CHAPTER 15
PERFORMANCE MEASUREMENT AND EVALUATION

This chapter deals with performance measurement and monitoring, and programme evaluation. These instruments and techniques have common purposes: improved programme management, increased accountability and better decision-making, as they feed back information on the outcomes and outputs of existing government policies and programmes in order to improve the design and implementation of such programmes in the present and the future.

A. Performance Measurement

1. What is performance measurement?

Performance measurement is an instrument for assessing progress against stated programme goals and objectives, assuming that the strategic objectives are known. It consists of the following activities:

- Documenting the “production process”, which consists of processes and activities used to turn inputs, which are the resources used by the programme, into outputs, which are the goods and services directly produced by the programme.

- Assessing the outcomes — the broader economic or social changes resulting from a policy or programme — and comparing them with the programme objectives.

Performance measurement may indicate in general terms the result of a policy measure or programme, but does not analyse why it has occurred or what changes may need to be made to activities or programme objectives. For this purpose, an in-depth assessment of the programme is needed. Programme evaluation, which is reviewed in Section B, extends beyond the tracking and monitoring of performance measures into an examination of the ways in which outcomes are affected by the programme concerned. Whilst performance measurement focuses on efficiency and effectiveness, evaluation covers in addition issues such as utility, relevance and the sustainability of the programmes concerned.

More specifically, it is often said that performance measurement covers the following five dimensions of performance (OECD, 1994):

- Efficiency, which is the relationship between the goods and services produced by a programme or an activity (outputs) and the resources used to produce them (inputs), and is measured by the cost per unit of output. Efficiency must be assessed against some benchmark, e.g. the unit cost of the activity in a previous period or the unit cost of carrying out a similar activity in another agency or establishment.
Effectiveness, which is the extent to which programmes achieve their expected objectives, or outcomes. Effectiveness is the most important element of value for money in the public sector. Goods or services may be provided efficiently but if they do not achieve their intended objectives, and give satisfaction to the users of public services, the resources used will be largely wasted.

Economy, may be defined as “the acquisition of the appropriate quality and quantity of financial, human and physical resources at appropriate times and at the lowest cost concerned”, and may be assessed through input measures, and comparisons with norms and standards.

Compliance. Agencies must comply with the budget or appropriation act and other laws/regulations e.g. in relation to the management of cash flows and timely payment of creditors. A tax collection agency, for example, may have specific performance targets (such as the amount of tax arrears collected). Such measures of financial performance are sometimes more related to efficiency than to compliance.

Service quality. In its broader sense, service quality refers to effectiveness. However, it is generally used in a narrower sense relating to the more immediate needs of users, such as timeliness, accessibility, reliability and continuity of services. As such, it refers to the quality of service delivery rather than of service outcomes. Development of a responsive client/consumer-oriented culture in public service delivery is on the reform agenda of most OECD countries and many others.

“Performance” is an amalgam of all these dimensions. Some individual dimensions interact and may conflict with each other. For example, it may be possible to improve service quality or compliance but only at higher cost and lower efficiency.

2. Performance measures and indicators

a. Types of measure or indicator

Performance can be measured through measures or indicators. Measures correspond to direct records of inputs, outputs and outcomes. Indicators are used as a proxy when direct measures are difficult or costly to obtain (e.g. the “street” price of illegal drugs as a measure of the outcomes of an anti-drug programme). In practice, as well as in this chapter, the terms “measures” and “indicators” are used interchangeably.

The categories of performance measures that support the assessment of the dimensions of performance enumerated above are as follows:

• Inputs. Measures or indicators of inputs concern the use of personnel, equipment, materials, etc. Inputs are usually expressed as the amount of expenditure or staff time. Measures of inputs concern the economy with which resources are used to deliver outputs and outcomes.

When expressed as a ratio of outputs and outcomes, input indicators are used to measure efficiency and cost-effectiveness. Inputs should include both current expenditures and the use of capital goods and, ideally, costs should be estimated on an accrual basis. For example, for a road maintenance programme a cost-effectiveness analysis should take into account the depreciation of equipment, which accounts for a significant share of the full costs. However, although the operational cash costs do not measure the full costs of a programme, generally the trend in such costs does not differ much from the full costs of the programme (Premchand, 1993).
Performance Measurement and Evaluation

• **Outputs.** Outputs refer to the goods and services produced by a programme or an activity (e.g. kilometres of road build, number of children vaccinated, etc.). Output indicators are used to assess efficiency. Efficiency can be measured by the ratios of inputs to outputs often expressed as the number of employees or amount of employees’ time per unit of output, and referred to as *unit cost* (e.g. the number of days expended per repair made, or the cost per kilometre of roads that were repaired to a satisfactory condition). **Productivity** is usually measured as the ratio of the amount of output to input, e.g. the number of prisoners transported divided by the cost of transportation. (Sometimes, however, productivity may refer to the ratio of outcomes to inputs). Workload or activity level measures are often used as a proxy for output measures (e.g. the number of inspections carried out).

• **Outcomes.** Outcomes correspond to the ultimate policy purpose, or the desired ends of a policy, that are achieved by producing the outputs (improved accessibility of remote areas, the reduction in the number of cases of a particular disease, etc.). Measures of outcomes concern *effectiveness.*

• **Intermediate outcomes** are expected to lead to the ends desired, but are not in themselves ends. In many programmes a progression or sequence of intermediate outcomes occurs. For example, in the case of an environmental programme the sequence of intermediate outcomes could be as follows: law passed; number of businesses that change their behaviour; reduction of hazardous wastes and pollutant counts. There are various terminologies that seek to capture similar distinctions. For example, the term “result” can be used in a similar sense to “intermediate outcome”, while “impact” may be used to describe results and (end) outcomes collectively. A recent guide published by the European Commission uses the term “impact” to describe the effects of a programme on society, and refers to “initial impact” as “results” and “long-term impacts” as “outcomes”.

Concerns about the quality of public service delivery are measured by indicators of *customer satisfaction* (e.g. the number of complaints received, surveys, and participative processes). **Service (delivery) quality** indicators measure the timeliness, accessibility, reliability and accuracy of services (e.g. police response time, compliance with transport timetables, hospital waiting times, etc.). Service quality often depends on processes.

**Work process measures** are indicators of the way work gets done in producing the output at a given level of resources, efficiency and effectiveness. Processes consist of a chain of activities or work practices, such as procurement procedures, technological processes for producing goods and peer reviews for policy formulation.

