



*Manual*

# **Performance audits**

02 February 2006

Audit Department II

European Affairs & Government-wide Performance Audit Division

Netherlands Court of Audit, Lange Voorhout 8, P.O. Box 20015, 2500 EA The Hague, The Netherlands

CONFIDENTIAL This draft document is the property of the Netherlands Court of Audit. It contains confidential information that is intended solely for administrative or official comment.



# Contents

	<b>Preface</b>	<b>6</b>
<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	Efficiency and effectiveness audits conducted by the Netherlands Court of Audit	8
1.1.1	Have the objectives been achieved?	8
1.1.2	Has the target group been reached?	9
1.1.3	Effectiveness audits	10
1.1.4	Efficiency audits	11
1.2	Summary of contents	12
<b>2</b>	<b>General framework for efficiency and effectiveness audits</b>	<b>14</b>
2.1	Introduction	14
2.2	Legislative framework	14
2.2.1	Duties of the Netherlands Court of Audit	14
2.2.2	Powers of the Netherlands Court of Audit	15
2.2.3	Administrative responsibility	16
2.2.4	From policy budgets to policy accountability	17
2.2.5	Dutch Central Government Performance Data and Evaluative Studies Regulations	17
<b>3</b>	<b>Audit subjects and types of audit</b>	<b>19</b>
3.1	Introduction	19
3.2	The public-sector production process	19
3.2.1	Outputs	20
3.2.2	Outcomes	23
3.2.3	Throughput	24
3.2.4	Input	25
3.3	Explaining the causes	25
3.4	Audit scope and methods	28
3.4.1	What steps should the audit involve?	28
3.4.2	How do you decide which type of audit to perform?	29
3.4.3	Ministerial responsibility	34
3.4.4	How do you deal with decentralised policy?	34



3.4.5	How should you audit outputs and outcomes for which non-central government bodies are responsible?	35
<b>4</b>	<b>Audit strategies</b>	<b>38</b>
4.1	Introduction	38
4.2	Objective, problem and audit questions	38
4.2.1	Objective	39
4.2.2	Problem	39
4.2.3	Audit questions	40
4.2.4	Hypotheses	40
4.2.5	Other details on the nature of the audit	41
4.3	Audit methods	41
4.3.1	Choice of audit methods	42
4.3.2	Methods	46
4.4	Methods matrix	62
<b>5</b>	<b>Analysing policy aims and policy information</b>	<b>64</b>
5.1	Introduction	64
5.2	Policy aims	64
5.2.1	General	64
5.2.2	Reconstruction and operationalisation of policy aims	65
5.2.3	Standards for policy aims	66
5.2.4	Assessing the formulation of policy aims	67
5.2.5	Recommendations	67
5.3	Policy information	68
5.3.1	General	68
5.3.2	What policy information?	68
5.3.3	Sources of information on outputs and outcomes	70
5.3.4	Quality of policy information (as a product)	72
5.3.5	Quality of policy information (as a process)	73
5.3.6	Assessing the use made of policy information	76
5.3.7	Checklists for assessing policy aims and policy information	77
5.3.8	Provision of information to the House of Representatives	78
5.3.9	Other points of interest	78
5.3.10	Recommendations	79
<b>6</b>	<b>Audits of the achievement of objectives (and the reaching of target groups)</b>	<b>81</b>
6.1	Introduction	81
6.2	Audits of the achievement of objectives	81
6.2.1	Problem definition and audit questions	81
6.2.2	Criteria	82
6.2.3	Audit design	82



