The Canadian Journal of Program Evaluation Vol. 12 No. 1 Pages 1–18
ISSN 0834-1516

EVALUATING EVALUATION IN THE EUROPEAN COMMISSION

Roger P. Levy
Robert Gordon University
Aberdeen, Scotland

Abstract: There is an increasing demand for the evaluation of expenditure and regulatory measures undertaken by the European Union in order to improve accountability and achieve "value for money" objectives. At the most general level, the task of organizing evaluation systems for these programs falls to the European Commission. Historically, the commission has focused on developing policies rather than monitoring or delivering them. With the maturing of certain policy areas, the commission's role is shifting in the direction of review and evaluation. From a systemic point of view, the management of EU policy presents particularly severe challenges in the area of evaluation. There are multiple actors located at local, national, and supranational levels; divergent administrative cultures and practices; variable quality of information, records, and capabilities; an attenuated system of reporting; and unclear lines of accountability. Joint funding of some programs creates additional problems by entangling program impacts and the audit purposes and management of national and EU institutions, respectively. The commission has still to come to a view on the parameters defining commonality and diversity in evaluation. In order to improve the situation, the commission has taken a number of initiatives, among them the setting up of an expert working group on the evaluation process. This article reports, in general terms, the findings of that group for 1994–95.

Résumé: On demande de plus en plus l'évaluation des dépenses et des mesures régulatrices prises par l'Union Européenne dans le but de mettre l'accent sur la responsabilité et de réaliser des objectifs de rentabilité. Il incombe à l'Union Européenne d'élaborer des systèmes d'évaluation pour ces programmes. Dans le passé, la Commission s'est concentrée surtout sur le développment des politiques plutôt que sur le contrôle de ces politiques et sur leur mise en place. Dans les domaines où les politiques sont en place depuis un certain temps, le rôle de la Commission change et va continuer à changer. Désormais sa tâche sera de les reviser et de les évaluer. Quant aux systèmes à adopter pour l'évaluation, la gestion de la politique de l'Union Européenne lance des défis complexes. Il y a de nombreux acteurs aux niveaux local, national et supranational, des cultures et des pratiques administratives divergentes, de l'information, de la gestion des dossiers et des capacités de qualité inégale, un système atténué de reportage et de l'imprécision dans les lignes hiérarchiques. Un financement commun de certains programmes crée des problèmes supplémentaires en mélangeant l'impact des programmes, les buts de la vérification, et la gestion des institutions aux niveaux national et européen. La Commission a encore à se prononcer sur les critères pour définir l'uniformité et la diversité dans l'évaluation. Pour améliorer la situation, la Commission a pris un certain nombre de mesures. Cet article présente les conclusions du groupe de travail d'experts sur le processus d'évaluation de 1994 à 1995.

In recent years, interest in the management of European Union (EU) programs has focused on the issue of budgetary fraud and irregularity — its level, causes, and ways of reducing it (legal, organizational, etc.). However, in concert with the development of the "new public management" (Farnham & Horton, 1993; Greer, 1994; Hood, 1991) already underway in the OECD countries, increasing attention has been paid to other management processes and techniques, including value for money (VFM)/comprehensive/performance auditing and program evaluation and review, aimed at enhancing and delivering accountability goals (Henkel, 1991; Power, 1994).

From a systemic point of view, the management of EU programs creates particularly severe challenges in the areas of both audit and evaluation. On the one hand, the management capabilities of the European Commission have been characterized as deficient because of basic structural and cybernetic weaknesses (Metcalfe, 1992, 1994), or as a result of a loss of power to the member states and a growing imbalance between responsibilities and capabilities (Kok, 1989; Van Heukelen, 1994). In common with other international organizations, the commission is perhaps expected to perform tasks for which it is not equipped (Jordan, 1995). On the other hand, the overall implementation environment is very complex, with multiple actors located at local, national, and supranational levels, divergent administrative cultures and practices, a variable quality of information, an attenuated system of reporting, and unclear lines of responsibility. Joint funding of many programs creates additional problems by entangling program impacts and audit purposes at national and EU levels (Levy, 1994, 1996).