**Process indicators** help in evaluating performance in areas where outputs or outcomes are difficult to measure and have also by themselves an independent value, notably for assessing the quality of public service delivery. In certain areas, where output indicators are not very meaningful and outcome indicators difficult to measure, performance is sometimes assessed through process measures. For example, peer reviews can be used to evaluate the process of providing policy advice to ministers. In some areas of public activity, such as law or politics, “due process” is a key element of good governance.

As indicated earlier, the **impact** of a policy measure or programme is often synonymous with its outcome. Sometimes, however, the term impact has a more precise definition. Sometimes a distinction is made between **gross outcomes** and **net impacts.** The net impacts are the outcomes truly attributable to the programme. They do not include the effects of factors external to the programme, and are estimated through the evaluation methods reviewed in Section B. Impact analysis refers to the assessment of the
Managing Public Expenditure - A Reference Book for Transition Countries

effects of an intervention on its surroundings (e.g. an environmental impact study), or shows the extent to which a programme actually produced certain effects on client populations. Sometimes, impact measures refer to how the outcomes of a particular programme affect other programmes or an organisation’s mission.

Social indicators may be used to assess the broad impact of certain government policies. They consist of measures at a highly aggregated level, such as infant mortality rates and adult literacy rates. In practice, it is not easy to use social indicators to assess the performance of a particular programme. There may be problems, for example, in determining causality (e.g. reductions of infant mortality may be due to a clean water programme, an immunisation programme, a nutritional programme, or to external factors that are not attributable to government policies). Programme outcome indicators are designed to focus on more detailed aspects of performance. Social indicators nevertheless provide useful background information for policy decision-making.

Performance measures may be either quantitative or qualitative. Qualitative measures can be transformed into quantitative ones by surveys, report cards, and other techniques of assessing the opinion of users. For example, the quality of education can be in part quantified by measuring the percentage of parents who are “fully satisfied” with their children’s school.

The ultimate purpose of the government’s programmes is to produce outcomes. However, defining outcomes and developing outcome measures can present difficulties. Sometimes, outcomes occur only after many years. Often, a programme is only one of many influences on an outcome. Attribution, which consists of determining how much of the outcome is truly attributable to the programme rather to other influences, is a challenging task. Compared to output indicators, outcome indicators are more relevant in assessing the achievements of programmes, but output indicators are generally easier to define and measure.

In some sectors, output measures can be used as a surrogate outcome indicator. For example, for a road construction programme, the number of kilometres built, which is the output of the programme, can be used for assessing its effectiveness. Nevertheless, in other sectors, notably the social sectors, outcomes can sometimes be so remote from outputs that the latter are not reliable indicators of the former. For example, an increased number of medical visits does not necessarily imply reduced illness. To provide useful feedback to decision-makers intermediate outcomes (i.e. results) should also be measured.

b. Performance as a relative concept

Performance is only a relative concept. By definition, assessing effectiveness requires comparing measures of outcome or output to the programme objectives. In practice, to assess whether results are good, bad or indifferent, every performance measure should be compared against some base or standard. Thus, performance is often measured against:

• What has been achieved in the past. Time series statistics are very useful, but do not take into account changes in efficiency or productivity due to technological factors.

• What other comparable programmes or organisations are achieving, or national/international standards in the field. The activities of other organisations provide useful benchmarks. The problem here is to find a strictly comparable organisation.3

• Targets set in the budget or other policy statements by the government.
Comparisons should be made only on a like-with-like basis. This requires defining properly the indicator and the basis of comparison. For example, in comparing the performance in examination results of different high schools, it may be appropriate to correct the gross measures (e.g. the ratio of exams passed per student) by various factors, such as differences in the social origin of students.

Although they have a useful role to play, performance measures and indicators need to be handled with care. Their meaning and interpretation must be systematically questioned and, if not used carefully, they may seriously distort the behaviour of organisations, managers and employers (see Likierman, 1993b). For example, a rigid focus on a small range of performance measures, with no provision for dialogue on their interpretation, may successfully achieve certain targets, whilst distracting attention from the attainment of broader organisational goals and objectives.

3. Functions of performance measurement

a. Different measurement systems for different purposes

The development and implementation of performance measurement should be adapted to local circumstances and concerns. The substance of performance measures differs according to the responsibilities of those whose performance is being measured and the requirements of those using the information. At the operational level, measures and indicators should be related to issues such as the management of resources and production processes. At a higher management level, the information should be related to issues of programme effectiveness, in order to inform decisions on policy formulation and resource allocation.

b. Organisational learning and programme management

Performance measurement is useful for evaluating administrative performance and organisational learning. It can be used to improve the operational efficiency of a complete organisation (e.g. a ministry or public agency); or by individual managers within an organisation to evaluate and strengthen the performance of a department, division, unit or other subdivision of that organisation.

Performance measurement can be a useful tool for managing entitlement programmes and investment projects or programmes. Performance information can help ensure that such programmes are implemented in conformity with their objectives, and in preparing new programmes. For example, in the road sector, performance measures covering issues such as mobility/accessibility, traffic flows, and safety and environmental factors can be useful in both preparing and supervising the implementation of programmes.

c. Performance contracts and agreements

Performance measurement can be used as an instrument for strengthening managerial accountability. Results-oriented management systems attempt to link the performance of managers to explicit or implicit contracts, which generally include performance targets. In theory, contracts should provide for both penalties and rewards. However, in most cases only rewards are included in the contracts, for example, the possibility of managers retaining some or all of any efficiency savings made; flexibility in resources (e.g. in staffing numbers); or a performance-related element in the pay of senior management and unit heads.

The dimensions of performance to be measured in defining such contracts are necessarily narrower than for programme evaluation. For example, it is possible to hold managers accountable for the output of a vaccination programmes, and to reward them accordingly. However, it is difficult to hold them
Managing Public Expenditure - A Reference Book for Transition Countries

responsible for the outcome of improving health, which may depend on factors outside their control, such as the quality of water or changes in tobacco consumption. Performance measures used for the purpose of accountability and control focus by necessity on inputs and outputs rather on programme outcomes.

Even where budget funds are provided on the basis of a contract or a performance agreement, the link between performance and resource allocation remains indirect. In most cases, the release of budget funds or grants is not conditional upon delivery of agreed performance results.

