436–7; Abma et al., 2001: 166). To what extent this learning at the individual level translates into collective learning within the national administrations largely depends on the role and status of the delegate in the country's policy networks and on his/her willingness and capacity to bring forward the lessons learned. For instance, changes in the delegate's position in the administrative hierarchy have in some cases greatly affected the potential of passing on the lessons in the reviewed country.

Somewhat paradoxically, it may well be the members of the *review team* themselves that learn the most through the dialogue during the review mission, since they are obviously less acquainted with the details of the reviewed country's policy. This could provide a potential foundation for learning through the accumulation

of review findings (see e.g. Pawson and Tilley, 1997). More use could be made of this potential, but this would require either more resources for the secretariat or a change of priorities and reallocation of resources so as to free resources for drawing lessons from the different reviews.

In the course of the review, the reviewed country officials from different sectors and levels of administration must find at least a temporary consensus concerning the issues on the review agenda. This process of enhancing inter-sectoral cooperation and 'development of professional networks' has been particularly valuable in countries whose environmental administration is still in its 'formative' phases (e.g. Mexico, Central and East European countries) or whose political culture has not traditionally favoured cross-sectoral interaction. Likewise, the review process has proven useful in helping persons that have recently entered the administration to establish contacts across sectors.

The *review report* is usually released in a press conference and a seminar, and the report is often sent for debate or for information to the National Commission for Sustainable Development, or to Parliament. In these discussions, the environmental administration usually pushes for the implementation of the recommendations.

These argumentation processes in which different parties put forward their views, trying to persuade other actors, 'delegitimize' the findings or defend their actions against critique, constitute the main forum through which the evaluation reports produce their impacts, through 'social learning'. In theory, this is where the goals of accountability and learning come together and have the potential of simultaneously promoting transparency, accountability and deliberative democracy through open, public debate and 'peer pressure'. In practice, EPRs usually generate rather weak interest, and the seminar in which results are discussed seldom stimulates substantial debate. Newspapers may even feature the review on their front page the day following the release event, but the debate usually seems to die out soon thereafter.

EPRs and Accountability

The fundamental strategic aim of EPRs and the main means through which they attempt to provide accountability is to enhance the *legitimacy* of environmental policies in the eyes of the public, other sectors of government and other stakeholders. By demonstrating the achievements and failures of the government in terms of its own policy objectives, international commitments and OECD policy principles, the reviews provide arguments in favour of stronger environmental policies. An aspect mentioned by virtually all of the interviewed persons was the importance of the EPR recommendations in giving more weight to critical environmental policy issues within non-environmental sectors of administration, notably the Ministries of Finance, Trade and Industry. This could hardly be achieved without the prestige and credibility that the OECD as an economic organization enjoys in these circles. The value of EPRs as an instrument for the environmental NGOs in their attempts to put pressure on the administration seems much less important. This is partly because of the 'soft' tone of the

recommendations, but also because of the difficulties of access to the EPR reports – the reports are expensive and not well publicized among most of the NGOs.

In a sense, EPRs can be considered as an example of empowerment evaluation (e.g. Fetterman, 1994, 2000; Fetterman et al., 1996; Weiss, 1998: 27), given their hidden agenda of 'empowering' the environmental authorities of the reviewed country. As in any empowerment evaluation, EPRs face the dilemma between advocacy – legitimacy in the eyes of the primary clients – and credibility in the eyes of other stakeholders. The goal of legitimizing environmental policies also raises the issue of evaluator ethics. To what extent should EPRs adopt the role of policy advocacy, given that most other, 'stronger' policy sectors already do so? In fact, one can argue that failing to advocate, and adopting a purely 'neutral' position would run against the ideals of deliberative democracy, since it would tend to hide the subtle use of power by powerful groups and strengthen the status quo (e.g. Saarikoski, 2000).