6.2.4	Performance indicators	83
6.2.5	Outcome indicators	84
6.2.6	Recommendations	85
6.3	Audits of the reaching of target groups	85
6.3.1	Problem definition	85
6.3.2	Audit questions	86
6.3.3	Criteria	88
6.3.4	Audit design	90
6.3.5	Target group indicators	90
6.3.6	Recommendations	90
<b>7</b>	<b>Effectiveness audits</b>	<b>91</b>
7.1	Introduction	91
7.2	Problem definition and audit questions	91
7.3	Criteria	94
7.4	Audit design	95
7.5	Audit methods	97
7.6	Effectiveness indicators	98
7.7	Recommendations	98
<b>8</b>	<b>Efficiency Audits</b>	<b>100</b>
8.1	Introduction	100
8.2	Format of efficiency audits	100
8.2.1	Objective	100
8.2.2	Problem definition and audit questions	100
8.2.3	Criteria	101
8.2.4	Recommendations	102
8.3	Terms	103
8.3.1	Outcomes, outputs and inputs	103
8.3.2	Comparison	103
8.3.3	Quality	104
8.3.4	Explanations	105
8.4	Defining the limits of the audit subject	105
8.4.1	General or operational objectives?	106
8.4.2	Outputs or outcomes?	107
8.5	Collecting data	108
8.5.1	Quality of data	108
8.5.2	Data on outcomes or outputs	110
8.5.3	Inputs	110
8.5.4	Quality	112
8.5.5	Finding and using explanations	113
8.6	Calculating efficiency	115
8.6.1	Efficiency indicators	117



8.6.2	FDH, DEA and SFA	124
8.6.3	Accounting for differences in efficiency	138
	<b>Annexe 1 Literature</b>	<b>147</b>
	<b>Annexe 2 Data collection: (longitudinal) surveys</b>	<b>156</b>
	<b>Annexe 3 VBTB budgets</b>	<b>158</b>
	<b>Annexe 4 Criteria for assessing the formulation of policy objectives</b>	<b>160</b>
	<b>Annexe 5a Criteria for assessing policy information and its usefulness</b>	<b>166</b>
	<b>Annexe 5b Details of the 'completeness' sub-criterion in relation to policy information (part of the 'relevance' criterion)</b>	<b>171</b>
	<b>Annexe 5c Examples of the use of criteria taken from previous audits</b>	<b>179</b>
	<b>Annexe 6 Description of the modus operandi method</b>	<b>190</b>
	<b>Annexe 7 Checklist: can you calculate an efficiency score yourself?</b>	<b>191</b>
	<b>Annexe 8 General conditions applying to DEA and SFA analyses</b>	<b>193</b>
	<b>Annexe 9 Glossary</b>	<b>196</b>



## Preface

The efficiency and effectiveness audits carried out by the Netherlands Court of Audit have changed greatly since the first *Performance Audit Manual* was published in 1996. Indeed, this process of change is still going on, mainly due to internal and external factors. The main external factors have been the enactment of the Government Accounts Act 2001 and the publication of a policy document entitled *From policy budgets to policy accountability*.

The most important internal factors have been the Court's mission and strategy. This manual has been conceived as a strategy-neutral document, and contains information on the analysis of policy objectives and policy information, and studies of outputs, outcomes and explanations. These subjects are likely to affect to a greater or lesser degree any strategy adopted by the Court. The idea is that auditors should be able to select from this manual those sections that have a bearing on the audit they are intending to perform.

The manual starts by outlining the *framework* in which the Court carries out efficiency and effectiveness audits, and by defining the most important *terms*.

The manual also describes the type of *decisions* you can expect to take when planning and conducting an efficiency or effectiveness audit. How you do this depends greatly on the nature of the problem and the domain in which the audit is to be conducted. For this reason, the manual does not contain any hard and fast rules for audit strategies. What you will need to do in each case is to explain in detail how far your audit intends to go (and this can range from simply finding out whether the objectives of a given policy have been achieved, to assessing the social impact and cost), what sort of data you intend to collect, how you are planning to collect them and how you intend to analyse them in such a way that you can answer the audit question.

The *standards* set by the Netherlands Court of Audit for government action are an important factor when it comes to formulating conclusions.



Efficiency and effectiveness audits come with fixed standards for assessing the formulation of policy objectives (see section 5.2.3) and the quality of policy information (see section 5.3.4). No general standards have been set for assessing the way in which outputs and outcomes have been formulated. In this case, you will need to formulate specific standards of your own, based on the stated objectives and the arrangements made in the relevant policy field (see also section 6.2.2). The same applies broadly speaking to the assessment of the effectiveness of government policy (see section 7.3). In order to guarantee consistency in the way in which standards are formulated, these have been codified in the form of a *standards database*; every audit team is obliged to consult this database when compiling an audit proposal and to refine it once the audit has been completed.