Contracts may offer a way of supplementing the over-formal nature of targets and indicators, since they place more emphasis on managing relationships than on simply collecting quantitative information about achievements (see Trosa, 1996a). However, generalising a contractualist approach beyond a few programmes or organisations would be difficult to apply in most transition countries at the present stage of their development.

Performance-related pay systems link some elements of remuneration to a specified level of activity or output and can contribute to improving operational performance. In transition countries, such systems could perhaps facilitate switching from an administrative culture based on command and control of the economy to a customer-oriented culture. However, extreme caution is required before considering whether to implement such systems. Selecting relevant indicators is a tricky issue. Making staff accountable for elements that are not fully under their control is questionable. But focusing on results more directly attributable to the efforts of staff and managers can encourage them to develop short-term responses that trigger rewards, to the detriment of actions that achieve wider programme and organisational objectives. Whether or not performance-related pay schemes actually improve performance is debatable. Motivational theories stress intrinsic motivation (i.e. the job itself, or the ethos of public service in some countries) rather than extrinsic motivation (money and benefits). Moreover, results-oriented personnel management systems may lead to undesirable outcomes where patronage or political considerations are dominant, since politicians and managers will tend to reward “their” people, rather than the best performing individuals.

d. External accountability

In some EU Member States, performance information is published to improve accountability to parliament and taxpayers, and to facilitate value for money audits. For example, in the UK, agencies are required to provide, in annual performance agreements with the minister concerned and in their published annual reports, data on performance spanning a number of years so that comparisons over time can be made. In several countries, the supreme audit institutions comment on the progress of performance measurement and its appropriateness, but in a majority of cases do not comment on the results themselves.

Caution is required before considering the publication of performance information for a wider audience. On the one hand, publication can be useful for both control purposes and its educational role, since it contributes to introducing a performance culture. On the other hand, there are risks of incorrect use of information, if its inherent limitations are ignored, and of demotivating staff by unfair criticism. Publication of performance information favours competition among similar organisations and therefore efficiency. However, it can also contribute to aggravate the gaps between good performers and weak performers, notably in the education and the health sector, by driving resources and the most socially favoured students and patients towards the schools and hospitals that appear to perform best.
4. Effective performance measurement

a. The need for caution

Using performance measures as a performance management tool (e.g. for contracts) or as vehicle for public and political accountability can be dangerous. The experience of centrally planned economies shows that the imposition of norms and standards tends to make the officials concerned focus too rigidly on the achievement of the specified targets. Quality can be sacrificed and the vital link between objectives and performance itself may not get the attention deserved. When standards are maintained regardless of resource availability, the likely result is a weakening of fiscal control. Standards and indicators should not be considered as immutable.

These problems are not exclusive to centrally planned economies and examples of undesirable outcomes in using (or misusing) performance indicators in market economies are numerous. The “law of unintended consequences” states that attempts at modifying behaviour may produce unintended behaviour, which may conflict with the goals and objectives of the policy or programme concerned. For example, if hospital subsidies are based on the length of patient waiting lists, hospital managers and doctors will have an incentive to keep non-critical cases waiting as long as possible whilst focussing their efforts on other cases (higher-quality care for some, little for others); if performance is assessed instead by number of patients treated, the overall quality of care may suffer.

Depending on the way they are set up and used, performance indicators present the following potential dangers:

- **Tunnel vision**, or emphasis on only the quantifiable, neglecting unquantifiable aspects of performance.

- **Measure fixation**, or concentration on what is being measured rather than the service that is being carried out.

- **Short-termism**, failure to attend to legitimate longer term objectives; and **suboptimisation**, or the production of a lower quality of service by concentrating on narrowly defined activities rather than wider organisational objectives.

- **Misrepresentation or deliberate corruption of data**; and misinterpretation or uncritical acceptance of the results of performance measurement.

- **Strategic management of behaviour** including deliberate under-performance in order to engineer targets that can be easily achieved.

- **Inflexible pursuit of defined performance objectives** set at one particular time.

- **Demoralisation** or loss of confidence and commitment amongst workers delivering services that are deemed less important than those targeted for performance measurement.

b. Criteria for good performance measurement systems

To avoid such pitfalls, some general guidelines and criteria can be used when setting up performance indicators (Shand, 1998):
• **Relevance and usefulness.** The measures should be defined properly in relation to the programme to which they relate and reflect the main goals and objectives of the programme. A manager’s performance should be measured only for those areas over which he or she has control.

• **Clarity and understandability.** Performance measures should be simple, well defined and be easily understood by users.

• **Cost effectiveness.** Performance measures must be established at reasonable cost. Data collection costs of introducing performance measures, and managing the system, must be assessed realistically and weighed against the expected benefits.

• **Capacity to monitor results.** As noted earlier, performance is a relative concept, and the measures must be applied consistently over time and between units in order to allow performance to be assessed in a systematic way.

5. **Benchmarking**

Benchmarking is a technique used in both the private sector and the public sector for comparing the performance of one organisation against a standard, whether absolute or relative to the performance of comparable organisations. It can be used to:

• Assess performance against the defined standard(s).

• Expose areas where improvement is needed.

• Identify processes activities that are carried out better in other organisations.

• Test whether measures taken to improve the efficiency or effectiveness of programmes have been successful.

There are two main techniques of benchmarking within the government sector:

• **Process benchmarking** applies to the processes and activities used to turn inputs into outputs. It consists either of benchmarking processes used by the organisation concerned against processes used in comparable organisations, or against processes as defined in a standard.

• **Results benchmarking** applies to actual results (outputs and outcomes). It consists of comparing the actual performance of different organisations using performance indicators, or of comparing actual performance against certain performance standards.

The two main types of benchmarking (process benchmarking and results benchmarking) are increasingly seen as complementary methods that can be used to reinforce each other. For example, results benchmarking can be used to identify discrepancies in results and process benchmarking can help explain why these discrepancies exist. Process benchmarking without results benchmarking tends to become inward-looking, leading to a focus on enhancing processes for their own sake, without checking whether or not the changes are relevant for users of public services and stakeholders.

Benchmarking may be used as a tool both for evaluation and continuous improvement. It is related to a number of management techniques, such as total quality management and process re-engineering,
and for performance comparisons and programme evaluation. Using benchmarking on a selective basis should be considered by transition countries. However, since benchmarking can involve a heavy investment in time and resources it is sensible to focus first on a few key organisations or processes.

