Evaluation

http://evi.sagepub.com

Spanish Evaluation Practice Versus Program Evaluation Theory: Cases From Five Policy Areas

Xavier Ballart

Evaluation 1998; 4; 149
DOI: 10.1177/13563899822208509

The online version of this article can be found at:
http://evi.sagepub.com/cgi/content/abstract/4/2/149

Published by:
SAGE Publications
http://www.sagepublications.com

On behalf of:

The Tavistock Institute

Evaluation
Copyright © 1998
SAGE Publications (London,
Thousand Oaks and New Delhi)
[1356–3890 (199804) 4:2; 149–170; 004815]
Vol 4(2): 149–170

Spanish Evaluation Practice Versus Program Evaluation Theory

Cases From Five Policy Areas

XAVIER BALLART
Universidad Autónoma de Barcelona

In this article, a contrast is made between evaluation theory and practice. For this purpose, a typology is constructed that is based on seven authors and nine dimensions, and presents examples of evaluations in various policy areas in Spain. It is anticipated that there will be a mismatch between theory and practice. This can be explained by the difficulties government evaluators have finding the data, time and resources to apply the techniques and conduct the types of evaluation that scholars consider academically excellent. It can also be explained by the fact that government evaluators and decision-makers, in a national context where there is no pressure for experimental methods, choose to produce information on a broader set of measures that gives them insights on both program effectiveness and organizational performance, but does not inform them on whether there is a causal relation between government programs and observed effects. From the five examples of the practice of evaluation in Spain, it is observed that practical evaluation can provide a good 'picture' of the relative effectiveness and efficiency of government programs. It is also observed that evaluation results fit the type of information which program managers and stakeholders expect, increasing the chances of evaluation studies being effectively used. Finally, for these five examples, the evaluators performed very much as controllers would do in the private sector, where there is no concern for the type of causality academics try to find between program features and the consequences of programs.

Introduction

Even in those countries where program evaluation and other forms of technical analysis have been more intensively applied, it is not clear whether they have been as powerful and useful as expected (Weiss, 1987; Majone, 1988; Lynn, 1989; Lindblom, 1990). Program evaluation has been even less fortunate than other forms of analysis when measured by the number of offices in government that carry out retrospective program evaluation. Often, evaluation will be ordered by a government office and carried out by external teams but the general impression is that there is very little systematic retrospective examination of policies.

With regard to evaluation in the European context, despite the fact that there may not be many formal offices with this title and evaluation activities may not be systematic enough and may vary in their level of scientific sophistication and rigor, there is an increasing evaluation activity that is sponsored and promoted both by European institutions and national, regional and local governments.[1] As Moore says, policy analysis and evaluation have 'increased the appetite for fact-based management about the extent public programs achieve stated objectives' (1995: 34), and indeed, program evaluation may have significantly influenced the political discourse in some policy areas.[2]

In this article, first, I try to condense a large body of essentially methodological writing on evaluation research into a typology based on seven authors and nine dimensions in order to present a clear idea of what the literature understands as 'program evaluation'. Second, I present a synthesis of evaluation practice in Spain, taking examples from five policy areas that have been central in the development of evaluation in basic education, training and employment, minimum income guarantees, higher education and research, and primary health care.[3] Third, I analyse how evaluation practice differs from theoretical models in this specific setting.

It is anticipated there will be a mismatch between theory and practice. However, the contrast is also intended to show: (1) *what authors and theoretical ideas are more present*, and (2) *what type of information government institutions obtain when they embark on evaluation studies.*

Drawing an analogy with the private sector, we are interested in approaches to evaluation that can provide information, not only on *program effectiveness*, but also on *organizational performance*. At the same time, as in the private sector, we also expect that public managers and program stakeholders might well accept having *less accurate information on the causal relation* between programs and observed effects if they can have a *broader set of measures* that gives them significant insights on the performance of their organizations and programs. For example, in studying a minimum income program, evaluation can provide information on the difference it makes to poverty, or evaluation can provide information on whether those who need this kind of 'vaccine' are receiving it, whether there are territorial differences in implementation, or the extent to which the program paper work imposes difficulties for social workers.

Evaluation in a specific context might therefore differ from theoretical models. This can be explained by the fact that the data, time, resources or techniques are not available to the kind of evaluation that scholars find truly excellent. It can also be explained by the fact that there are always choices to be made in evaluation. Government evaluators and decision-makers may choose to maximize the production of information that best meets their needs, both for internal and political management. This choice also has relevance for the eventual utilization of evaluation results.

Theoretical Models in Program Evaluation

Program evaluation is often defined as a 'practice-driven field' that emerges as a response to the public sector need for accountability. It is also often categorized as an 'art' given the limitations that a purely 'scientific' approach encounters when a research is conducted in the real context of public policies with all the alliances and conflicts that may arise. This does not mean that program evaluation does not have a theoretical base. Indeed, the field has moved towards greater theoretical sophistication since it was first pointed out that it would be a serious mistake to overlook the importance of theory in program evaluation (Chen and Rossi, 1983). Program evaluation theory evolves from the concepts and methods of other disciplines (psychology, statistics, education, economics, public policy) but it is also the result of the lessons learned in trying to respond to demands for evidence on the value created with specific public policies (Shadish et al., 1991).

However, there is great diversity in practice. There is diversity in the ways of approaching evaluation tasks as a response to diverse evaluation demands and situations. There is also diversity in the activities conducted under the rubric of evaluation as a result of the different professional standing of evaluators—private firms, university professors, bureaucrats—and the variety of analytical tools they may use according to their substantive field of training. This diversity fuels a debate that goes well beyond the differences of methods making it more difficult to understand and categorize theoretical models.

Given the difficulties of making a selection, and the fact that some of the best theorists have already debated these issues, the framework for this paper is based on Shadish et al.'s *Foundations of Program Evaluation* (1991). The seven authors they selected to explain trends in the development of evaluation theory are taken as fundamental references to reflect diverse positions from which related authors depart. Despite the risk of oversimplification, their evaluation models are described in terms of the nine dimensions reflected in Table 1.[4] In addition to these nine dimensions, the first cell in each column is used to identify the author with key terms they introduced and to include other important authors that could be seen as related in the way they approach evaluation.[5]

Looking at Table 1, one can find multiple alternative ways to approach an evaluation ranging from social experiments to case studies. From this general perspective, the choice for one model generally means that an alternative research logic will not be pursued and therefore that the evaluation will be categorized accordingly. Evaluators and their work, or administrations and their requests for proposals, can be associated to Cook and Campbell's book on quasi-experimentation, to Guba and Lincoln's fourth generation evaluation, or to management control systems of performance, to name just three examples.