This *Performance Audit Manual* replaces the *Performance Audit Manual* (1996), the *Policy Outputs Audit Manual* (1997) and the *Efficiency Audit Manual* (2001). The authors of this manual have used material from all three previous manuals. The manual has been designed to tie in with the material used in the course on audit methods and techniques.



# 1 Introduction

## 1.1 Efficiency and effectiveness audits conducted by the Netherlands Court of Audit

The efficiency and effectiveness audits conducted by the Netherlands Court of Audit are designed to measure outputs and outcomes. There are a number of different ways of looking at outputs and outcomes, however: we can analyse the degree to which the body in question has achieved its objectives, we can analyse the effectiveness of policy and we can analyse the efficiency of outcomes and outputs.<sup>1</sup> These different types of audit are layered: in order to audit the effectiveness of government policy, the assumption is that you already have information on the degree to which the government has been successful in achieving its policy aims. In addition, an awareness of the effectiveness of policy is required in order to reach a judgement on the efficiency of the outcomes. Against this background, the different types of audit – ranging from measuring a particular body’s success in achieving its policy objectives to measuring its efficiency in doing so – are growing both ever more complex and ever more demanding in terms of the quality of the policy information on which they are based.

### 1.1.1 Have the objectives been achieved?

An audit of the achievement of objectives involves analysing the actual social situation after a policy has been implemented, and comparing this with the policy objectives, i.e. the social situation the policy was intended to create. It is not possible, however, to establish a causal *link* between the government’s performance and the degree to which certain effects have been produced. In other words, it does not allow you to make any pronouncements about the extent to which a minister’s policy has been successful or not. This type of audit is particularly relevant to the Court’s remit (i.e. improving the operation of central government and/or legal

---

<sup>1</sup> There is also a fourth type of audit: an analysis of the extent to which the target group has been reached. This type of audit is discussed in chapter 6; see also section 1.2.2 below.





persons with statutory tasks) if it is reasonable to expect the government body in question to have enjoyed only limited success in achieving its aims or if you want to build a platform for a more in-depth analysis of the effectiveness of government policy.

For this reason, the strategy described in chapter 6 for assessing the achievement of objectives is as open as possible to more precise effectiveness audits.<sup>2</sup>

In some cases, it is either too early or impossible for other reasons to establish whether the ultimate objectives have been attained, in which case it may be useful to try and establish whether the auditee has reached certain predefined milestones along the way. This may also be important where a minister is not actually responsible for the achievement of a given aim, but is responsible for ensuring that certain milestones are reached.

This applies in particular to *facilitation audits*. The key question here is whether the conditions have worked that the minister has put in place in order to ensure that government policy is properly implemented by regional, provincial or local authorities. In other words, facilitation audits look at intermediate stages in the policy chain, the stages between the role performed by central government in creating the right conditions for policy and the work performed by local authorities and organisations.

In a setting in which policymaking has been largely decentralised, facilitation audits form an excellent complement to audits of central government performance. The minister in question is directly accountable if the work of his or her ministry is found not to have had any effect upon the possibilities of local authorities. This may explain why certain social effects have not been achieved or have not been achieved to a sufficient degree.

### **1.1.2 Has the target group been reached?**

In certain practical situations, it may prove difficult to establish whether certain general objectives, such as 'conserving nature in the Netherlands', have been achieved. What you can do in such cases is establish the auditee's success in achieving certain secondary aims or milestones. One of these is its ability to *reach its target group*.

---

<sup>2</sup> See also Lulofs and Schuddeboom (1991).



An audit of the government's ability to reach its target group may be appropriate where the policy in question is expected to result in either individuals or certain groups displaying a particular type of behaviour. In such cases, policy is effective if it has an impact on the right people. This means that the predefined target group has got to come into contact with the policy in some way or another.

Whilst reaching the target group is not a sufficient condition for also achieving the policy objectives, it is very unlikely that these will be achieved if either the target group or the relevant intermediaries have not been reached. Theoretically, it is possible to include in any audit of the impact of government policy an examination of the extent to which the auditee has been successful in reaching either its target group or the relevant intermediaries.

### **1.1.3 Effectiveness audits**

When the Court performs an effectiveness audit, the purpose is to establish whether the government policy in question has had the desired effect. The term 'effect' covers both outputs and the outcomes achieved by means of these outputs.