Accountability through EPRs mostly takes the form of 'informal' coercion, since poor performance in EPRs warns the individual member countries about the risk of being excluded from the 'good company' of other OECD countries. The interviews confirm the experiences from other fields of OECD activities (Beyeler, 2002; Marcussen, 2001: 8; Strang and Chang, 1993) showing that these indirect coercive learning mechanisms are most effective with regard to the 'laggard' countries, while countries with a well-developed policy system in the evaluated field often find reviews of relatively little value. One of the more extreme cases is that of the Netherlands, in which none of the interviewed persons consider accountability a function worth pursuing in EPRs, many of them arguing that the country's own systems of follow-up and evaluation of policies already provide sufficient accountability. Moreover, such coercive effects are weakened by the fact that the recommendations are expressed in a rather 'diplomatic' language, and are too general to be easily followed up. In reality, probably no country genuinely risks being excluded from the 'good company'.

EPRs and the Negative Side-Effects of Performance Measurement

So EPRs have the potential to combine both learning and accountability functions by involving key stakeholders in an active debate backed by an international organization. Before jumping to the conclusion that EPRs do, indeed, successfully integrate the two functions within a single evaluation framework, two questions need to be answered: first, have EPRs succeeded in avoiding the negative consequences of focus on accountability; and, second, have the potential positive effects been realized in practice?

The problems of performance indicators related to the inadequate design and technical complexities are greatly attenuated by three factors. First, the indicator model employed in EPRs does not assume a direct cause–effect relationship between the indicators of policy responses and outcomes (pressures on the environment or environmental quality). Second, indicators constitute only one source of evaluative information in EPRs, while the main body of evaluation

consists more of qualitative assessment and contextual information obtained through multiple sources, including question and answer sessions with stakeholders. Finally, an EPR is carried out only every seven or eight years in a country, and therefore 'goal displacement' is less likely than when an ongoing performance measurement system is in place.

The EPRs display three major shortcomings generally attributed to performance measurement systems. First, their policy relevance and capacity to engender learning within the environmental administrations are compromised by the fact that they pay scarce attention to the 'why' and 'how' questions, i.e. the causal relationships between policies and outcomes. However, learning from EPRs is not built primarily on increased understanding concerning the underlying causal mechanisms, but rather on peer pressure, and the processes of argumentation and dialogue triggered by an evaluation carried out by a prestigious international organization. Moreover, EPRs apply a slightly extended notion of performance, including not only outcome measurement but also the performance of procedures (e.g. citizen participation, use of a range of policy instruments, 'polluter pays' principle). Also accountability is interpreted more broadly, encompassing the commitment by the reviewed country to continuous evaluation and dialogue.[6]

Second, EPRs are still often perceived as external control, engendering resistance and defensive attitudes on the part of the national authorities. This is the case despite the considerable efforts by the OECD to ensure mutual trust and confidence in the process through obtaining the personal involvement and commitment of high-level policy makers (i.e. the Minister of the Environment), and consensus among participants on the performance measurement methods: indicators and the EPR guidelines are jointly agreed upon by relevant OECD committees. The increased risk of cheating and dissimulation created by greater openness is not the most serious problem, since any outright distortion of information is likely to be discovered, given the transparency of the process. More problematic is the capacity of the civil servants in the reviewed countries to conceal sensitive issues or present them in a manner that hides more than it reveals, and thus makes it more difficult for the review team members to ask the most relevant questions.

The third problem with the EPR peer review approach is perhaps the most fundamental one. As seen above, performance measurement has been claimed to be incompatible with deliberative democracy, because of the closed, authoritative manner of representation, which does not encourage dialogue and deliberation. Indeed, the OECD reviews adopt a top–down perspective, and present the issues in a highly standardized manner, leaving little room for alternative ways of representation adapted to the needs of local-level actors. Even though this diminishes the potential for learning through intimate interaction and dialogue, one may argue that the real potential of EPRs lies elsewhere, in their capacity to enhance learning through public argumentation and debate. However, the crux of the matter is that, as has been noted, in most cases EPRs seem to fail to engender policy debate in the reviewed country, which is a precondition for both learning and accountability.

Why don't EPRs Generate Debate?