Table 1 suggests some convergence in the theories on how to address questions that typically concern evaluators at the beginning of any study. Indeed, this is the way theories are directly useful to practice.

- If the question is *what* to evaluate, theories offer various possibilities that range from net impact to social needs, including intermediate processes or cost-effectiveness. But they may also relate the object of an evaluation to the nature of the program, the stage in its development or the availability of data or resources.

Table 1. Theoretical Models in Program Evaluation

	Scriven	*Campbell*	*Weiss*	*Wholey*	*Stake*	*Cronbach*	*Rossi*
New terms	Science of valuing	Quasi-experiments	Enlightening model	Performance management	Responsive, naturalistic	Analysis of UTOS (people, treatments, observations settings)	Theory-driven model
Related authors		Boruch, Cook, Riecken, Rossi	Cronbach, Chelimsky, Palumbo, Wholey	Abramson, Bellavita, Horst, Nay, Scanlon	Guba, House, Lincoln, Patton		Berk, Chen, Freeman, Wright
Concept	Judgement of value	Produce information to improve decision-making	Clarify questions, ‘enlighten’ decision makers and stakeholders	Produce information for results management	Produce practical and useful information for program members	Clarify questions about: problem, program, context	Produce information to improve decision-making
Prioritized interest	Public interest	Agency, funders	Legislators, executives	Managers, legislators, executives	Program beneficiaries, and personnel	Legislators, executives	Agency, funders
Role of evaluator	Judge	Methods expert	Educator	Change agent	Service provider	Educator, knowledge diffusion	Methods expert
Object	Effects	Effects, causal relations	Effects, processes, context	Effects, activities, products	Activities	Effects, processes, context	Concepts, implement utility (effects, cost-benefit)
Source of criteria to determine value	Social needs	Objectives	Objectives, documents, stakeholders	Stakeholders	Clients, professional/ technical personnel	Stakeholders	Stakeholders, scientific literature
Methods	Driven by the question	Randomized experiment. quasi-experiment	Quantitative and qualitative methods	Evaluability assessments, performance indicators, rapid feedback evaluation	Qualitative methods Case studies	Quantitative and qualitative methods	Randomized experiment Quasi-experiment, qualitative methods for non-causal analysis
Perspectives on utilization	Instrumental	Instrumental	Conceptual, policy formulation	Instrumental	Instrumental	Conceptual, understand social reality	Instrumental, conceptual, political, theory formulation
When should evaluation be done?	Cost-free evaluation	Serious reform initiative, after initial implement. Mistakes	According to the three Is: interests, available information, ideologies	Cost-effectiveness	Useful for program members	According to relevance and need of information on the problem	Innovative programs, pilot-test stage
How to organize evaluation process?	External operation	External operation	Involve personnel, without being obtrusive	Integrated in the results-oriented management system	Permanent contact with the program	In close cooperation with the program	Paying attention to the timing and ways managers work

- Theories also offer several answers to the question of *who should decide* what is to be evaluated. It is usually the political leaders or the executive managers of a program that make decisions about its evaluation. However, some theoretical models argue that others, such as legislators, clients or professional staff, should either answer this question or at least be taken into account by decision-makers. Independently of who is the client of the study, there seems to be agreement on the need to work on stakeholders' values. It is also very important to consider information needs of decision-makers and to work closely with them and with the people working in the program in order to foster utilization.
- If the question is *how* to evaluate, answers reflect the methods division. At the same time, there is a degree of agreement on some ideas. For example, it may well be argued that case studies are not the best way to approach causality or that evaluating interventions in a context where knowledge is not well structured requires a more qualitative approach or at least some kind of statistical-observational analysis.
- Closely related to methods is the question on *when* to evaluate. Often, the answer is determined by technical considerations. However, the concern for utilization has given more weight to considerations of a political (interests, ideologies, hidden purposes) or administrative nature (urgency, perception of the evaluation by the personnel, extent that it may affect the regular delivery of services).
- A final question *is what should be expected from the evaluation*. Theories can be quite contradictory: for some, an evaluation only makes sense if it is directly useful; for others, an evaluation will be really useful if it contributes to understand a social problem, changes perceptions or suggests ways for reforms in public intervention. Again, there is some agreement in that there are evaluations that are closer to the idea of a service to the program while others are more an opportunity to generate a discussion of a higher theoretical level.

Evaluation Practice in Five Policy Areas

Elementary and Secondary Education

In October 1990, the Spanish Parliament approved the main piece of legislation that was introduced by the Socialist government of Prime Minister Felipe González to reform the education system. The new law, known as LOGSE (General Organization of the Education System Act) substantially modified previous legislation regulating education in Spain and started a process of reform that would take some years to complete. The change was important for several reasons:

- The new legislation extended by two years—until the age of 16 years—obligatory schooling and advanced the grade at which students make a choice concerning whether to follow an academic track leading to higher education or a professional training.[6]
- It established basic curriculum requirements in the context of a new situation characterized by decentralization, where self-governing regions and schools have different levels of authority to develop the content of their educational programs.
- It focused the reform on quality issues and on accountability. The emphasis on

> quality was developed by reference to elements such as school management, parent participation, teacher-training, pedagogical practice and curriculum development. Accountability was defined as a need to inform the society about the overall performance of the system which remains a fundamental responsibility of the central government.

In June 1993, a National Institute of Quality and Evaluation (INCE) was created, within the Ministry of Education, with the mission to develop the evaluation of the education system. The INCE started its activities with a plan to evaluate in four years the new primary education, the new secondary education and the new professional training. The three projects use a similar methodology and have been developed in parallel by the INCE with the collaboration of regional administrations and private firms for field work.

The INCE Approach to Evaluation The purpose of the evaluation undertaken by the INCE is 'to obtain indicators of the overall functioning of the education system', in order to be able to 'provide information to the education agencies and to the citizens on the extent the system achieves the objectives' set by the authorizing legislation. The INCE sees itself as working for education agencies and the 'social actors' participating in the—so-called—'School Council'. It also refers to providing information that allows various stakeholders to form an 'opinion on the state of education' (INCE, 1994a: 4).