In recognition of these problems, the commission has in recent years taken a variety of regulatory, managerial, and policy initiatives to improve auditing and evaluation. This article reviews the background to these initiatives and focuses on the 1994–95 work of the evaluation study group established by the commission's Directorate General for Budgets (henceforth DG19). After analyzing the general forces for change, I look at the immediate events that catalyzed the initiative. I then outline the terms of reference of the study, the coverage of the policy areas, and the methodology employed. This is followed by an analysis of the evaluation process, the findings on the structure of evaluation within the commission, and some of the general recommendations made. I conclude with an assessment of the future for evaluation within the commission in the context of the management of change.

CATALYSTS FOR CHANGE

The catalysts for increased evaluation of commission-managed programs are numerous and diverse, but can be grouped under five general headings: internal institutional, legal regulatory, macropolitical, financial, and policy life-cycle.

Internal institutional pressures refer to those emanating from a broadly defined accountability alliance within the EU institutions and from managerial spillover effects. The accountability alliance is made up principally of the European Court of Auditors (ECA), the Budgetary Control Committee (BCC) of the European Parliament, and some parts of the commission itself (those directorates general particularly concerned with budget formulation and monitoring and control [DG19 and DG20], and monitoring units within policy DGs). The ECA has been engaged in VFM audit since its inception in 1977, raising awareness of efficiency and effectiveness issues in its special and annual reports (the commission must respond to these). Although "VFM audit" and "program evaluation" are not synonyms, they do border the same territory (Gray, Jenkins, & Segsworth, 1993; Roberts & Pollitt, 1994) and occasionally even stray into each other's territory (Goldsworthy, 1995); and this has been an important factor in this case (for example, see interview with the acting director general of DG20 in Pratley, 1995).

Acting mostly although not always in concert with the ECA, the BCC has also put pressure on the commission through committee hearings and, more significantly, via the recommendations contained in its annual budget discharge decision. The latter have placed a series of obligations on the commission to improve its financial management and evaluation capabilities. For example, the discharge resolution of May 31, 1993, demanded "increased monitoring, control and evaluation operations by the commission" of structural funds, and more *ex ante* evaluation generally (European Parliament, 1993). The BCC spends much of its time on straightforward legality and regularity issues, particularly budgetary fraud and the routine requests from the commission for budgetary virement, so it does not have a sole focus on either VFM or program evaluation.

The accountability units within the commission itself have attempted to raise the profile of evaluation by a variety of means, which we will review towards the end of this article. However, the spread of multiyear programming from the European Development Fund (EDF — the EU's principal foreign aid program) to other areas of EU spending — principally research and development and structural actions of regional economic assistance — can be seen as an example of managerial spillover that allows for the development of review processes scarcely possible under annual budgeting systems. Since the establishment of the five-year financial perspective from 1988, the idea of multiannuality has become generalized. The certainties of the system make it both possible and desirable to carry out *ex ante* and *ex post* program evaluation work, although this does not mean these activities actually take place. National, linguistic, and functional faultlines criss-cross the organization of the commission, fracturing the management process between (currently) 23 DGs and 15 nationalities.

Legal regulatory pressures refer to instruments such as the financial regulation of 1990, which requires the EU institutions to manage for "la bonne gestion financière" (broadly, VFM) and to measure outputs; and the structural fund regulations of 1988 and 1993 with their heavy requirements for *ex ante*, ongoing, and *ex post* evaluation by both the commission and the member states. Funding recipients must stipulate what arrangements they have made for evaluation in their original submissions, which have to be submitted by evaluation units within the commission policy directorates involved. These proposals are subject to further scrutiny by DG20 (Financial Control), which can refuse endorsement if review and evaluation schedules are deemed insufficient. Within common agricultural policy (CAP) legislation, there is far more emphasis on specific control measures than on evaluation, and the position is decidedly underdeveloped from a regulatory point of view.