Effectiveness audits are complex, partly because *the policy chain may be long* and partly because there may be *external factors* at work that can affect the ultimate aims of the policy in question. There are often a whole series of steps between government outputs and the social situation that is ultimately created.<sup>3</sup> The more steps there are in the chain leading to a given output or outcome, the harder it is to establish a causal link with government policy. In many cases, the outcome is not simply the result of the policy pursued by the minister, but is affected by a range of other factors. For example, whether mature women are successful in returning to the labour market depends not only on measures taken to improve placement services for such women, but also on a variety of socioeconomic factors.

---

<sup>3</sup> Environmental subsidies are a case in point. The chain from outputs to social impact includes a series of intermediate steps such as the payment of subsidies, the target group's awareness of the existence of the subsidies, the decision taken by the target group to buy the product in question, the use of the product by the target group, and environmental changes.



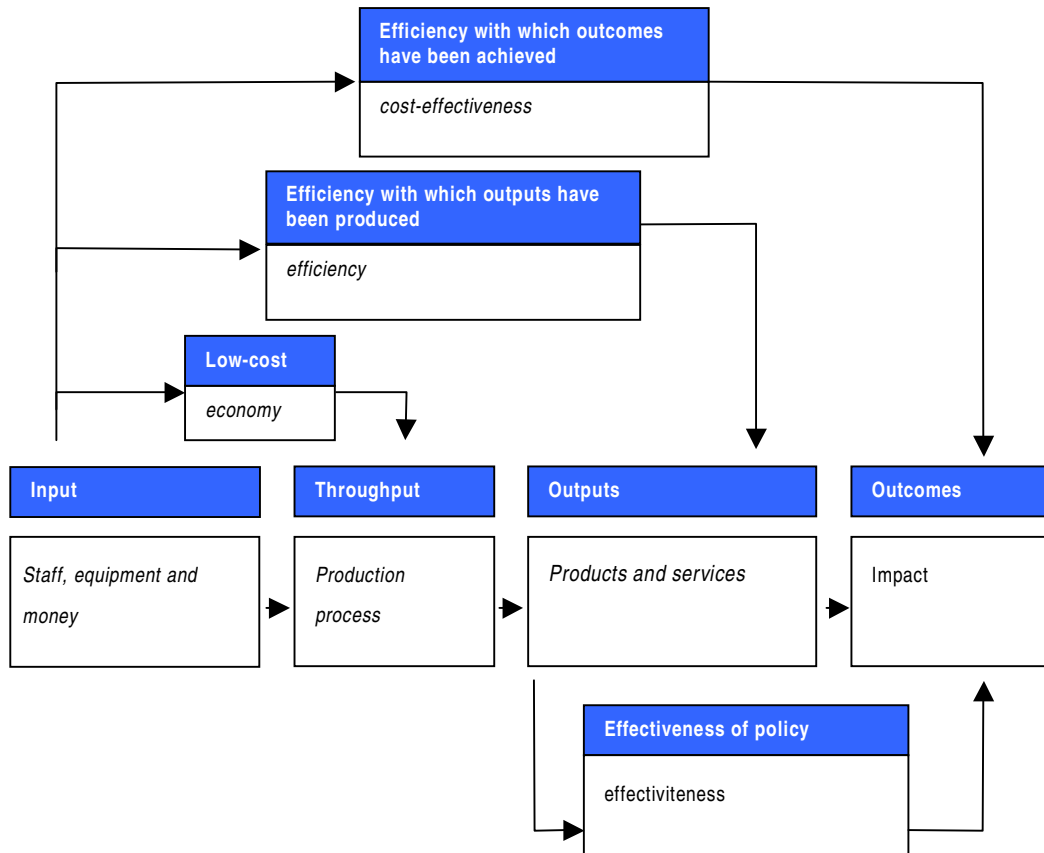
#### 1.1.4 Efficiency audits

Citizens are interested not just in whether certain policy outputs have been achieved, but also in how much they have cost and whether the cost can be reduced. Hence the importance of conducting efficiency audits, in addition to audits of outputs and outcome. Such audits may either concentrate on the outcomes or on the outputs produced by the implementation of government policy. In the former case, the audit is geared towards establishing the cost-effectiveness of the outcomes of government policy, whilst the aim in the latter case is to measure the efficiency of the outputs produced by the implementation of government policy. The question is whether:

- the same outcomes or outputs could have been achieved with fewer resources; or
- the same resources could have produced more outcomes or outputs.