The object of the evaluation is rather ambitious, focusing on both *education outcomes* and the *implementation of the reform*. Education outcomes include academic performance and fundamental educational values emphasized by the legislation such as 'the development of the individual and social personality of the student, the education in the respect of rights and liberties, the exercise of tolerance and freedom, the respect of linguistic and cultural pluralism, the education for peace, cooperation and solidarity' (INCE, 1994b: 4, similar to ideas in INCE, 1995a, 1995b). The study of the reform implementation process aims to provide information to education agencies on critical factors of educational success by focusing on various elements such as teacher-training, human and material resources, curriculum development and education practice, attitudes and expectations of the various parties involved, student habits and school atmosphere.[7]

The reform attempts to change many aspects of schooling: curriculum, student population access to the new curriculum, management and pedagogical practices. Ideally, reformers would like to affect student outcomes, changing general characteristics of the schools and concrete educational practice at the classroom level. From this perspective, it may be judged as too ambitious and even bound to failure given the difficulty to change the basic teaching-learning process: 'the way teachers teach, the way students are expected to learn, and the ways knowledge is defined in schools' (Elmore, 1987: 61).

The basic instrument to collect data in the INCE evaluation is a macro-survey that is addressed to students, teachers, school management teams, education agencies and parents.[8] The questionnaires for each group are comprehensive and include the tests of educational achievement for students.[9] In some cases, qualitative techniques are also used to complete the tests (for example, on oral expression) or to assess more elusive aspects such as student attitudes or teachers' morale. Overall, it is a massive effort to reach the significant number of individuals involved as providers, managers or clients of educational services, to capture either how they have been affected by education processes or their perceptions.

Some aspects of the reform are amenable to straightforward measures: for example, the use of specific resources such as teaching training and support, new adapted textbooks or increased availability of subject area specialists at a younger age (previously only accessible at 14 years for students following an academic track—*bachillerato*). However, most of the substantive aspects the reform is supposed to affect are very difficult to grasp and measure—do teachers collaborate more among themselves? Are students more positively orientated? Does writing a curriculum project change the actual teaching in the classroom? The evaluation of these and other important changes requires high levels of inference from observations that are difficult to structure and standardize. Solely measuring educational results will be a challenge given the lack of agreement on how to define educational goals for the newly structured educational periods beyond legal definitions or when compared with educational goals of the main stages of the educational process before the reform.

Minimum Income and Emergency Aid

In May 1990, the Parliament of the Basque Country in the north of Spain passed a law introducing a new plan to fight poverty based on three programs:

- Along the lines of the French *revenue minimum d'insertion* and other European experiences with so-called 'social salaries', the Basque government would guarantee as 'a last resort', a minimum income—known as *ingreso mínimo de inserción* or IMI—to those who no longer benefited from unemployment, had insufficient pensions or were in any other circumstances that led them to experience financial difficulty.
- The regional administration, independently of the IMI, would pay for specific expenses of a basic nature (rent, utilities, essential furniture and debt incurred to pay those basic expenses) to residents in the three Basque provinces who could not afford them, in order to avoid social exclusion and family destruction. The program was known as *ayudas de emergencia social* or AES.
- Given the concern for dependency on public welfare, beneficiaries would sign an agreement—known as *convenio*—with social workers at the local level where both parties would stipulate the actions to be taken by the beneficiary to solve their problem in terms of training or job search. The plan did not allocate additional resources for employment and training programs but, rather, would foster cooperation with existing programs. This third pillar was fundamental in making possible the idea of an exchange of help for a commitment to be active in the search for a solution.

The plan was the response to the growing concern for new forms of poverty that emerged as a consequence of the severe economic crisis the Basque provinces suffered in 1975–85. In a traditionally rich region, the restructuring of heavy metallurgic and chemical industries meant hard times for many families who had been relatively prosperous.[10] The Parliament asked the regional government for an evaluation of the new programs before June 1992. The evaluation was conducted by the Labour and Social Security Department (DTSS) on the basis of management records and some specific studies contracted out with private consultants. The main instruments of the evaluation were a panel of beneficiaries that were monitored during two years (1990–2) and various surveys with social workers and participating administrations at the local level.

The Evaluation of DTSS According to the DTSS (1992), the main goal of the new plan was to avoid the destruction of personal, family and social life of those affected by the economic crisis. The basic concern regarding each of the three pillars of the new plan was: (1) to reach all potential beneficiaries, making sure every person who was entitled to IMI would receive it; (2) that people with urgent basic needs saw them satisfied through the AES; and (3) that instead of becoming welfare dependent, participants were actually activated to enroll for training in new professions or to directly search for jobs.

In order to determine the extent to which these goals were achieved, the evaluation produced a few, simple, numeric indicators that were reported for each of the three provinces. Fundamentally, these were data on coverage, cost and dependency:

(1) The proportion of homes that could potentially benefit from IMI was high. 17.9 percent of Basque families (11,323) had benefited until December 1991 from IMI. The IMI was fundamentally reaching unemployed without subsidies and single women that were either separated or divorced.

(2) A total of 9255 families had received AES by December 1991, assistance having been spent on housing (65 percent of total expenses), extreme personal need (24.9 percent) and debt (10.1 percent). AES beneficiaries were slightly different from IMI beneficiaries, including a more senior population with insufficient pensions. Many IMI beneficiaries (35 percent) did not qualify for AES benefits with important differences across the three provinces.

(3) Only 22.5 percent of the beneficiaries had signed agreements with local social workers and 45 percent of the social workers were reporting difficulties in finding programs to which beneficiaries could be referred. In both the number of agreements signed and availability of resources, the three Basque provinces presented important differences. As many as 24 percent of beneficiaries remained in the program, while 27.8 percent left the program after more than one year and 48.1 percent had had some contact with employment during the study period.

(4) However, few jobs were legal. The majority were temporary jobs, not declared to the employment and social security agencies as required by law. Those who received more training stayed longer in the program but remained economically independent after they found a job.

The evaluation also included aggregate information on budget growth and a report on management problems. For example, AES appeared to be growing faster than expected because it tended to be used instead, or as a complement to IMI. On the management side, social workers seemed to be overwhelmed by the new programs and felt that their jobs had changed to focus exclusively on the paperwork that IMI and AES required.

One final check on the overall effectiveness of the plan was made by comparing living conditions in the Basque provinces before and after the plan, using EUROSTAT data. European statistics showed an increase in absolute poverty and a decrease in relative poverty.