Macro-political pressures refer to the growing demands for greater accountability from net contributor member states (the U.K. and Germany in particular) and from taxpayer interests. As the number of member states who are net contributors to the budget has increased, so the demands have intensified (the issue was raised during the French presidential campaign of 1994, for example); and with the accession of Austria, Sweden, and Finland in January 1995 it is quite possible that net contributors will for the first time outnumber net beneficiaries. This creates a potential coalition for change until the next enlargement, which will in all probability again reverse the balance. Within member states, the demand for accountability of co-funded programs (most notably in structural actions) also has a spillover effect on the EU contribution to the total.

Political and financial pressures are closely linked. The 1980s and 1990s have seen public expenditure programs in the member states come under severe pressure from two opposing directions: ever-increasing demands, particularly in the fields of health and welfare as demographic changes work through, versus the desire to reduce taxes or at least keep them stable. This has resulted in an explosion of interest in VFM and program evaluation throughout the OECD countries. Added to this is the requirement within the Treaty on European Union of 1992 that national budget deficits must be limited to 60% of GDP and public borrowing to 3% of GDP in any one year for a state to qualify for entry to full monetary union in 1999. At the same time, commitments to EU programs (especially structural actions) have grown rapidly as a result of the single-market project. Agreement to raise finances for these programs has been accompanied by exhortations to secure effective use of the money.

The *policy life-cycle influence* is technical rather than political and is based on a model of organizational ageing. From a historical point of view, the commission has focused on developing and establishing policies and policy frameworks rather than on monitoring or delivering them. The latter tasks have for reasons of practicality or political expediency essentially been left to agencies in the member states. Although the doctrine of subsidiarity provides no simple guarantee that policy making is always the function of the Council of

Ministers or commission and implementation is always the task of member state agencies, there is a substantial degree of accuracy in this distinction. The main exception to this general rule is the commission's role in drawing up annual (and latterly multiyear) budgets for approval and amendment by the Council and, to a lesser extent, the European Parliament. But even the commission's responsibility under the treaty to "implement the budget" (Article 205) has a quite restricted meaning, as it is the member state agencies that actually administer and manage the funds at ground level.

A recent House of Lords report titled "Financial Control and Fraud in the Community" referred to the predominance of a spending culture within the commission (House of Lords, 1994), and one can argue that the influence of accountability imperatives is quite weak within such an organization. With the maturing of certain policy areas, it is arguable that the commission's role is changing in the direction of more review and evaluation, and it is significant that the president of the commission, Jacques Santer, has referred to the "profound reform in financial management practices" underway in the organization (Santer, 1995). The evaluation study is only one component in a range of initiatives — for example, the three-phase sound financial management program launched by the commission in 1995 — designed to bring about a change in culture. Thus, we can expect that resources will be directed away from the production of policies and towards the monitoring and evaluation of outputs.

THE DG19 EVALUATION STUDY, 1994–95

The immediate background to DG19's decision to establish an expert working group on the evaluation capabilities of the commission as a whole has to be seen in the overall context outlined above. The effectiveness requirement in the 1990 financial regulation, Declaration 18 of the treaty to evaluate EU legislation "taking into account costs and benefits," and the commitment of the Edinburgh summit to ensure VFM through prior appraisal and periodic review were all cited as reasons for the study by DG19, which demonstrated both the legislative and macropolitical impetus behind the initiative and the clear "financial" focus of the exercise.

On the other hand, institutional pressure emanating from three sources seemed to play a more immediate role in the timing. First, there was a request from the European Parliament to furnish a report describing the systems of evaluation and monitoring for each expenditure program in order to show that objectives were being achieved at reasonable cost. Second, the draft budget agreed by the Council of Ministers (the EU's main legislative authority) for 1995 stated that the "effectiveness of Community actions can only be appropriately appreciated if they have clear and verifiable objectives and if there are procedures for monitoring and evaluating progress with regard to these objectives" (European Council, 1994).

From the commission side, the cabinet of the then-commissioner for budgets led the initiative despite some misgivings elsewhere, and other evaluation initiatives either underway (in D69) or already completed (by DG20). Significantly, the project was managed by the Directorate General for Budgets (DG19) rather than Financial Control (DG20).