Box 15.1 presents some examples of performance indicators used in the UK health sector.

### Box 15.1. PERFORMANCE INDICATORS IN THE UK HEALTH SECTOR

<table>
<thead>
<tr>
<th>Areas and categories covered</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Health improvement</strong></td>
<td></td>
</tr>
</tbody>
</table>
| The overall health status of the population, reflecting social and environmental factors and individual behaviour as well as care provided by the NHS and other agencies i. Deaths from all causes (people aged 15-64) | i. Deaths from all causes (people aged 15-64)  
ii. Deaths from all causes (people aged 65-74)  
iii. Cancer registrations |
| **II. Fair access**         |            |
| Access to elective surgery  | i. Surgery rates  
ii. Conception rate for girls aged 13-15  
iii. People registered with an NHS dentist |
| Access to family planning services | i. Conception rate for girls aged 13-15  
ii. People registered with an NHS dentist |
| Access to dentists          | i. Conception rate for girls aged 13-15  
ii. People registered with an NHS dentist |
| Access to health promotion  | i. Early detection of cancer  
ii. Disease prevention and health promotion  
iii. Inappropriately used surgery  
iv. Cost-effective prescribing |
| Access to community services| i. Early detection of cancer  
ii. Disease prevention and health promotion  
iii. Inappropriately used surgery  
iv. Cost-effective prescribing  
v. Acute care management  
vi. Chronic care management  
vii. Mental health in primary care  
viii. Cost-effective prescribing  
ix. Discharges from hospital |

(Cont'd)
### Box 15.1. PERFORMANCE INDICATORS IN THE UK HEALTH SECTOR (cont'd)

<table>
<thead>
<tr>
<th>Areas and categories covered</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IV. Efficiency</strong></td>
<td></td>
</tr>
<tr>
<td>Maximising use of resources</td>
<td>i. Day case rate</td>
</tr>
<tr>
<td></td>
<td>ii. Length of stay in hospital</td>
</tr>
<tr>
<td></td>
<td>iii. Unit costs</td>
</tr>
<tr>
<td></td>
<td>iv. Generic prescribing</td>
</tr>
<tr>
<td><strong>V. Patient/career experience of the NHS</strong></td>
<td></td>
</tr>
<tr>
<td>Accessibility</td>
<td>i. Patients who wait more than 2 hours for emergency admission</td>
</tr>
<tr>
<td></td>
<td>ii. Patients with operations cancelled for non-medical reasons on the day of, or day after, admission</td>
</tr>
<tr>
<td>Co-ordination and communication</td>
<td>iii. Delayed discharge from hospital for people aged over 75</td>
</tr>
<tr>
<td></td>
<td>iv. First outpatient appointments for which patient did not attend</td>
</tr>
<tr>
<td>Waiting times</td>
<td>v. Outpatients seen within 13 weeks of written referral</td>
</tr>
<tr>
<td></td>
<td>vi. Inpatients admitted within 3 months of a decision to admit</td>
</tr>
<tr>
<td><strong>VI. Health outcomes of NHS care</strong></td>
<td></td>
</tr>
<tr>
<td>Success in reducing level of risk</td>
<td>i. Conception rate for girls aged 13-15</td>
</tr>
<tr>
<td>Success in reducing level of disease, impairment and complication of treatment</td>
<td>ii. Decayed, missing and filled teeth in 5 year olds</td>
</tr>
<tr>
<td></td>
<td>iii. Avoidable diseases</td>
</tr>
<tr>
<td></td>
<td>iv. Adverse events/complications of treatment</td>
</tr>
<tr>
<td></td>
<td>v. Emergency admission to hospital for people aged improving quality of life for patients and over 75</td>
</tr>
<tr>
<td></td>
<td>vi. Emergency psychiatric readmission rate</td>
</tr>
<tr>
<td>Success in optimising function and improving quality of life for patients and careers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vii. Infant deaths</td>
</tr>
<tr>
<td>Success in reducing premature death</td>
<td>viii. Survival rates for breast and cervical cancer</td>
</tr>
<tr>
<td></td>
<td>ix. Avoidable deaths</td>
</tr>
<tr>
<td></td>
<td>x. In-hospital premature deaths</td>
</tr>
</tbody>
</table>

(cont'd)
Box 15.1. PERFORMANCE INDICATORS IN THE UK HEALTH SECTOR (cont’d)

<table>
<thead>
<tr>
<th>Areas and categories covered</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>VII. Example of breast cancer disease</td>
<td>i. Cancer registrations</td>
</tr>
<tr>
<td></td>
<td>ii. Cancer registrations plus interval cancers by stage at first diagnosis</td>
</tr>
<tr>
<td></td>
<td>iii. Incidence of avoidable complications — (recurrence, complications of therapy, etc.)</td>
</tr>
<tr>
<td>NHS success in reducing level of disease, impairment and complications of treatment</td>
<td>iv. Measured using a self-assessment questionnaire or other appropriate measure</td>
</tr>
<tr>
<td>NHS success in restoring function and improving quality of life of patients</td>
<td>v. 5 year survival</td>
</tr>
<tr>
<td>NHS success in reducing premature death</td>
<td>vi. 5 year survival standardised for age and stage of disease</td>
</tr>
</tbody>
</table>

B. Programme Evaluation

1. Definition and objectives

a. What is programme evaluation?

Programme evaluation focuses on the assessment of a programme’s achievements against its objectives. The term “policy evaluation” is also used but this has a wider scope, since it can cover several programmes, the regulatory framework, the analysis of interrelations between programmes and regulations, etc. However, many of the issues described below also apply to policy evaluation. Indeed, some countries do not distinguish programme evaluation from policy evaluation.

Programme evaluation can encompass different stages in a programme life-cycle:

- *Ex post* evaluations are carried out when the programme has been in place for some time to study its effectiveness and judge its overall value. These evaluations are typically used to assist in allocating resources or enhancing accountability. Questions of outcome and the overall relevance of the programme are expected to be addressed.

- *Intermediate* evaluations are usually undertaken during the implementation of the programme. The purpose is to support and improve the management and implementation of the programme. Emphasis is put on operational questions.
b. Objectives

The goal of evaluation is to improve decision-making and resource allocation by providing reliable data about the effects of policies and programmes. Uses of programme evaluation may cover the following:

- **Assisting in resource allocation and identifying desirable policy changes.** Evaluation provides information on the impact of existing policies. It therefore assists policy makers in assessing the value of public programmes and identifying areas where policy changes and/or shifts in the allocation of resources between different programmes may be necessary.