The information summarized was used by DTSS to present IMI and AES as a considerable success in terms of coverage, while recognizing the relative failure of the plan in helping beneficiaries to find their way back to work. The evaluation was also useful in signaling management problems and possible solutions. For example, the report included the demand from social workers not to have to modify payments each month according to changes in the earnings of beneficiaries as a way to reduce administrative burdens.

Employment and Training Programs

The Programs Employment and training programs in Spain can be administered and financed by central, regional or local governments. Actions are often jointly financed by the European Social Fund (ESF). When this is the case, programs and their evaluation follow European guidelines that, on the whole, have served to restructure and simplify the wide variety of actions undertaken by the various levels of governments. Nevertheless, training actions jointly financed by ESF allow for a considerable degree of freedom to accommodate programs to the local characteristics of labour markets.

In jointly financed actions, the Spanish Ministry of Labour and Social Security is the administration unit for the ESF in Spain and is responsible for the evaluation which can be contracted out according to a procedure that includes in the proposal to the ESF the organization that will actually undertake the evaluation. A team of economists of the General Foundation of the Autonomous University of Madrid (FGUAM) was chosen to evaluate training programs aimed at objectives 3 and 4 of the ESF—long-term unemployed and young people—for a standard year, 1992.

The Evaluation. The evaluation of FGUAM (1995) had two main objectives: (1) to determine effects on individual beneficiaries, and (2) to produce information that helped administrations to improve program management. With regard to the first objective, the researchers focused on one basic indicator—finding a job—and drew inferences from their knowledge of economic theories, with regard to effects of training in labour markets. They used a control group, built from employment office records, to compare job candidates who received and did not receive employment training. They also compared regional results. The second objective was pursued through a survey of regional administrations where they were asked their opinion on implementation issues.

(1) The evaluation found an association between employment and professional training: 45 percent of participants obtained jobs, half of those within six

months, and another 30 percent before 12 months. Approximately one-third were keeping their jobs when the surveys were conducted, 15 months after the training took place. Impact was higher among those who had work experience, male and young—under 25 years.

(2) There was a significant difference with the control group: the employment rate was 13.1 points higher among job candidates who received training. The differences were slightly higher among women than among men, and among those under 25 years compared with unemployed between 25 and 45 years. According to the final report: adjustments due to economic recession would have fallen more on young people had it not been for the program, 'as it is shown by the differential job procurement rates among participants and non-participants', and 'among young people and adults among the participants' (FGUAM, 1995: 16).

For the researchers, there was a substitution effect that favoured program participants when it comes to final employment. They theorize that training actions may have a positive effect on reducing the degree of dispersion in wages across production sectors and in reducing the job offers that did not find suitable job candidates.

(3) Concerning management issues, there was a perception in the field that decentralization had allowed for extensive participation of business associations and labour unions, with the overall effect of better adaptation to local conditions and needs.

The results of this evaluation are consistent with other studies, confirming that in a situation where the supply of people willing and able to take jobs exceeds the demand for these jobs, training and employment programs might affect *who gets* these jobs. If, however, as the evaluators suggest, the supply of people willing to take jobs exceeds the demand, but the jobs are not filled or they are filled with candidates that are not adequately prepared, then a training and employment program can have a significant social value; even more so in a labour market where a significant number of those unemployed lack the skills that employers are looking for.

Universities and Research

Basic Reforms The main pieces of legislation affecting universities, science and technology were passed during the first and second mandate of the Socialist government, in 1983—the University Reform Law—and in 1986—the Law of Science. The former changed the fundamental regulation of universities, giving them a considerable degree of autonomy from the administration. The latter created an Interministerial Commission for Science and Technology (CICYT) in order to coordinate resources and programs through a National R & D Plan that is renewed every four years after a process of consultation with various constituencies and in accordance with European Union framework programs.

Evaluation Instruments Innovation processes in higher education and science and technology have resulted in various forms of evaluation. In the case of the universities, three approaches have been developed:

(1) The main instrument has been a form of *individual evaluation* that allows university professors to submit reports of their research activity in periods of six years to a national commission with the possibility, if the evaluation is positive, to be granted a monthly bonus that is added to their salaries. The national commission in charge of this procedure aggregated the results of individual evaluations in order to identify the first rank of Spanish research institutions by discipline.
(2) The Council of Universities (Consejo de Universidades, 1995) promotes the *institutional evaluation* of universities financing part of their cost. The institutional evaluation includes self-evaluation and an external evaluation. It has to be completed in one year, and must cover teaching, research and administrative-support services in relation to the unit of analysis which can be one or more degrees or one or more departments. The Council has developed a guide and forms to collect information and opinions on specific issues.[11].
(3) Some universities proposed CICYT, the central coordinating agency for science and technology, to be evaluated by ANEP, the administrative unit created to coordinate the ex ante peer review evaluation of research proposals. ANEP developed a methodology to evaluate the research activity of university departments on the basis of a simple, yet comprehensive system of indicators that is applied to a standard year. As shown in Table 2, the system is based on two types of information collected by a panel of experts.

For the evaluation of the science and technology programs financed by the National Plan, CICYT publishes, on an annual basis, an extensive report that concentrates almost exclusively on input–output indicators for the whole country in comparison to other industrialized nations. There have been efforts to discuss and publicize research findings, but within the strict limits of committees formed by area experts. On the technological side, the Centre for Industrial Technology Development (CDTI), within the Ministry of Industry, has used surveys as a way to gauge private sector perceptions on public policies. The work of ANEP, rating scientific projects and institutions, can also be regarded as an implicit form of ex post evaluation because there is a strong reliance on the record of the research team.[12]

Health Primary Care

The health system in Spain is fundamentally based on public provision and universal coverage. At the same time, there is an important private sector that is managed by private insurance companies. Public health can be administered either by a central government agency or by regional agencies in self-governing regions to which health services were transferred. The case of Catalonia is interesting because the region had an important tradition of local health plans and community hospitals which provided an excellent basis for shifting an important part of the publicly controlled provision of health services to private providers. Partly to acomplish this, the Catalan government created an agency, the SCS (Catalan Health Service), that can either contract with private hospitals and health providers or with the public agency that inherited the public health system in the region.