Aims and Terms of Reference

The general aim of the study was to obtain an informed appraisal from a group of outside experts of the role and quality of program evaluation as it related to the formulation, budgeting, and implementation of EU policies, in order to improve program management and the value for money obtained from expenditure. Nine papers were commissioned, of which eight focused on four specific policy areas. The ninth paper was to provide a more discursive overview of evaluation capabilities and prerequisites, bringing in experiences from an international perspective. In the four policy areas, one study focused on the systems and capabilities within the area as a whole, and the other was designed to provide a detailed account of the evaluation of one program or policy sector.

The prescribed structure for the studies was divided into four parts by the terms of reference. This assumed a rational framework for evaluation based on traditional assumptions of organizational structure and behavior, which were as follows:

1. an analysis of the role of evaluation in the area concerned and how it fitted in with other parts of the management system (policy formation, objective setting, programming, *ex ante* appraisal, budgeting, implementation, monitoring, *ex post* evaluation, feedback);
2. an appraisal of the regulatory and logistical frameworks (legal requirements, financial and human resources utilized, training and reporting of internal evaluators, use of out-

side contractors, frequency, partnership with agencies in the member states);
3. an assessment of the substance of extant evaluations within the policy area with specific regard to the nature and utilization of effectiveness measures, impact measures, efficiency measures, and option appraisal;
4. an analysis of the use made of the evaluations by commission program managers, the Council of Ministers, the European Parliament, and the end users of the programs.

The terms of reference were very broad, attempting to take in the management framework, legal basis, methodology, and substance of evaluation reports, and the use of evaluations within the prescribed areas. The inclusion of the use of evaluations and user perceptions is interesting. As Knox and McAlister (1995) observe, although policy makers see the involvement of users in evaluation as increasingly important, simply defining who end users are — let alone assessing their demands — is a daunting task. In this case, the study allowed for a reasonable envelope for the first three objectives, but it was never going to be possible to survey and analyze the end users of evaluations within that framework.

Methodology

The methodology of the study group is worth looking at from a number of angles. In the first place, the policy areas chosen for analysis comprised the biggest spending items in the total budget (structural actions, agriculture, R&D, and the European Development Fund [EDF]), although the EDF of course remains outside the normal budget. This made sense from the point of view of coverage of as much as the budget as possible, but can also be seen as a nondiscriminatory methodology designed to offend everybody equally and nobody in particular. Given the relatively slim resources available for the study, there may have been a case for a more focused approach; this may, however, have met far more resistance in the targeted areas. The strategy chosen is perhaps symptomatic of organizational dynamics where central managing units are relatively weak and coordination is poor.

Second, the composition of the group was diverse both nationally (an invariable characteristic of EU working groups) and professionally, and included policy analysts, economists, auditors, and evaluators. Although the initial briefing sessions were designed to develop and refine the terms of reference and agree on common definitions (the original terms of reference were significantly modified as a result), it was inevitable that different approaches would be followed, especially as the group was dispersed geographically and was unlikely to meet spontaneously to hone a common methodology. It can be argued that the diversity of the policy sectors would have made such an endeavor futile, as it precludes a fruitful methodological pluralism. On the other hand, a greater attention to methodology at the outset may have refocused effort on some of the key issues and brought group learning together earlier rather than later.

At the technical level, the fieldwork methodology comprised a combination of elite interviewing and document analysis in varying proportions, with the experts bringing their own theoretical predilections to bear on the data (hence the variations in outputs). Access to officials and documents was limited because of time and budgetary constraints, and there was the added complication of restricted access to some internal files. On a quantitative level documentation was plentiful, occasionally overwhelmingly so. It fell into six main categories:

1. internally produced evaluations circulated internally indicating internal evaluation capacities, methodologies, concerns, and focus;
2. externally produced evaluations circulated internally indicating policy concerns at the time and the sources of outside learning;
3. internally produced evaluations circulated externally (end user documents in the public domain) that were important for comparison with (1) and (2) above;
4. externally produced evaluations circulated externally that showed the range of expert evaluation sources in the area and its divergence/convergence with internal work;
5. legislation relevant to the sector setting out evaluation requirements, if any (these are extensive in some sectors and nonexistent in others);
6. supporting documentation, such as budget figures and projections.