Rather than attributing the failure of EPRs to engender genuine interest and debate solely to the authoritative way of representing, alternative explanations can be proposed. First, the review reports are often considered too cumbersome to read: they are filled with numbers and employ rather dry, impersonal and technical language. Second, in an attempt to build confidence and ensure a spirit of trust, the tone in the reports, in the recommendations in particular, is relatively 'soft', consensual, diplomatic and general. Many actors consider this to be a major reason for the failure of EPRs to stimulate meaningful debate in the reviewed country. Others have pointed out that stronger wording would risk diminishing the legitimacy of EPRs in the eyes of their main clients, the national governments. This also might undermine support that these reports rely on in the longer term.

Third, there seem to be notable problems of dissemination: EPRs are little known beyond the administrative circles or even beyond environmental authorities. This results partly from the OECD publication policy, which prevents the review reports from being available free of charge or free copies being widely distributed in the reviewed country. Partly, the reviewed country officials are to blame, since they could make stronger efforts to distribute reports among stakeholders. The OECD press releases at the occasion when the review report is published have likewise been criticized for failing to clearly bring forward the main messages and issues of potential interest to the media.

A fourth major defect identified in EPRs is their low frequency – once in every seven or eight years – which reduces their capacity to stimulate learning.[7] The likely numerous changes of government and of the administrative personnel mean that the 'momentum' created by the review is lost. However, shortening the review cycle would require additional resources, unlikely in the OECD's present budget situation, or significant reorganization and reprioritization of work in the secretariat.[8] This brings us back to the issue of the OECD's role and current 'identity crisis', and invites one to place EPRs within the context of the whole of the OECD activities. A major reason for the reduction in the influence of the EPR is the decline in the organization's perceived role and prestige. In Japan, the OECD was very useful in the 1970s and early 1980s, when the country still had little contact with Europe and the US. Likewise, the OECD was perhaps the only international organization that accepted Portugal as a member during the period of dictatorship, and therefore enjoyed a high political status in the country. For Finland, membership in the EEC would have been politically inconceivable during the cold war, whereas the OECD membership (since 1969) provided the country with a 'window to the West'. Now, Portugal is no longer a dictatorship, the cold war has ended, Japan has several pathways for contacts with the outside world and EU integration has become the highest political priority in most European countries – all these reasons have contributed to reducing the OECD's political weight. However, the importance of the OECD at the expert and technical level is still considered high, and the OECD reviewing team experts are, with few exceptions, highly appreciated.

While the OECD no longer seems to play a crucial role in most of the old member countries, the recently joined and the likely future member countries

constitute a potential clientele for EPRs. However, as the organization's membership becomes increasingly heterogeneous and the main beneficiaries of EPRs would be those with relatively less developed environmental policy institutions, the more advanced countries – usually the main contributors to the OECD budget – would perceive EPRs as giving little 'value-added'. Moreover, one must bear in mind the double role of EPRs: on one hand, they are a mechanism for the OECD – as a relatively independent policy actor, but also as an instrument for its most powerful member countries – to influence policy making in its member countries, disseminate a certain perspective on sustainable development and create a certain image of the organization itself. On the other hand, EPRs' policy influence crucially depends on the image of the OECD as a credible, independent actor. If the OECD – and the EPR programme – became perceived as entirely manipulated by the powerful countries and particular interest groups, the potential of EPRs to promote learning or provide accountability would suffer.

Conclusions

Most of the learning and accountability effects stemming from an evaluation pass through the processes of argumentation and dialogue. Dialogue as an instrument of rational exchange of information is particularly important for accountability and for 'social learning', whereas dialogue as open conversation is crucial for cognitive development and process use. A sufficient public forum for dialogue and argumentation is indispensable for any evaluation (Valovirta, 2002). However, the OECD experience illustrates the formidable difficulties involved in attempts to combine the partly contradictory functions of learning and accountability in a single evaluation framework. The existence of a forum for debate is not enough if the evaluation itself fails to provide sufficient stimulus and 'raw material' for debate. Indeed, many of the shortcomings of the OECD Environmental Performance Reviews in this perspective seem to stem from their attempt to serve too many audiences and fulfil too many functions at a time.