- **Improved programme management and organisational learning.** As noted earlier, feedback mechanisms contribute to the learning process of those managing and implementing programmes and can be used to improve their operational performance.

- **Enhancing public accountability.** Evaluation can improve transparency and accountability by shedding light on the impact of government policies.

c. Evaluation, monitoring, and audit

Evaluation is different from other feedback mechanisms such as monitoring and performance measurement, since it is generally conducted as a single exercise and gathers information in greater depth. While performance measurement focuses on efficiency and effectiveness, and often even narrower issues, evaluation studies assess also whether the programme complies with the needs and socio-economic problems it was designed to address. Evaluation studies often include a detailed review of attribution and causality issues, while performance measurement deals with more roughly assessed outcome or output indicators.

However, regular monitoring and performance measurement systems can provide useful information for successful evaluation. For example, some programmes include pre-determined milestones that record the achievement of certain goals and objectives, these can be used as “anchors” for a more detailed examination of the achievements and failures of the programme.

Evaluation and external audit are historically separate functions carried out by separate institutions. Audit is closely related to the parliamentary oversight function, and underlines the importance of legal compliance and the accountability of public organisations. As discussed in Chapter 14 the independence of supreme audit institutions is crucial, while evaluation is generally an activity carried out under the responsibility of the executive. Although its uses for enhancing accountability are increasing, evaluation is primarily an instrument for strengthening programme management and supporting decision-making. However, in practice, the boundaries between evaluation and audit are becoming blurred, since traditional financial audits are being supplemented with value for money audits, which are similar in methodological terms to programme evaluations.

2. Key evaluation issues

a. The programme logic

A key issue in programme evaluation is to examine the programme’s “intervention logic” (i.e. the basic rationale for analysing a programme in order to examine to what extent it has achieved its goals and objectives); some manuals on evaluation refer to the “programme theory”.
Programmes are always conceived with a given set of needs in mind. These needs are the socio-economic objectives and issues that the programme seeks to address, expressed from the point of view of its particular target population. According to the logical framework approach adopted by the European Commission\textsuperscript{7}, these objectives can be divided into:

- **General objectives**, which are expressed in term of end outcomes.

- **Specific objectives**, which are expressed in term of intermediate outcomes.

- **Operational objectives**, which concern planned inputs and outputs.

A structured approach should consist of the following steps: (i) description of the programme; (ii) clarification of its objectives, and the needs that the programme is aimed at addressing; (iii) identification of the possible causal relation between programme activities and effects; (iv) identification of the possible level of outcomes that can be evaluated (intermediate outcomes and/or end outcomes); (v) identification of outcome indicators and the criteria that will be used to assess effectiveness; and (vi) identification of the factors that may effect the outcomes. Box 15.2 shows an example of such a logical framework approach for an evaluation study, based on the EC’s methodology.

As in the case of performance measurement systems described in Section A, relevant indicators must be set up. Similar problems can arise to those discussed in that earlier section. The attribution problem,

---

**Box 15.2. EXAMPLE OF A LOGICAL FRAMEWORK APPROACH TO EVALUATION**

<table>
<thead>
<tr>
<th>Intervention Logic</th>
<th>Verifiable indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>End outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Modern, private and transparent banking sector, characterised by a sound risk-reward proportionality</td>
<td>Performance rating of commercial banks</td>
</tr>
<tr>
<td><strong>Intermediate outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Restructuring and reorganisation of commercial banks into viable institutions, and preparation for subsequent privatisation</td>
<td>Number of banks privatised</td>
</tr>
<tr>
<td></td>
<td>Prudential and performance ratios of banks, bad debt ratios, levels of provisioning, etc.</td>
</tr>
<tr>
<td></td>
<td>Acceptance of guarantees</td>
</tr>
<tr>
<td></td>
<td>Implementation of corrective action plans</td>
</tr>
<tr>
<td></td>
<td>Organisational structure of the banks</td>
</tr>
</tbody>
</table>

*(cont’d)*
which is related to determining whether and to what extent the programme concerned caused the effects observed, is particularly crucial for evaluating programme outcomes. The evaluation of a programme requires the comparison of results with the targets established in the programme design, or with specific benchmarks. If the programme goals and objectives are stated clearly, when formulating budget policies, their evaluation is significantly easier.

b. Other key issues

Figure 15.1 shows the logical framework and the main issues to be addressed in performance evaluation studies. These key issues can be grouped into the following categories:

- **Continued relevance.** The extent to which a programme is relevant to government priorities. To what extent are the objectives and mandate of the programme still relevant? Are the activities and operational outputs consistent with the programme’s mandate and plausibly linked to its objectives and other intended results?

- **Utility.** How does the intended impact of the programme compare with the needs of the targeted population?

- **Sustainability.** To what extent can any positive changes arising from a programme be expected to last after the programme has been terminated?

- **Efficiency.** How economically and efficiently have the various inputs been converted into outputs?

### Box 15.2. EXAMPLE OF A LOGICAL FRAMEWORK APPROACH TO EVALUATION (cont’d)

<table>
<thead>
<tr>
<th>Operational objectives</th>
<th>Outputs</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained management and staff</td>
<td>Functioning equipment</td>
<td>Technical assistance (training, pre-privatisation support, etc.)</td>
</tr>
<tr>
<td>Manuals</td>
<td>Restructuring plans available and discussed</td>
<td>Number of days of technical assistance by contractors and consultants</td>
</tr>
<tr>
<td>New organisational structure</td>
<td>Organisational structure accepted by management</td>
<td></td>
</tr>
<tr>
<td>Installed equipment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrective action plans</td>
<td>Number of staff trained</td>
<td></td>
</tr>
<tr>
<td>Restructuring action plans</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Source: Commission of the European Communities, evaluation of a Phare banking programme, 1998*
• **Effectiveness.** Were programme objectives achieved? What client benefits and what broader outcomes, both intended and unintended, resulted from carrying out the programme? Does the programme complement, duplicate or overlap with other programmes?

• **Cost-effectiveness.** Are there alternative solutions or programmes that might have achieved the desired objectives and intended results more cost-effectively? Are there more cost-effective ways of delivering the existing programme?