PRINCIPAL FINDINGS AND OUTCOMES

It will already be evident that the process of evaluating commission evaluation was beset by significant logistical, technical, and methodological difficulties. This resulted, for example, in a number of alternative “scoring” frameworks for evaluation studies reviewed; other papers, however, chose not to devise such systems, instead analyzing evaluations according to the criteria specified in the terms of reference. Inevitably, there was variation in the amount of documentation reviewed and officials interviewed in each study.

Given the fracturing of the commission as a whole and the closely guarded autonomy of each DG, it was perhaps predictable that the exercise did not receive a uniformly positive welcome from the DGs affected, and that access and cooperation varied. It is possible to speculate that other DGs viewed the purposes of the study and its author, DG19, with suspicion, seeing the process as a zero sum game that would result in more centralization, loss of departmental autonomy, and possible displacement of departmental function. In addition, some of the relevant documentation was of a politically sensitive nature and so could not be circulated.

Despite the heterogeneity of the study group and the theoretical base of its investigations, there was a substantial degree of consensus on the strengths and weaknesses of the current system, and on recommendations for the future. It is also possible to isolate a number of key features of the overall system from these observations.

Strengths

Four main areas of strength were identified. As far as the major policy areas were concerned, there were developed mechanisms in place for both requesting and executing program evaluation. In the area of structural actions, for example, a large number of evaluation reports (in excess of 300) had already been produced and supplied to the commission, and partnership strategies between the commission and member state agencies had been developed for evaluation. In the case of the EDF, the project cycle management (PCM) model for the evaluation of aid programs had been implemented in DG8 since the early 1980s, and a studies unit to evaluate different sectors of the common agricultural policy had been created in DG 6 (agriculture) in the late 1980s. Second, the use of external contractors was well developed in some areas, with regular and safe systems in place for contract selection and management. Third, the development of multiannual programs had provided a major structural incentive for program evaluation. Finally, there had been a discernible improvement in the clarification of objectives and in identifying impact indicators in some areas.

Weaknesses

Areas of weakness were predictably more numerous, although not necessarily so different from experience elsewhere (for example, see Hyndman & Anderson, 1995, on performance management systems in U.K. executive agencies). As was pointed out in a number of studies, problems of clarifying objectives, measuring impacts, designing indicators, and establishing a link between budgetary allocations and outputs are not unique to the commission. Indeed, they have been acknowledged by the member states themselves when questioned about their own evaluations of, for example, structural actions (House of Lords, 1991; see also Carter, Klein, & Day, 1992, for a more general discussion).

Second, there was reasonable frequency of evaluation of multiannual programs, but this was complicated by a program management cycle in which new programs started before old ones had finished, or where delays in program starts "squashed" the cycle into a much shorter time frame than had been intended. In these latter cases, midterm evaluations in particular would tend to show poor take-up and utilization. The overall result was that there was limited opportunity for evaluation results to be taken into account either in existing programs or in the design of the next cycle. As far as annual programs were concerned, there was no rolling program of evaluation at all.

Third, although the larger program areas did possess an evaluation capability, smaller programs did not. In addition, there was a discernible "missing link" between the skills in the evaluation units on the one hand and evaluation capabilities (i.e., the lack of them) among program managers on the other. This resulted in basic misunderstandings about the nature of evaluation within DGs. In some cases, routine management tasks and what might be termed microlevel evaluation were mistaken for program evaluation.

Fourth, although the use of external contractors was well developed in some areas, this might have come at the expense of developing evaluation capabilities among officials in the DGs. Furthermore, lack of evaluation expertise at this level meant that the product delivered by external contractors was of variable quality because of variation in specification design. Taken together, these factors could result in the marginalization of evaluation as part of the management process.