While it is certainly possible to promote learning and accountability at the same time, and while the OECD EPRs have succeeded in avoiding many of the most serious pitfalls related to performance measurement, the OECD experience highlights the importance of clearly identifying one primary function and main client for each evaluation exercise. In the context of the OECD EPRs, and given the basic concept of 'peer pressure' underlying the reviews, it would be most logical to focus on providing accountability through 'social learning'; in the end, the reviews are designed to put pressure on governments, and the main source of such pressure is open dialogue and argumentation. The empowerment aspect of the peer reviews does not change this conclusion: empowerment can take place through learning – helping the empowered entity to develop its policies – or through accountability – by increasing the legitimacy of policies. In order to avoid conflicts among the OECD countries, each with their own needs and priorities, another option could be differentiating the reviews so that accountability is emphasized in some reviews and learning in others. The downside of this option

would be the weakening of comparability across countries. Whatever the option chosen, despite its limitations the concept of peer pressure is worth conserving as a unique evaluation approach that carries great potential in terms of simultaneously promoting both learning and accountability notably in international contexts.

Acknowledgements

Notes

1. This work also included participation in workshops preparing the 'second cycle' of EPRs, and a survey concerning the experiences from the 'first cycle' of EPRs, carried out by the author of the present paper together with the OECD secretariat (OECD, 1997b).
2. Personal communication with a senior staff member from the OECD Environmental Performance and Information Division.
3. See e.g. the special issue on learning, *Evaluation* 6(3) (2000); Baron (1999).
4. See e.g. the special issue on dialogue, *Evaluation* 7(2) (2001).
5. 'Les processus d'apprentissage d'une entité (individuelle ou collective) se caractérisent par une complexification de ses modèles mentaux issue d'une acquisition, d'une création ou d'un transfert interne de connaissances et se concrétisent par un accroissement de ses capacités de réponse aux perturbations de son environnement.'
6. These are among the proposed measures to achieve a proper complementarity between evaluation and performance measurement (e.g. Blalock, 1999; Davies, 1999; Feller, 2002: 450; Owen, 2002; Perrin, 1998: 376).
7. E.g. Bukkems and de Groot (2002) have emphasized that the learning effects of performance audits greatly depend on their regularity.
8. Of course, the usefulness of increasing the frequency of the reviews depends on the indicators and criteria employed to assess performance. Carrying out reviews every one or two years would not be a wise use of resources if the reviews focus on indicators of the state of the environment, which change relatively slowly. Therefore, a rather broad consensus exists among WPEP delegates about a desirable review frequency of 4–5 years.

References

Abma, T. A. (1998) 'Text in an Evaluative Context: Writing for Dialogue', *Evaluation* 4(4): 434–54.
Abma, T. A., J. F. Greene, O. Karlsson, K. Ryan, T. A. Schwandt and G. A. M. Widdershoven (2001) 'Dialogue on Dialogue', *Evaluation* 7(2): 164–80.
Albaek, E. (1989/1990) 'Policy Evaluation: Design and Utilization', in R. C. Rist (ed.) *Policy Evaluation: Linking Theory to Practice.* Cambridge: Edward Elgar Publishing Co.
Argyris, C. and D. Schön (1996) *Organizational Learning II: Theory, Method, and Practice.* Reading, MA: Addison-Wesley.

Baron, G. (1999) 'Evaluation, participation, apprentissage: Une conception de l'action public avec rationalité limité. Analyse et étude de cas d'une évaluation d'un programme européen.' Doctoral dissertation, Université de Rennes I.

Bernstein, D. J. (1999) 'Comments on Perrin's "Effective Use and Misuse of Performance Measurement"', *American Journal of Evaluation* 20(1): 85–93.

Beyeler, M. (2002) 'OECD Surveillance and Welfare States in Western Europe: Concepts and Methods of a Comparative Research Project', paper presented at COST A15 Second Conference 'Welfare Reforms for the 21st Century', 5–6 April, Oslo.

Billig, M. (1987) *Arguing and Thinking: A Rhetorical Approach to Social Psychology*. Cambridge: Cambridge University Press.

Biott, C. and T. Cook (2000) 'Local Evaluation in a National Early Years Excellence Centres Pilot Programme: Integrating Performance Management and Participatory Evaluation', *Evaluation* 6(4): 399–413.