In addition to these issues, an evaluation study should also assess whether a programme has resulted in negative effects or inefficiencies, technically referred to as the effects of “deadweight”, “displacement” and “substitution”. Deadweight refers to effects which would have arisen even if the programme had not taken place. For example, a retraining programme aimed at the long-term unemployed may benefit some people who would have undertaken training even without the programme. Displacement and substitution are used to describe situations where the benefits of a programme on a particular individual, group or area are only realised at the cost of disbenefits to other individuals, groups or areas. For example, in the

**Figure 15.1.  PROGRAMME LOGIC, PERFORMANCE MEASUREMENT AND EVALUATION**

![Programme Logic, Performance Measurement and Evaluation Diagram](image-url)
case of a programme to provide employment subsidies, substitution will happen if, in the enterprises that benefit from the programme, subsidised workers take the place of unsubsidised workers. Displacement will occur where an enterprise benefiting from employment subsidies wins business from other firms that do not participate in the scheme. Thus, the jobs created in the participating firm may be offset by job losses in other firms.

3. Preparing an evaluation study

There are two major phases of an evaluation study:

• **Design.** Identifying the main issues and questions to be addressed and developing a methodology for gathering and analysing information.

• **Implementation.** Collecting and analysing data, drafting a report that presents the findings of the study and makes recommendations.

Most evaluation guides emphasise the importance of the design phase. The key steps in preparing an evaluation study are the following:

• **Identify the goals of the evaluation.** An important initial question is: for what purpose is the evaluation being launched? To improve management? To improve policy decision-making? To improve accountability?

• **Define the scope of the evaluation.** When a programme includes several objectives and target groups, it can sometimes be cost-effective to restrict the evaluation to some particular aspects of the programme.

• **Identify the questions to be asked.** If the purpose of the study is mainly to improve management, it should focus on screening programme implementation and service delivery. If the issue of accountability is being evaluated, the study should focus on the effectiveness of the programme. To provide feedback to decision-makers, the evaluation may need to include a cost-effectiveness analysis and an assessment of the continued relevance of the programme and its utility.

• **Establish the programme logic.**

• **Set benchmarks.** Evaluation is about assessing the “value” of a programme. This involves making judgements on the degree to which the programme’s performance has been “good” or “bad”. Predetermined and transparent benchmarks are needed to ensure that such value judgements do not become arbitrary.

• **Draw up the analytical agenda.** The agenda consists of defining the evaluation design, the data collection methodology and the data analysis techniques. An evaluation design describes the methods that will be used to gather information and draw conclusions on the results that can be attributed to the programme. The design framework depends on both the type of information to be retrieved and the type of analysis to which this information will be subjected.

• **Take stock of available information.** For most programme evaluation work, the monitoring system should be the first source of information. However, this information may need to be supplemented with a review of other statistical sources, questionnaires, user surveys, etc.
• Prepare a work plan and estimate evaluation costs.

• Prepare the terms of reference.

• Select the evaluator.

4. Evaluation design

a. Is there a golden rule?

There are several ways of carrying out an evaluation study. Reviewing them is necessary in order to define an appropriate strategy for evaluation. Literature on evaluation contains many debates on the advantages and disadvantages of particular evaluation methods or approaches. However, as Cheminsky (1995) puts it:

The choice of methods for evaluation and of instruments and data depends more on the question being asked than on the qualities of a particular method. Does the question involve description? Or does it involves reasoning from cause to effect? In a descriptive study, we would not need to worry about the problems of, say, experimental methods because they would not be appropriate. In a cause-and-effect study, many different methods might be applicable depending on the policy question. This centrality of the question, rather than the method, pushes us in the same direction we had to take because of the weaknesses of individual methods: toward complementary and reinforcement. For example, we mitigate the superficiality of a survey by adding case studies, we humanise a time-series analysis by conducting a survey or a set of interviews, and we integrate a process evaluation within an outcome study. Focusing on the question rather on the method has liberated us somewhat with regards to our methodological choices and brought a new emphasis on pluralism that sits rather uneasily with the evaluation chapels of our recent past.

Or as the EC evaluation guide (Commission of the European Communities, 1997) succinctly puts it: “golden rule: there is no golden rule”.

Possible approaches to the design of an evaluation study and methods for data collection and analysis are reviewed briefly below. More technical and detailed reviews of evaluation methods will be found in the abundant literature on the subject.

b. Experimental and quasi-experimental designs

Evaluating a programme requires understanding what results the programme has caused. For this purpose, the most common approach is to assess the effects of a programme against what would have happened in the programme’s absence, i.e. a counterfactual. Because this hypothetical outcome cannot be directly observed, however, the evaluator must apply some techniques to identify the counterfactual against which programme impact is to be measured.

Experimental evaluation techniques are aimed at better identifying the results that can be directly attributed to the programme. They apply the methodology of the natural sciences to public programmes (notably, to social programmes). The experimental design involves randomly assigning programme beneficiaries to two groups. This process is intended to make the two groups as similar as possible in all respects. One of the groups, called the “experimental group” or the “treatment group”, participates in the programme under examination. The other group, called the “control group” does not participate. In a properly constructed experiment, the differences in outcomes between the two groups can be attributed to the effects of the programme or policy.
An evaluation study based on an experimental (randomised) design has the advantage of producing results in which there should be a high degree of confidence. Unfortunately, it is often very difficult to obtain such reliable results. Full comparability of control and treatment groups is hard to achieve. Experimental designs must be implemented from the start of the programmes concerned, otherwise differences will exist between those who have benefited from the programme and those who have not. The notion of control groups can pose ethical problems. For example, it can be illegal or unethical to grant social benefits to one group and not to another group.