Table 2. Indicators for Evaluation of Research in University Departments, Based on An Evaluation Report for Universidad Autónoma de Barcelona from Spanish Agency for Evaluation and Perspective

Quantitative indicators		*Qualitative indicators*
Scientific resources	(1) Human resource structure (2) External funds by source (3) Average external funds per researcher (4) Average cost of one publication	Potential resources Overview of collaborations with other institutions, research meetings, mobility of personnel
Scientific production	(1) Published papers in scientific journals classified in four groups of decreasing impact (2) Books and chapters in collective books (3) Doctoral dissertations (4) Patents, European/US patent offices	Score for each publication according to the scale of the group in which the journal or publisher is classified
Scientific performance	(1) Average quality (total score after classification of publications) (2) Average scientific productivity (relation total score and number of researchers) (3) Gross productivity (average number of articles per researcher) (4) Relative participation of the research personnel in publications	Lines of research actually developed, according to publications

The Management Control System With government contracting out and creating the conditions for competition between public and private providers, accountability became a key issue. This was required to remain closely associated with the agency that controls public funds. With regard to health in primary health care, SCS developed a system of indicators that would help them to draft and monitor contracts. Indicators are divided into various groups that refer to the main areas negotiated with providers. Table 3 includes a few examples. The system is used to monitor outputs, population served and cost. It is also used to monitor the implementation of the specific policies that the funding agency is promoting in order to improve the quality of services. Data collected through this system can be used to take corrective action, to redistribute resources or to plan alternatives in dealing with basic health problems.

The Contrast: Practice Versus Theory

One way to capture the overall characteristics of evaluation in the five policy areas is by filling out the table that was used to present the basic theoretical models. As can be seen in Table 4, the Spanish practice of evaluation in these five policy areas does not correspond to one theoretical model, but takes elements from various models. At the same time, some of the elements that are fundamental for some theorists (e.g. true experiments, analysis of 'UTOS' (see Cronbach, Table 1), case studies and naturalistic approaches) are hardly found.

Only employment and training programs seem to follow the traditional quasi-experimental approach to assessing effects, that could be associated with Campbell or Rossi in Table 1. In the education cases—primary and secondary education, and higher education—there is an effort to produce information that helps decision-makers and stakeholders to form an opinion on the state of the education system. The 'spirit' of the evaluation is closer to Weiss than to any other author. The aim is to 'enlighten' and clarify questions, without reaching the point of mapping out the processes and

Table 3. Indicators for Health Service Contracts. Based on a list of indicators from Catalan Health Services Administrative Offices

Population covered	(1) Implementation of a system to identify clients in a standard year (2) Population attended in relation to assigned client population by categories of clients in a standard year (3) Dental review (primary education students) (4) Hypertension diagnosed/under control (5) Patients with cholesterolemia, diagnosed/under control (6) Patients with glycemia diagnosed/under control (7) Patients with *diabetes mellitus* diagnosed/under control (8) Smokers (9) Excessive alcohol drinkers
Use of records: population covered	(1) Individual records for primary care (in relation to assigned population) (2) Vaccination (3) Arterial tension (in relation to adult population) (4) Cholesterolemia (in relation to target population) (5) Glycemia (in relation to target population) (6) Smoking (in relation to adult population) (7) Excessive alcohol drinkers (in relation to adult population)
Availability of protocols	(1) High blood pressure (2) Non-smoking (3) Follow-up of a healthy child (4) Dental care
Client oriented systems	(1) Implementation of a system to program visits to other centers/specialists (2) Implementation of a system to take care of the paper work for visits to other centers/specialists (3) Implementation of a system to collect client complaints or comments
Health care activity	(1) Number of visits per inhabitant per day (2) Number of visits by category (general/pediatrics/dental care, nurse-general/nurse-pediatrics, social worker) (3) Use of laboratories/radiology/referrals to other centers/specialists (4) Prescriptions (number, cost per inhabitant, cost per prescription, relative use of certain categories)

Table 4. Evaluation in Five Policy Areas

	Elementary and secondary education	*Minimum income and emergency aid*	*Employment and training programs*	*Universities, science and technology*	*Health primary care*
Concept	Clarify questions, 'enlighten' decision-makers and stakeholders	Improve accountability with authorizing environment	Improve accountability with authorizing environment	Clarify questions, 'enlighten' decision-makers and stakeholders	Management control
Prioritized interest	Ministry of Education and Science, educational community, citizens	Legislators, Basque Department of Labor and Social Security	European Social Fund, Ministry of Labor, Regional administrations	Ministry of Education and Science, university chancellors, citizens	Health Department
Role of evaluator	Analyst, methods expert	Analyst, methods expert	Methods expert	Analyst, expert reviewer	Service and financial analyst
Object	Effects, processes, context	Coverage, cost, effects	Effects, implementation	Activities, performance, processes	Activities, cost, procedures
Source of criteria to determine value	Legislation	Legislation	Program goals and European guidelines	International practice, literature review	Legislation, contracts with providers
Methods	Survey, achievement tests, performance indicators	Survey, performance indicators	Quasi-experimental approach, survey	Performance indicators, expert panels, peer-review, surveys	Performance indicators
Perspectives on utilization	Conceptual, policy formulation	Policy formulation	Instrumental (funding)	Conceptual and instrumental (funding)	Instrumental (funding)
When should evaluation be done?	Periodically, as part of central management	Before new authorization	Periodically, as part of funding process	Periodically, as part of funding process	Periodically, as part of management
How to organize?	Internal, involving stakeholders	Internal, involving stakeholders	External, involving stakeholders	Internal and external	Internal

establishing links, as would be prescribed in a theory-driven model. In the regional cases—minimum income in the Basque provinces and primary health care in Catalonia—there is a managerial approach to evaluation that would be represented by Wholey in Table 1. Coverage-output measures are used as a proxy for performance, but evaluation does not seek to analyse causality.

Making Sense of Evaluation Practice

First, we can conclude that the evaluations portrayed in this paper are not academically orthodox, especially when they are examined from the traditional evaluation perspective that places so much emphasis on net impact and effectiveness analysis. If the purpose is to produce a powerful analysis of effects, these evaluations may not be very useful. At the same time, in each of the five evaluations, one can find pieces of information that are clearly valuable not only to administration managers, but also to commissioning authorities and stakeholders. Table 5 summarizes the potential use of each of the five evaluations.