Blalock, A. B. (1999) 'Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration?', *Evaluation* 5(2): 117–49.

Bukkems, G. and H. de Groot (2002) 'Learning by Auditing: Mission Impossible?', Paper presented at the Conference of the European Evaluation Society 'Three Movements in Contemporary Evaluation: Learning, Theory and Evidence', 10–12 Oct., Seville.

Chelimsky, E. (1997) 'The Coming Transformations in Evaluation', in E. Chelimsky and W. R. Shadish (eds) *Evaluation for the 21st Century*, pp. 1–26. Thousand Oaks, CA: SAGE.

Chelimsky, E. (1998) 'The Role of Experience in Formulating Theories of Evaluation Practice', *American Journal of Evaluation* 19(1): 35–55.

Davies, I. C. (1999) 'Evaluation and Performance Management in Government', *Evaluation* 5(2): 150–9.

Dryzek, J. S. (2000) *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. New York: Oxford University Press.

Eggers, H. and E. Chelimsky (1999) 'Purposes and Use: What can we Expect?', *Evaluation* 5(1): 92–6.

Feller, I. (2002) 'Performance Measurement Redux', *American Journal of Evaluation* 23(4): 435–52.

Fetterman, D. M. (1994) 'Empowerment Evaluation', *Evaluation Practice* 15(1): 1–15.

Fetterman, D. M. (2000) *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: SAGE.

Fetterman, D. M., S. J. Kaftarian and A. Wandersman, eds (1996) *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. Thousand Oaks, CA: SAGE.

Forss, K., C. C. Reubien and J. Carlsson (2002) 'Process Use of Evaluations: Types of Use that Precede Lessons Learned and Feedback', *Evaluation* 8(1): 29–45.

Giddens, A. (1990) *The Consequences of Modernity*. Cambridge: Polity Press.

Greene, J. (1999) 'The Inequality of Performance Measurements', *Evaluation* 5(2): 160–72.

Haas, P. M. (1992) 'Introduction: Epistemic Communities and International Policy Co-Ordination', *International Organisation* 46(1): 1–35.

Hajer, M. A. (1995) *The Politics of Environmental Discourse: Ecological Modernisation and the Policy Process*. Oxford: Clarendon Press.

Hanberger, A. (2001) 'What is the Policy Problem? Methodological Challenges in Policy Evaluation', *Evaluation* 7(1): 45–62.

Henry, G. T. and M. M. Mark (2003) 'Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions', *American Journal of Evaluation* 24(3): 293–314.

Joas, M. (2001) *Reflexive Modernisation of the Environmental Administration in Finland: Essays of Institutional and Policy Change within the Finnish National and Local Environmental Administration*. Abo: Abo Akademi University Press.

Kirkhart, K. (2000) 'Reconceptualizing Evaluation Use: An Integrated Theory of Influence', in V. Caracelli and H. Preskill (eds) *The Expanding Scope of Evaluation Use*, New Directions for Evaluation 88, pp. 5–24. San Francisco, CA: Jossey-Bass.

Lampinen, O. (1992) *The Utilization of Social Science Research in Public Policy*. Suomen Akatemian julkaisuja 4. Helsinki: VAPK-kustannus.

Leeuw, F. L. (2002) 'Evaluation in Europe 2000: Challenges to a Growth Industry', *Evaluation* 8(1): 5–12.

Leviton, L. and E. Hughes (1981) 'Research on the Utilization of Evaluations', *Evaluation Review* 4(5): 525–48.

Litfin, K. T. (1994) *Ozone Discourses*. New York: Columbia University Press.

Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.

Marcussen, M. (2001) 'The OECD: Transnational Regulation through the Idea-Game', paper prepared for the SCANCOR Workshop, 'Transnational Regulation and the Transformation of States', Stanford University, 22–3 June.

Martin, S. and I. Sanderson (1999) 'Evaluating Public Policy Experiments: Measuring Outcomes, Monitoring Processes or Managing Pilots?', *Evaluation* 5(3): 245–58.