These difficulties led to the deployment of quasi-experimental techniques, of which there are several types. The more “robust” quasi-experimental design consists of pre-programme/post-programme comparisons with a non-equivalent (non-randomised) control group. This approach is broadly similar to the experimental design method described above. However, the method of selecting the control group is less rigorous, and statistical techniques may be used to adjust for any initial differences between the two groups. Other techniques that can be used include time-series analysis and post-comparisons between several groups. The first of these techniques involves the collection and analysis of time-series information in order to identify changes or trends in behaviour that may be attributable to the policy or programme under examination.

c. Causal model approaches

An alternative approach to the experimental models described above involves the use of causal models. Such models attempt to measure the impact of a range of factors (independent variables) on the outcomes of a government policy or programme (dependent variables). The programme itself is only one of the factors that determine these outcomes. Different sorts of model may be used, when appropriate, such as simulation models; input-output models; microeconomic models; macroeconomic models; and statistical models. For example, several models are or have been used to assess the effects of the EU Structural Funds policy. Computable General Equilibrium Models (CGE) are sometimes used to assess the impact of tax policies or programmes on income redistribution. Causal models tend to be used in situations where there is already evidence of a relationship between the independent and dependent variables. However, using such models presents risks and should be used with care. When the model is not well specified and coherent, and not based on a sound analysis of the relationship between the variables and parameters, its results are often misleading. Generally, causal models should be used as a complement to other evaluation methods.

d. Economic evaluation

Economic evaluation introduces information on costs and benefits into the evaluation methodology. It is either conducted separately or as a complement to other evaluation methods. Thus, cost-benefit and cost-effectiveness analysis methods can be used ex post, to assess whether the actual costs of the programme were justified by the actual benefits.

e. Non-causal approaches

In some cases, there is a problem of circularity that makes it difficult to establish a direct relationship between a programme and its effects. The identified outcomes may be due partly to the programme’s influence, but the programme can also be influenced by the external factors that contributed to these outcomes. For example, it is often found that cities served by a motorway experience more rapid economic development than other cities. However, this may be due to the fact that the motorway route was designed in order to serve cities with the highest potential for economic development. Neither the experimental approach nor the causal model approach can deal with this problem of circularity.
In such cases a pragmatic approach is often adopted. The evaluation does not attempt to find a counterfactual, but provides a thorough description of the programme and makes extensive use of interviews with stakeholders, case studies, analysis of documents, assessments by experts, etc. Identifying the behaviour of different stakeholders and explaining how it affects programme processes and outcomes can cast light on the underlying causal relationships and is important for successful achievement of policy objectives.

In evaluating Research and Development (R&D) programmes, it should be recognised that by their nature the outcomes of such programmes are uncertain — indeed, were the outcomes ensured in advance, there would be no need for the research. Peer review is commonly used as an evaluation technique in this field and can contribute to an ex post assessment of whether the appraisal of risks against potential benefits was reasonable.

Naturalistic (or qualitative) evaluation methods are based on the principle that the world is socially constructed and constantly changed by the interaction of individuals. Such an approach may be appropriate where the evaluator is not committed in advance to a particular set of values or outcomes and is prepared to work with stakeholders to identify those values and the relevant “facts” that are not objectively based or causally determined. In this approach, evaluation cannot provide objectively “correct” answers but instead can act as a facilitating mechanism to produce a consensus among stakeholders. Naturalistic evaluation methods include participant observation, ethnographic methods, informal interviewing procedures, case studies, etc. In public sector applications, the importance of the political dimension in decision-making complicates the issue of identifying the role of different stakeholders so that making use of elements of the naturalistic approach alongside other methods may be more relevant than a pure naturalistic study.

Assessing the views of programme beneficiaries is increasingly used in evaluation studies. It consists of participant observation, qualitative interviewing and related techniques to gauge beneficiary values and preferences. Such approaches can derive information on many factors at the household and community levels that would be beyond the scope of more quantitative techniques.

5. Data collection and analysis

Once the general approach and design of an evaluation study is agreed, the next step is to define the data needed to obtain the necessary information. Data are facts and statistics that can be observed and recorded. Deciding which data are most relevant raises the questions of measurement and attribution discussed earlier in Section A on performance measurement.

If reliable data cannot be obtained from a secondary source, primary data collection becomes necessary. Primary data collection, however, will generally cost more than reliance on secondary data and should therefore be avoided if possible. A plan to extract primary data typically involves selecting a collection technique (such as natural observation and surveys), developing measurement techniques (such as questionnaires, interview guides and observation record forms) and preparing a sampling plan.

Case studies may be used when it is impossible, for budgetary or practical reasons, to choose a large enough sample, or when in-depth data are required. Such studies allow the evaluator to perform detailed analysis and, therefore, can generate valuable information and explanatory hypotheses for further analysis. Case studies may also be used to examine a number of specific activities or projects, through which the evaluator hopes to reveal information about the programme as a whole. Alternatively, a case study may be chosen because it is considered a particularly relevant example, or to compare the functioning of an organisation or the implementation of a programme with “best practice.”
Depending on the type of analysis required and the availability of data, specific data analysis methods must be determined (such as cost-benefit, multiple regression, analysis of variance). Statistical analysis involves the manipulation of quantitative or qualitative data to describe phenomena and to make inferences about relationships between variables. Non-statistical analysis is carried out, for the most part, on qualitative data—such as detailed descriptions of activities or processes or the transcripts of group discussions. Several types of non-statistical analysis exist (content analysis, inductive analysis, etc.). Non-statistical data analysis relies on the evaluator’s professional judgement to a greater degree than is the case with statistical analysis.

Reporting the findings of evaluation studies often involves the presentation of a large volume of data in a concise manner. Statistical tabulations, graphical displays and simple statistical analysis, such as the mean or the variance, can be used to highlight key characteristics of the data.

6. Evaluation reports

An example of the information that should be included in an evaluation report is given in Box 15.3. However, the structure of these reports needs to match the particular goals of the evaluation study and the needs of report users: there is no universally applicable model.

7. Problems of method and implementation

Methodological problems are intrinsic in all approaches to evaluation, but can be dealt with when the limitations are recognised and the issues are properly addressed. This requires specific knowledge and skills that can be gained by training staff and commissioning external expertise to conduct evaluations. However, such problems do not imply that carrying out evaluation studies is a worthless activity. Even if evaluation cannot provide definitive answers, it can add useful information to the discussion about the design and implementation of government policies and programmes.

a. Difficulties and threats

Problems related to causality are common to social sciences in general. Conclusive evidence of cause-effect relationships can rarely be established, since controlling all relevant variables is seldom possible. Experimental evaluation design is often difficult, expensive and lengthy, if not impossible to apply in practice. Even if experimental evaluation design is used, generalising the results beyond the conditions of the experiment is usually uncertain. Causal relationships between a programme and an observed outcome often cannot be unequivocally proven, mainly because of the intractability of the measurement and attribution problems discussed earlier.