Fifth, it emerged that only a very small percentage of each budget was directly devoted to evaluation. This percentage was difficult to measure without a full breakdown of salary costs, but appeared to fall within 0.05–0.1% of program budgets. In the case of agriculture, there was considerably less spending on evaluation either relatively or absolutely compared with, for example, structural actions.

Sixth, progress in evaluation had tended to take place in isolation, with only limited spillover from one DG to another. For example, the PCM methodology had spread from DG8 to DG1, but no further. There was great scope for both the transfer of knowledge and inter-DG cooperation. Similarly, although partnership with member state agencies had been developed in some areas, there was still scope for improvement.

Finally, there was the issue of lack of transparency and the general failure to identify and involve a meaningful range of stakeholders in program evaluation. Consultation tended to involve only a limited range of interests (or even the consideration of a limited range of possible interests), which narrowed further as evaluations evolved from internal documents and files to option papers and finally to public policy documents. In some areas, consultation structures are well defined, hierarchical, and exclusive, in which established DG "clients" — including member state interests — have a disproportionate influence over the final shape of evaluations of key policy areas, to the exclusion of other stakeholders.

The overall picture that emerges is of a system that has considerable evaluation capability and potential but that is fragmented, suboptimal, and lacking in transparency. It is characterized by complexity and by poor integration and communication within the commission and between the commission and other agencies; a learning culture in which experience is either not shared or is discounted by others; an absence of agreed performance measures and systems; incomplete, nonstandardized, and unreliable data; and a variable quality and quantity of human resources available for evaluation.

Recommendations

The common recommendations focused on addressing these perceived weaknesses. Thus, considerable emphasis was placed on improving the human and organizational support structure for evaluation (inevitably involving more resources) and for developing evaluation methodologies. Concretely, this would mean training officials in specialized evaluation skills and spending a greater proportion of program budgets on evaluation. In addition to strengthening evaluation units in the program directorates, there was also support for the establishment of a centralized horizontal evaluation support unit that could act as a catalyst for learning and the spreading of best practice, and could perhaps carry out across-the-board evaluations of its own. In general, communication channels within and between directorates in the commission needed to be increased in number and widened in nature to better disseminate results.

On the methodological side, it was recommended that a greater range of both quantitative and qualitative indicators be developed that included the collection and use of fieldwork data, focused on impacts, were fit for the specific programs being evaluated, and were developed in conjunction with ground-level implementation agencies. The latter would mean working with national agencies, including joint training of staff and so on. The payoff would be the development of suites of measuring instruments that enable genuine cross-national comparisons of program results. These methodologies should themselves be subject to regular external review and appraisal. The use of independent experts could be extended (not a view taken by DG19) and the management of them improved.

As for the problem of fitting evaluations into the project cycle and choosing appropriate subjects for evaluation, it was suggested that a more pragmatic approach was needed, involving targeting some areas and leaving out others altogether. The need to improve communications alluded to earlier also related to the key recommendation to improve transparency. Stakeholder access and participation could be improved by opening up the consultation process earlier and creating new forums for the discussion of evaluation findings.

FUTURE PROSPECTS

The fate of the expert papers and their recommendations of course lies in the hands of the original client. A process of consultation is anticipated in which individual directorates respond to the reports, with recommendations emerging only after this has taken place. Should such a process become part of the more general reform underway in the commission, it will add to the management-of-change dynamic that is already discernible.

In the context of the diverse and fractured nature of the management system, what kind of model or combination of models is appropriate, and what strategy is necessary to achieve them? Is a centrally based system organized on the principles of rational hierarchy (typical of most performance management systems) either possible or desirable, or is some more devolved, self-regulating system more appropriate? (Elements of both approaches are present in the expert papers.)

At the level of high politics and legality, the principle of subsidiarity (reinforced by the Treaty on European Union), which appears to reaffirm the role of the member states as the most appropriate level of program management, makes a centralist solution unlikely. At the level of low politics, the administrative capacities of the commission also preclude this option. Nevertheless, there seems to be a more general tendency among decision makers to search for these kinds of answers despite all the evidence in the organizational theory literature. Surveying the management of change in the U.K. public sector, Stewart and Walsh observed that "there was a faith in hierarchy and bureaucratic rationality" (1992, p. 509). Similarly, in her study of the effect of the fraud issue on EU institutional structures, Mendrinou (1992) noted that the commission had been very active in promoting centralist solutions to the problem.