There is a pattern in Table 5. On the one hand, except for the evaluation of employment and training programs which focusses on placements, evaluations attempt to produce various pieces of information that help managers and stakeholders to judge better their policies and programs. This kind of information is adequate to generate a public debate on a policy, to inform quality judgments and to have an idea of what is going on. The conclusions users may reach are of the type: 'this university department has been doing very well lately', 'emergency aid is out of control in two provinces, and it is used as a substitute for minimum income', 'parents do not seem happy about doing less academic courses'. They tend to detect problems or express effectiveness and efficiency in relative terms. On the other hand, evaluations appear to find it difficult to trace the relationship between public intervention and observed effects in participants or in the situation that they try to ameliorate.

This could simply be explained by the technical difficulty in pursuing that kind of analysis. However, government managers and evaluators face a trade-off between dedicating their work and effort to the accurate analysis of effectiveness, or producing a set of measures on cost, coverage, management problems and performance. A set of measures looking at different aspects of performance can send a powerful message about the relative success of a program or organization, and this seems to be the option that is preferred. From this perspective, public organizations are closer to the forms of accountability used by private organizations, with the difference that they need much more creativity and effort to develop a coherent and comprehensive set of measures.[16]

The trade-off is important in terms of the adaptability of evaluation efforts to the way it is conceptualized by program managers and stakeholders. As can be seen in Table 4, in Spain, (1) evaluation is conceptualized with a clear aim to *produce information that is useful to various social actors* and administrations. The first client is the government institution in charge of the program, but evaluation efforts attempt to respond to the concerns of overseers and stakeholders; (2) there is consensus that *evaluation results will only be reliable over time*, and that systems of information need to be settled and adjusted. In this context, trends, signals and perceptions are given more attention than

Table 5. Potential Use of Evaluations in Five Policy Areas

	Evaluation is not useful to determine	*Evaluation can be useful to find out about*
Elementary and secondary education	Whether the student–teacher interaction has changed The extent to which students internalize new values Whether differences of educational results can be attributed to curriculum variations and decentralization The relative success of the reform with respect to previous programs[13]	Dubious elements of the reform, as viewed by education community and social actors Territorial variations in implementation Curriculum, teaching and context elements influencing quality[14] Sources of problems
Minimum income and emergency aid	Whether the program has made a difference The utility of working agreements The behavioral response to public aid	Whether those who need help are getting it Variations among the three provinces Quality of jobs for welfare recipients Trends in expenditure Sources of waste, fraud and abuse Management problems
Employment and training programs	Whether the program has made a difference[15] The extent to which programs adapt to local differences The key components for employment and training programs to place participants	Reasonable levels of employment for participants Time required to find employment Characteristics of participants more likely to find jobs
Universities and research	Changes in learning processes Relative effectiveness of differences in the curriculum The performance of administrative services The returns to society of research	Research quality Scientific productivity Good performers, area leaders Institutional performance in broad terms
Health primary care	The relation between changes in the process of delivering medical services and health status (Brook and McGlynn, 1991)	Demand for medical services The level of compliance with forms that guarantee a minimum of quality Relative efficiency of providers

exact figures; and (3) evaluations tend to be rather ambitious in terms of gathering data about a wide array of questions in an effort to *grasp every angle of a reality* that tends to have many different faces.

The way evaluation is approached, combining the development of systems of indicators with surveys and narrative accounts that can take the form of a self-evaluation or the simple aggregation of a whole set of observations and measures, fits well this concept of evaluation. Certainly, it needs to be improved to become a good functional equivalent of financial analysis in the private sector, but in producing the kind and the volume of information that is valued by government decision-makers, commissioners and stakeholders, increases the likelihood of effective use. For evaluation results to be used, evaluation projects need to adapt. In Spain, where the demand for the use of experimental approaches is much less intense, it was found that evaluation approaches that are closer to the production of management control indicators or analysis of perceptions are a better way to inform decision-making in the policy process.

Conclusions

Evaluation is conceptualized at a national or regional level in Spain with the aim to produce information that is useful to various social actors and administrations. The effort to collect data from stakeholders does not translate into organizational arrangements that actively involve stakeholders in the use of results. The purpose of evaluation is closer to the idea of providing information to clarify issues, informing quality judgments and having a perception of what is going on, than to making a summative judgment. Therefore, it is not expected that evaluation results can determine the fate of a program in the short-term. Evaluation results need to be consistent over time and to cover as many issues as possible to give a complete picture of the way an organization or a program is functioning.

The way evaluation is conducted, putting the emphasis on producing data on various issues—costs, coverage, management problems, activities, perceptions on effectiveness—and using a combination of methods and approaches—management indicators, surveys, interviews, narrative accounts—fits well with the concept of evaluation and the expectations of its results. It is the government response to the legislative imperative to evaluate. And it is pursued within the constraints that lack of data, time and technical expertise impose on evaluation projects. The end-result brings evaluation closer to the private sector model of aggregating information to have a better idea of effectiveness and efficiency.

From this perspective, program evaluation can be, as Moore (1995) says, the equivalent in the public sector of profitability in the private sector. Both focus on the past, make it possible to hold organizations and their managers accountable for what they do and, indeed, develop information. These approaches develop understandings about past performance in ways that do not generally examine causality between public interventions and social needs.

Notes

1. The literature on institutionalization of evaluation through government reform efforts has been one way to look at evaluation activity in Europe (Rist, 1990). In Spain, except for a proposal to reorganize the central government in modules that would be controlled by results (MAP, 1990), there has not been a series of reforms such as those in the UK or France. For a recent account on the UK, see Pollitt (1993). For France, see Duran et al. (1995).