Another difficulty is deciding whether to focus only on the officially recognised objectives of a programme (i.e. those included in statements of government policy) or to take a broader view and study all the effects of the programme. The latter approach gives a more comprehensive picture of the outcomes of the programme but is more complex and time-consuming. Setting an appropriate time period over which the programme is evaluated is difficult but critical, as relevant outcomes should have sufficient time to mature. However, the information’s usefulness may diminish if the evaluated programme is changed before the evaluation report is finalised or the evaluation findings can be applied.

Assessing whether evaluation findings can be generalised is of particular importance when evaluation is expected to contribute to future policy decisions. However, the conditions under which the programme took place are not necessarily representative of future conditions.
b. Criteria for successful evaluation

The conclusions of an evaluation study should be based on a comprehensive coverage of the relevant issues. The evaluator should try to get as accurate a picture as possible of the issues of concern and explore them as far as time and financial resource constraints allow. However, a focus on breadth is important. If breadth is sacrificed for greater depth of analysis in the issues covered, the conclusions reached may be narrowly accurate but lacking in perspective.

Given that evaluation is an aid to decision-making, the criteria for selecting an appropriate evaluation method must ensure that useful information is produced. This implies an understanding of the decision-making environment to which the evaluation findings will be introduced.

---

**Box 15.3. AN EXAMPLE OF AN EVALUATION REPORT STRUCTURE**

*Executive summary*

- An overview and summary of the entire report.
- A discussion of the strengths and weaknesses of the chosen evaluation methods.

*Introduction*

- Description of the programme in terms of needs, objectives, delivery systems, etc.
- The context in which the programme operates.
- Purpose of the evaluation in terms of scope and main evaluation questions.
- Description of other studies that have been done.

*Research methodology*

- Design and implementation of the research and collection of data.
- Analysis of data.

*Evaluation results*

- Findings.
- Conclusions.
- Recommendations.

*Annexes*

- Terms of reference.
- Additional tables.
- References and sources.
- Glossary of terms.

In developing an evaluation method, it is necessary to take into account basic considerations such as practicability, affordability and ethical issues. An approach is practicable to the extent that it can be applied effectively without adverse consequences and within time constraints. Affordability refers to the cost of implementing an evaluation study. Implementing the method most appropriate to a given situation might be unrealistically expensive. Objectivity is of paramount importance in evaluative work. It should always be clear to the reader what the conclusions are based on, in terms of the evidence gathered and the assumptions used. Evaluation information and data should be collected, analysed and presented so that if others carried out the same exercise and used the same basic assumptions, they would reach similar conclusions. Evaluators may frequently be called on to provide advice and recommendations to the client who commissioned the study. In these circumstances, it is important to maintain a distinction between the objective findings of the study, and programme recommendations derived from the evaluation itself or from other sources of information, such as policy directives. When conclusions are ambiguous, it is particularly important that the underlying assumptions are spelled out.

Resistance may be encountered to making full use of evaluations. Politicians are often reluctant to allow sensitive areas of policy to be evaluated, discuss findings of evaluation studies, or to formulate policy goals precisely. Managers may fear to be criticised. In most countries, evaluations have to gain support and need champions. This requires dialogue with decision-makers and stakeholders when carrying out evaluation studies. The stakeholders are those with an interest in the outcome of the evaluation, such as those operating a programme under examination. They should be consulted in defining the issues at stake and planning the evaluation, as they are typically expected to supply data to the evaluator and often play a major role in interpreting the results and implementing the recommendations. However, stakeholders sometimes feel that their interests are threatened by an evaluation. If they become actively opposed, they can sometimes sabotage the project.

8. The role of evaluation in transition countries

Developing an evaluation culture needs time, and the development and the development of such work in OECD countries is uneven and not systematically carried out. It is not therefore recommended that transition countries should set as an immediate objective the development of a comprehensive system of evaluation; these countries have higher priority tasks in related areas such as building up an effective system of external audit.

However, transition countries need generally to make shifts in the composition of their expenditure programmes, and evaluation studies could provide information and analyses that are useful in preparing the ground for such changes. Thus, in a number of countries, the preparation of evaluation studies in areas such as social assistance, health or education could be desirable.

Institutional arrangements for carrying out evaluation studies vary from one country to the other. Box 15.4 shows some examples of the arrangements in some EU Member States. In transition countries, it might be possible to establish a small unit at the central level, perhaps in the ministry of finance, to provide expertise and methodological guidance to line ministries, and to assist them in the preparation of their evaluation work and in drafting the terms of reference for the studies.
**Box 15.4. INSTITUTIONAL ARRANGEMENTS FOR EVALUATION**

**France.** A National Council of Evaluation was established in 1999 and is responsible for preparing an annual evaluation programme on the basis of proposals formulated by line ministries and local governments. This Council is composed of scientists, other experts and representatives of local authorities. The evaluation studies are financed by a National Fund for Evaluation, and are published. Besides the activities co-ordinated by the National Council, line ministries and sector evaluation committees also carry out evaluation studies.

**Netherlands.** Budget directorates within line ministries are responsible for co-ordinating the programme of evaluation studies and ensuring that necessary advice, guidance and research expertise is provided. They draw up evaluation programmes for individual projects, encourage the periodic evaluation of policies and monitor the quality of the analyses carried out and its application. Other directorates provide support on issues such as personnel and organisational management, auditing and legislation. The Court of Audit reviews the quality of the evaluation methodology and the organisation and management of evaluation studies, and publishes reports on these matters.

**United Kingdom.** Organisational arrangements for evaluation are diversified and have a “polycentric” character. The National Audit Office, the Audit Commission, HM Treasury, line ministries, executive agencies and many local authorities undertake evaluation studies. There is no single organisation that is responsible for supervising or co-ordinating this work, though the Treasury has published some guidance documents. Evaluation is well developed in some areas and findings of evaluation studies are used in setting (and adjusting) policy priorities and in budget management.

**Sources:**
NOTES


3. Where government programmes are operated through a network of comparable institutions in different regions or localities (e.g. schools or social security benefit offices), “internal” benchmarks can be established, i.e. school A can be compared with schools B and C. This technique has been used in countries such as the UK to create an internal market, promote competition and raise service standards in areas such as health care, education, tax collection and the payment of social benefits.


6. E.g. in France “public policies evaluation” refers to both programme evaluation and policy evaluation. (Conseil Scientifique de l’évaluation, 1996).


9. See, for example, Weiss (1998), especially Chapter 11 on “qualitative methods”.

10. Squire (1995) has argued that, to constitute an evaluation, such information would have to be used in the context of either an experimental or quasi-experimental evaluation.