Mol, A. P. J. and G. Spaargaren (2000) 'Ecological Modernisation Theory in Debate: A Review', *Environmental Politics* 9(1): 17–49.

Morabito, S. M. (2002) 'Evaluator Roles and Strategies for Expanding Evaluation Process Influence', *American Journal of Evaluation* 23(3): 321–30.

Muller, P. and Y. Surel (1998) *L'Analyse des politiques publiques*. Paris: Montchrestien.

O'Connor, M., S. van Hove, J. O'Neill and C. Hue (2001/2002) *Le Modèle politique de la démocratie délibérative*. C3ED, Département de Sciences Economiques et Management. Université de Versailles Saint-Quentin-en-Yvelines. EEE-SDP Module 'Gouvernance'. 50 pp.

OECD (1997a) *OECD Environmental Performance Reviews: A Practical Introduction*. OECD Environment Monograph. OCDE/GD(97)35. Paris: OECD.

OECD (1997b) *Workshop on the Second Cycle of Reviews. Lessons from the First Cycle of Reviews* (document prepared by the Finnish Delegation). OECD Environment Directorate/Environment Policy Committee, Group on Environmental Performance. Paris, 17 Nov. 1997. Room document No. 12 for the workshop.

OECD (2001a) *OECD Environmental Indicators: Towards Sustainable Development*. Paris: OECD.

OECD (2001b) *Indicateurs clés d'environnement*. Paris: OECD.

Owen, J. M. (2002) 'Organisational Learning as an Accountability Commitment: Lessons from Innovative Practice', paper presented at the Conference of the European Evaluation Society 'Three Movements in Contemporary Evaluation: Learning, Theory and Evidence', 10–12 Oct., Seville.

Pagani, F. (2002) *Peer Review: A Tool for Co-operation and Change. An Analysis of an OECD Working Method.* Paris: OECD. SG/LEG(2002)1.

Patton, M. Q. (1997) *Utilization-Focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks, CA: SAGE.

Patton, M. Q. (1998) 'Discovering Process Use', *Evaluation* 4(2): 225–33.

Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: SAGE.

Perrin, B. (1998) 'Effective Use and Misuse of Performance Measurement', *American Journal of Evaluation* 19(3): 367–79.

Perrin, B. (1999) 'Performance Measurement: Does the Reality Match the Rhetoric? A Rejoinder to Bernstein and Winston', *American Journal of Evaluation* 20(1): 101–11.

Perrin, B. (2002) 'Towards a New View of Accountability', paper presented at the Conference of the European Evaluation Society 'Three Movements in Contemporary Evaluation: Learning, Theory and Evidence', 10–12 Oct., Seville.

Pitarelli, E. and E. Monnier (2000) 'Benchmarking: The Missing Link between Evaluation and Management?', paper presented at the European Evaluation Society Fourth Conference 'Taking Evaluation to the People: Between Civil Society, Public Management and the Polity', 12–14 Oct., Lausanne.

Pollit, C. (1998) 'Evaluation in Europe: Boom or Bubble?', *Evaluation* 4(2): 214–24.

Radaelli, C. M. (1995) 'The Role of Knowledge in the Policy Process', *Journal of European Public Policy* 2(2): 159–83.

Saarikoski, H. (2000) 'Environmental Impact Assessment (EIA) as Collaborative Learning Process', *Environmental Impact Assessment Review* 20(6): 681–700.

Sabatier, P. A. (1987) 'Knowledge, Policy-Oriented Learning and Policy Change: An Advocacy Coalition Framework', *Knowledge: Creation, Diffusion, Utilisation* 8(4): 649–92.

Sabatier, P. A. and H. C. Jenkins-Smith (1993) *Policy Change and Learning: An Advocacy Coalition Approach.* Boulder, CO: Westview.

Sabatier, P. A. and H. C. Jenkins-Smith (1999) 'The Advocacy Coalition Framework: An Assessment', in P. A. Sabatier (ed.) *Theories of the Policy Process*, pp. 117–66. Boulder, CO: Westview.

Sairinen, R. (2000) *Regulatory Reform of Finnish Environmental Policy*. Espoo: Helsinki University of Technology, Centre for Urban and Regional Studies Publications A 27.