2. A good example is the discussion of the Massachusetts Employment and Training Program (ET Choices) by Behn (1991).
3. Examples are from Spain, a country that typically reproduces trends experienced by other western democracies, a few years later but very intensively. Evaluation was introduced in the 1990s, which allows learning from the experience of other countries, while adapting European practices.
4. Table 1 is based on my own readings and on the interpretation of the work of the seven authors by Shadish et al. (1991). The nine dimensions do not appear as such in their work.
5. Again, there is a risk in including some names but not others. There is also a risk in relating some authors to one of the seven. Furthermore, reading the columns, one can easily see the influences that the seven have among themselves. Almost all the authors are North American despite some interesting contributions from Europe (for a general approach, see Hellstern, 1986; on performance measurement, see Carter et al., 1995). However, European authors seem to either conduct applied research in one field (e.g. training and employment programs) or diffuse theories that were first presented in the American context.
6. The former eight grades of what was known as General Basic Education were reduced to six in the new Primary Education, while all the students had to do four more years of Secondary Education, until they were able to take either *bachillerato* leading to higher education or professional training. The change implies that high schools would now have younger children (12 years old versus 14 with the older system) and that school would be obligatory for them until the age of 16 years. The new legislation also mandated the public provision of pre-school from age three years and regulated different forms of special and non-regular schooling (INCE, 1994b, 1995a).
7. In the case of primary education, the study focusses on second graders of the new system and sixth graders of the old system. This choice is justified on the basis that second graders started their formal education with the new system, while the study of sixth graders can provide a reference point that can be later used to compare the educational results achieved by the new primary education. In the case of secondary education and professional training, the study focusses on students of the last year of those periods (INCE, 1994b, 1995a, 1995b).
8. The survey is stratified by education center and student group. It is representative at the regional level (INCE, 1994b, 1995a, 1995b).
9. The evaluation of academic performance is adapted to different stages of educational development: second graders are tested on basic cognitive skills—communication and reasoning in language and mathematics—while sixth graders and secondary education students are subject to more specific tests of wider subject matter: language and social sciences, maths and natural sciences. Questionnaires also differ in terms of the fundamental values they stress: in the case of secondary education, there is a concern for student attitudes towards studying or for perceptions on the connection between what is studied and real world problems. In the case of professional training, several items refer to students' perceptions on whether they are prepared to work. All students are also asked about studying methods, teaching practice or school atmosphere. Parents are asked about studying methods, school atmosphere, school management and participation. The questionnaires for teachers include other issues such as resources, teaching practice, training and expectations on educational results at each level. Questionnaires for school management teams and administrative agencies are similar to those for teachers. In the case of professional training, the survey is extended to 'social agents' such as union and business leaders (INCE, 1994b, 1995a, 1995b).

10. Industrial investment in the Basque country, in relation to total figures for Spain, went from 13.7 percent in 1975 to 4.7 percent in 1983. Unemployment rose to 22.5 percent in 1984. The province of Vizcaya, where Bilbao is, was second in income per capita in 1972, and only the twenty-first in 1985. A study of the Basque government for 1986 estimated that 133,700 (29.5 percent) of the number of families were affected and saw a significant reduction in their income having a direct effect on their consumption capacity. However, 54,100 (9 percent) lacked basic material goods and had trouble keeping their housing. The study also estimated that 0.5 percent of families were living in a miserable condition (DTSS, 1992).
11. For example, on teaching, universities must collect information and external evaluators must review: curriculum requirements, student evaluations, characteristics of the student body and level of satisfaction, characteristics of the faculty and selection procedures, support personnel and tasks, material resources (classroom space, availability of computers for students, libraries, laboratories), student success rate, drop-out rate, average time to finish a specific degree (Consejo de Universidades, 1995: 26–7).
12. For a more extensive description of these forms of evaluation, see Luís Sanz-Menéndez (1995).
13. The evaluation does not take full advantage of a unique opportunity to make a comparison between students that attend reformed and non-reformed schools: for some time—during the time it takes regions to make the new policy universal—only a fraction of the student population—some students in a school, some schools in a locality, or some schools in the region—will participate in the new progam. Indeed, it is an ideal situation because there is no political or ethical problem: services are not denied, it is the lack of resources and organizational capacity that prevents the administration fully implementing the policy at a given moment.
14. The INCE evaluation addresses some of the central features of the schooling context, as it has been defined in the literature (Oakes, 1989; Porter, 1991), which can be critical in understanding what factors influence quality.
15. The evaluation of employment and training programs has a weakness in the way the control group was constructed. The evaluators used a sample of job candidates who did not participate in any training program with a similar composition of sex, age and education level. However, if the participants found more jobs, it could certainly be because they were more motivated, had better interpersonal skills or were more employable in terms of their talents or past experience.
16. Utilization is not guaranteed because Spanish government institutions have failed to establish institutional links that facilitate the feedback of evaluation results to overseers and stakeholders. Collecting data from stakeholders is not accompanied by a strategy of dissemination of results and involvement of stakeholders.This is clear in the education cases. Only in primary health care and employment and training programs could evaluation data be directly used for control purposes, while in the Basque program of minimum income and emergency aid, utilization of evaluation results did not necessarily follow the debate in the Parliament.

References

Behn, R. (1991) *Leadership Counts*. Cambridge, MA: Harvard University Press.

Brook, R.and E. McGlynn (1991) ‘Maintaining Quality of Care’, in E. Ginzberg (ed.) *Health Services Research.* Cambridge, MA: Harvard University Press.

Bryk, A. (ed.) (1983) *Stakeholder-Based Evaluation*. San Francisco, CA: Jossey-Bass.