This activity has had a number of incarnations. At the operational level, it has involved the standardization of internal procedures and the introduction of evaluative tools to improve cost effectiveness. Article 2 of the financial regulation has proved particularly useful in this regard (Pratley, 1995, p. 253), and in structural actions the MEANS program, which is designed to develop a common methodology for evaluation, has been underway since 1992. The financial management initiative referred to earlier is a recent example of this strategy. Another development has been the creation or strengthening of centralized management agencies within the commission (e.g., the antifraud unit [UCLAF], the Inspectorate General, central control units in the directorates). However, it is questionable whether administrative proliferation of this type is helpful; it may instead be symptomatic of the characteristics of the system identified earlier. The failure to use and build on existing structures is perhaps indicative of the low trust/low learning culture.

The final aspect of the centralist strategy has been the imposition of legally binding evaluation requirements at treaty, program, and subprogram implementation level. The 1993 regulations governing the structural funds are the most well known example, but there are new provisions within the treaty as well as a range of specific regulations in the agricultural sector that in one way or another impose financial management requirements on member state agencies. As with the creation of centralized units, this approach has its limitations. The annual and special reports of the European Court of Auditors and the commission's annual review of the operation at the structural level provide ample evidence of failures in implementation. Ever more stringent legal or administrative imperatives from the centre are not in themselves sufficient to produce good auditing and evaluation on the ground.

The centralist strategy is not, however, the only one at work. As we have seen in connection with the structural funds (Levy, 1995), the commission has tried hard to develop joint evaluation based on commonly agreed principles across levels of management. Indeed, the MEANS program is also trying to address the issue of how to realise decentralization and partnership in evaluation, and to use evaluation to strengthen partnership (European Commission/Centre d'evaluation, 1995). In addition, there are some legislative instruments that provide material incentives in this direction or simply insist that it happen. As Bowerman's (1994) study of liaison between the National Audit Office and the Audit Commission in the U.K. shows, cooperation and coordination in auditing has proved extremely difficult even in unitary states with a long tradition of and developed infrastructure for state auditing. Imagine the difficulties in trying to coordinate joint evaluations with a variety of partner agencies all with varying cultures, methodologies, and capabilities in a policy environment that is often adversarial, as is the case, for example, in the drawing up and approval of structural fund programs at the outset of the cycle. Although the absence of agreed measuring instruments and indicators may be advantageous in allowing greater flexibility in moving to agreed solutions, this is offset by the competitive relationship between actors in the process. Thus, movement toward partnership has been painfully slow and subject to many distortions in the area of structural fund evaluation (Smith & Spenlehauer, 1994).

With all its difficulties, an evaluation partnership between the commission and managing agencies in the member states is the only realistic way forward given the disposition of overall evaluation capabilities. As with auditing, the most the commission can really hope to achieve is a meta-evaluation function. To the extent that the evaluation study discussed here looked at the issue of encouraging and sharing learning within the commission service itself, this agenda has been advanced. To take the process a step further, however, means extending the scope of the evaluation of evaluation to look at how the commission and member state agencies can more productively cooperate in evaluation. As stated in Osborne, Bovaird, Martin, Tricher, and Waterston's (1995) critique of the rationalist model, it also means revisiting the issue of stakeholder input to get a more rounded view of performance management.

REFERENCES

Bowerman, M. (1994). The NAO and the Audit Commission: Cooperation in areas where their VFM responsibilities interface. *Financial Accountability and Management, 10*(1), 47–64.

Carter, N., Klein, R., & Day, P. (1992). *How organizations measure success: The use of performance indicators in government*. London: Routledge.

European Commission/Centre d'evaluation. (1995, November). *MEANS methods for evaluating structural policies for improved evaluation of community programmes: Results of the 1995 MEANS Programme*. Brussels: Centre d'evaluation.