Schubert, U., A. Martinuzzi, E. Huchler, M. E. Langer and B. Obermayr (2000) 'Evaluation Nachhaltiger Entwicklung: Herausforderung für Methodenentwicklung und Evaluationspraxis', in M. Härtel and R. Stockmann (eds) *Berufliche Umweltbildung und Umweltberatung: Grundlagen, Konzepte und Wirkungsmessung*. Bielefeld: Buchreihe des Bundesinstituts für Berufsbildung (BIBB). Bertelsmann-Verlag. http://www.sustainability. at/easy.

Schwandt, T. A. (2001) 'A Postscript on Thinking about Dialogue', *Evaluation* 7(2): 264–76.

Siebenhühner, B. (2001) 'How do Scientific Assessments Learn? A Comparative Study of LRTAP and IPCC', paper presented at 'Frontiers 1' European Society for Ecological Economics Conference, 4–7 July, Cambridge.

Söderbaum, P. (2000) *Ecological Economics: A Political Economics Approach to Environment and Development.* London: Earthscan.

Söderbaum, P. (2001) 'Political Economic Person, Ideological Orientation and Institutional Change: On Competition between Schemes of Interpretation in Economics', *Journal of Interdisciplinary Economics* 12: 179–97.

Strang, D. and P. M. Y. Chang (1993) 'The International Labour Organisation and the Welfare State: Institutional Effects on National Welfare Spending, 1960–1980', *International Organization* 47(2): 235–62.

Tamminen, R., A. Manninen and J. Pellinen (1999) 'The Concept of Accountability: From Stewardship to Discipline', *Hallinnon tutkimus (Administrative Studies)* 18(1): 33–47.

Taylor Fitz-Gibbon, C. (2002) 'Evaluation in an Age of Indicators: Challenges for Public Sector Management', *Evaluation* 8(1): 140–8.

Thoenig, J.-C. (2000) 'Evaluation as Usable Knowledge for Public Management Reforms', *Evaluation* 6(2): 217–29.

Uusikylä, P. (1999) 'Politiikan ja hallinnon arviointi (Evaluation of Policy and Administration)', in R. Eräsaari, T. Lindqvist, M. Mäntysaari and M. Rajavaara (eds) *Arviointi ja asiantuntijuus* (Evaluation and Expertise), pp. 17–30. Helsinki: Gaudeamus.

Valovirta, V. (2002) 'Evaluation Utilization as Argumentation', *Evaluation* 8(1): 60–80.

Van der Knaap, P. (1995) 'Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation?', *Evaluation* 1(2): 189–216.

Van der Knaap, P. (2000) 'Performance Management and Policy Evaluation in the Netherlands: Towards an Integrated Approach', *Evaluation* 6(3): 335–50.

Vedung, E. (1991) *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.

Weale, A. (1992) *The New Politics of Pollution*. Manchester: Manchester University Press.

Weiss, C. H. (1980) 'Knowledge Creep and Decision Accretion', *Knowledge: Creation, Diffusion, Utilisation* 1(3): 381–404.

Weiss, C. H. (1987) 'The Circuitry of Enlightenment: Diffusion of Social Science Research to Policymakers', *Knowledge: Creation, Diffusion, Utilisation* 8(2): 274–81.

Weiss, C. H. (1998) 'Have we Learned Anything New about the Use of Evaluation?', *American Journal of Evaluation* 19(1): 21–33.

Weiss, C. H. (1999) 'The Interface between Evaluation and Public Policy', *Evaluation* 5(4): 468–86.

Wimbush, E. and J. Watson (2000) 'An Evaluation Framework for Health Promotion: Theory, Quality and Effectiveness', *Evaluation* 6(3): 301–21.

MARKKU LEHTONEN is finalizing his PhD on environmental policy evaluation at the Centre d'Economie et d'Ethique pour l'Environnement et le Développement at the University of Versailles Saint-Quentin-en-Yvelines. Please address correspondence to: 60, rue de la Colonie, 75013 Paris, France. [email: makelehtonen@yahoo.com]