Campbell, D. (1969) 'Reforms as Experiments', *American Psychologist* 24: 409–29.
Campbell, D. and R. Boruch (1975) 'Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects', in C. A. Bennett and A. A. Lumsdaine (eds) *Evaluation and Experiments: Some Critical Issues in Assessing Social Programs*, pp. 195–296. New York: Academic Press.
Campbell, D. and J. Stanley (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.
Carter, N., R. Klein and P. Day (1995) *How Organizations Measure Success. The Use of Performance Indicators in Government*. London and New York: Routledge.
Chelimsky, E. (1987) 'The Politics of Program Evaluation', in D. Cordray, H. S. Bloom and R. Light (eds) *Evaluation Practice in Review*. San Francisco, CA: Jossey-Bass.
Chen, H. and P. Rossi (1983) 'Evaluating with Sense: The Theory-Driven Approach', *Evaluation Review* 7: 283–302.
Consejo de Universidades (1995) *Programa de Evaluación Institucional de la Calidad de las Universidades*. Madrid: Consejo de Universidads.
Cook, T. and D. Campbell (1979) *Quasiexperimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
Cronbach, L. (1982) *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey-Bass.
Cronbach, L., S. Ambron, S. Dornbusch, R. Hess, R. Hornik, D. Phillips, D. Walker and S. Weiner (1980) *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass.
Departamento de Trabajo y Seguridad Social (1992) *Informe de Evaluación del Plan Integral de Lucha contra la Pobreza en Euskadi* (1988–92). Vitoria: Publicaciones del Gobierno Vasco.
Dilulio, J., Jr (1994) *Measuring Performance When There is No Bottom Line*. Center for Public Management Working Paper 94–2. Washington, DC: The Brookings Institution.
Duran, P., E. Monnier and A. Smith (1995) 'Evaluation *à la française*', *Evaluation* 1: 45–63.
Elmore, R. (1987) 'Reform in the Culture of Authority in Schools', *Educational Administration Quarterly* 23: 4.
Fundación General Universidad Autónoma de Madrid (1995) *Ex-Post Evaluation of Training Actions Co-Financed by the European Social Fund.* Madrid.
Guba, E. and Y. Lincoln (1989) *Fourth Generation Evaluation*. Beverly Hills, CA: Sage.
Hellstern, G. (1986) 'Assessing Evaluation Research', in F. Kaufman, G. Majone and V. Ostrom (eds) *Guidance, Control and Evaluation in the Public Sector*. Berlin: De Gruyter.
Horst, P., J. Nay, J. Scanlon and J. Wholey (1974) 'Program Management and the Federal Evaluator', *Public Administration Review* 34: 300–308.
House, E. (1980) *Evaluating with Validity*. Beverly Hills, CA: Sage.
Instituto Nacional de Calidad y Evaluación (1994a) *Documento-Marco de Funcionamiento del Instituto Nacional de Calidad y Evaluación*. Madrid: Ministerio de Educación y Ciencia.
Instituto Nacional de Calidad y Evaluación (1994b) *Proyecto de Evaluación de la Educación Primaria.* Madrid: Ministerio de Educación y Ciencia.
Instituto Nacional de Calidad y Evaluación (1995a) *Evaluación Piloto de la Educación Secundaria Obligatoria*. Madrid: Ministerio de Educación y Ciencia.
Instituto Nacional de Calidad y Evaluación (1995b) *Proyecto de Evaluación de la Formación Profesional*. Madrid: Ministerio de Educación y Ciencia.
Lindblom, C. (1990) *Inquiry and Change: The Troubled Attempt to Understand and Shape Policy*. New Haven, CT: Yale University Press.
Lynn, L. E. (1989) 'Policy Analysis in the Bureaucracy: How New? How Effective?', *Journal of Policy Analysis and Management* 8: 3.

Majone, G. (1988) 'Policy Analysis and Public Deliberation', in R. Reich (ed.) *The Power of Public Ideas*. Cambridge, MA: Ballinger.

Ministerio de Administraciones Públicas (1990) *Reflexiones para la Modernización del Estado*. Madrid: Ministerio de Administraciones Públicas.

Moore, M. (1995) *Creating Public Value*. Cambridge, MA: Harvard University Press.

Oakes, J. (1989) 'What Educational Indicators? The Case for Assessing the School Context', *Educational Evaluation and Policy Analysis* 11(2): 181–99.

Palumbo, D. (1987) *The Politics of Program Evaluation*. Beverly Hills, CA: Sage.

Patton, M. (1980) *Qualitative Evaluation Methods*. Beverly Hills, CA: Sage.

Pollitt, C. (1993) 'Occasional Excursions: A Brief History of Policy Evaluation in the UK', *Parliamentary Affairs* 45(3): 353–62.

Porter, A. (1991) 'Creating a System of School Process Indicators', *Educational Evaluation and Policy Analysis* 13(1): 13–29.

Rist, R. (ed.) (1990) *Program Evaluation and the Management of Government*. New Brunswick, NJ: Transaction Publishers.

Rossi, P. and H. Freeman (1989) *Evaluation: A Systematic Approach*. Beverly Hills, CA: Sage.

Sanz-Menéndez, L. (1995) 'Research Actors and the State: Research Evaluation and Evaluation of Science and Technology Policies in Spain', *Research Evaluation* 5(1): 2–10.

Scriven, M. (1972) 'The Methodology of Evaluation', in C. H. Weiss (ed.) *Evaluating Action Programs: Readings in Social Action and Education*, pp. 123–36. Boston, MA: Allyn & Bacon.

Scriven, M. (1980) *The Logic of Evaluation*. Inverness, CA: Edgepress.

Shadish, W. and R. Epstein (1987) 'Patterns of Program Evaluation Practice. Among Members of the Evaluation Research Society and Evaluation Network', *Evaluation Review* 11: 555.

Shadish, W., T. Cook and L. Leviton (1991) *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.

Stake, R. (1975) 'To Evaluate an Arts Program', in R. E. Stake (ed.) *Evaluating the Arts in Education: A Responsive Approach*, pp. 13–31. Columbus, OH: Merrill.

Stake, R. (1978) 'The Case Study Method in Social Inquiry', *Educational Researcher* 7: 5–81.

Stake, R. (1980) 'Program Evaluation, Particularly Responsive Evaluation', in W. Dockrell and D. Hamilton (eds) *Rethinking Educational Research*, 72–87. London: Hodder & Stoughton.

Weiss, C. (1972) *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.

Weiss, C. (1973a) 'The Politics of Impact Measurement', *Policy Studies Journal* 1: 179–83.

Weiss, C. (1973b) 'Where Politics and Evaluation Research Meet', *Evaluation* 1: 37–45.

Weiss, C. (1983a) 'Toward the Future of Stakeholder Approaches in Evaluaton', in A. S. Bryk (ed.) *Stakeholder-Based Evaluation*. San Francisco, CA: Jossey-Bass.

Weiss, C. (1983b) 'Ideology, Interest, and Information: The Basis of Policy Decisions', in D. Callahan and B. Jennings (eds) *Ethics, the Social Sciences, and Policy Analysis*, pp. 213–45. New York: Plenum.

Weiss, C. (1987) 'Evaluating Social Programs: What Have We Learned?', *Society* 25: 40–5.

Weiss, C. and M. Bucuvalas (1980) *Social Science Research and Decision-Making*. New York: Columbia University Press.

Wholey, J. (1983) *Evaluation and Effective Public Management*. Boston, MA: Little Brown.

Wholey, J., M. Abramson and C. Bellavita (eds) (1986) *Performance and Credibility. Developing Excellence in Public and Nonprofit Organizations*. Lexington, MA: Lexington.

XAVIER BALLART is 'Profesor Titular' of Political and Administrative Sciences at the UAB in Barcelona. He was a visiting researcher at the Kennedy School of Government, Harvard University, with the support of a grant from the Spanish Ministry of Education and Science, during the academic year 1995–6.