European Council. (1994). *Draft budget 1995*. Luxembourg: Office for the Official Publications of the European Communities.

European Parliament. (1993, May 31). Discharge for 1991 budget. *Official Journal of the European Communities*, C150 Vol. 36.

Farnham, D., & Horton, S. (Eds.). (1993). *Managing the new public services*. London: Macmillan.

Goldsworthy, D. (1995, September). *Evaluation in an audit body: Experiences from the National Audit Office.* Paper presented to the U.K. Evaluation Society, London.

Gray, A., Jenkins, B., & Segsworth, B. (Eds.). (1993). *Budgeting, auditing and evaluation: Functions and integration in seven governments*. New Brunswick, NJ: Transaction.

Greer, P. (1994). *Transforming central government*. Buckingham, U.K.: Open University Press.

Henkel, M. (1991). The new evaluative state. *Public Administration*, *69*(1), 121–139.

Hood, C. (1991). A public management for all seasons. *Public Administration, 69*(1), 3–19.

House of Lords, Select Committee on the European Communities. (1991). *4th Report*. Regional Development Policy, Session 1991-92, HL paper 20. London: HMSO.

House of Lords, Select Committee on the European Communities. (1994). *Financial control and fraud in the community*. Session 1993-94, 12th Report, HL paper 75. London: HMSO.

Hyndman, N., & Anderson, R. (1995). The use of performance information in external reporting: An empirical study of the UK executive agencies. *Financial Accounting and Management*, *11*(1), 1–19.

Jordan, A. (1995). Designing new international organizations: A note on the structure and operation of the global environment facility. *Public Administration*, *73*(2), 303–312.

Knox, C., & McAlister, D. (1995). Policy evaluation: Incorporating users' views. *Public Administration*, *73*(3), 413–436.

Kok, C. (1989). The Court of Auditors of the European Communities: The other European court in Luxembourg. *Common Market Law Review, 26*, 345–367.

Levy, R. (1994). Audit and accountability in a multi-agency environment: The case of the common agricultural policy in the UK. *Financial Accountability and Management*, *10*(1), 65–75.

Levy, R. (1995, May). Subsidiarity, accountability and the management of EU programmes. Paper delivered at the European Community Studies Association, Charleston, SC.

Levy, R. (1996). Managing value for money audit in the European Union: The challenge of diversity. *Journal of Common Market Studies, 34*(4), 509–529.

Mendrinou, M. (1992). *European Community fraud and institutional development*. Manchester: European Policy Research Unit, University of Manchester.

Metcalfe, L. (1992). After 1992: Can the Commission manage Europe? *Australian Journal of Public Administration, 51*(1), 117–130.

Metcalfe, L. (1994). International policy co-ordination and public management reform. *International Review of Administrative Sciences, 60*, 271–290.

Osborne, S., Bovaird, T., Martin, S., Tricher, M., & Waterston, P. (1995). Performance management and accountability in complex public programmes. *Financial Accounting and Management*, *11*(1), 19–37.

Power, M. (1994). *The audit explosion*. London: Demos.

Pratley, A. (1995). Auditing, financial management and evaluation: An interview with Alan Pratley. *Evaluation*, *1*(2), 251–263.

Roberts, S., & Pollitt, L. (1994). Audit or evaluation? A national audit office VFM study. *Public Administration, 72*(4), 527–550.

Smith, A., & Spenlehauer, V. (1994). Policy evaluation meets harsh reality. *Evaluation and Program Planning, 17*(3), 277–287.

Stewart, J., & Walsh, K. (1992). Change in the management of public services. *Public Administration*, *70*(4), 499–518.

Santer, J. (1995, October 20). The future of Europe: What role for the Commission? President's Speech to the Jean Monnet Conference, Florence, Italy (Reference: Speech/95/218).

Treaty on European Union. (1992). Luxembourg: Office for Official Publications of the European Communities.

Van Heukelen, M. (1994, December). The evaluation of European public policies. Speech to the First European Evaluation Society Congress, The Hague.