In practice, the terms 'policy efficiency' and 'operational management efficiency' are used too. The former relates to the policy outcomes, whilst the latter relates to policy outputs.

Figure 1.1 shows the relationship between efficiency and effectiveness audits on the one hand and the various aspects of the public-sector production process on the other. The figure does not include audits of the achievement of objectives or the reaching of the target group, as these may be regarded as sub-types of effectiveness audits.



**Figure 1.1 The public-sector production process** (source: Ministry of Finance, 2001)

Both the intended and the actual outputs and outcomes play a role in the above figure.

## 1.2 Summary of contents

Chapter 2 of this manual discusses the legislative framework as this affects efficiency and effectiveness audits. Under the Dutch Constitution, the Netherlands Court of Audit enjoys an independent position in the Dutch constitutional system. The Court's duties and powers are regulated by the Government Accounts Act. This chapter explains how the Court's constitutional position and its duties and powers affect its work in conducting efficiency and effectiveness audits.

Chapter 3 discusses the object of the audits: what are the various stages of the production process in the public sector, and what types of outputs and outcomes does the Court examine in its audits? This chapter also explains how to choose between different types of audit: audits of the



achievement of objectives, audits of the reaching of the target group, effectiveness audits and performance audits.

Chapter 4 is all about collecting and analysing data for the purpose of efficiency and effectiveness audits. Dependent on the nature of the problem in hand and the type of audit chosen, you can use various data collection methods to make pronouncements about outputs and outcomes. Depending on the data collection method you adopt, you can then use a variety of analytical methods to assess possible explanations for discrepancies in outputs and outcomes. The various methods are discussed in this chapter.

Chapter 5 outlines various methods for auditing policy objectives (i.e. what outputs and outcomes is the policy intended to produce?) and explains the standards used for assessing policy information.

Chapters 6 to 8 explain how you can use either policy information (provided it is of a sufficiently high standard) or data you have collected yourself to audit outputs and outcomes. Problem definitions, standards, audit designs and possible recommendations are discussed for all the various types of audit.

We should like to stress that it is by no means necessary for every audit to keep strictly to the procedure set out in this manual. Auditors are encouraged to select those items that are relevant to the audit they are planning to perform.



## 2 General framework for efficiency and effectiveness audits

### 2.1 Introduction

The Dutch Constitution guarantees the independence of the Netherlands Court of Audit in the country's constitutional system. The Court's duties and powers are regulated in the Government Accounts Act 2001.<sup>4</sup>

The Court's audits are based on the administrative responsibility of ministers, state secretaries and the boards of organisations which have links with central government (see section 2.2.3).

### 2.2 Legislative framework

#### 2.2.1 Duties of the Netherlands Court of Audit

The Dutch Constitution states that the Netherlands Court of Audit 'shall be responsible for examining the State's revenues and expenditures' (article 76). Under the Constitution, the organisation, composition and powers of the Netherlands Court of Audit are regulated by Act of Parliament, i.e. the Government Accounts Act 2001. The latter Act also contains details of the audits which the Netherlands Court of Audit is entitled to perform. The Constitution also states that special laws may be enacted in order to assign additional duties to the Court.

Under section 85 of the Government Accounts Act 2001, the Netherlands Court of Audit is required to 'examine the effectiveness and efficiency of the policy pursued, and the efficiency of financial and material management, of the records kept for this purpose and of the organisation of central government'.<sup>5</sup>

---

<sup>4</sup> Bulletin of Acts and Decrees 2002, 413.

<sup>5</sup> The legislation does not actually make clear what exactly is meant by 'the efficiency of financial and material management, of the records kept for this purpose and of the organisation of central government'. For example, the explanatory memorandum accompanying the Government Accounts Act 2001 refers for further details on this section to



Under section 91 of the Government Accounts Act 2001, the Court has certain powers of audit in relation to legal persons with statutory tasks or having financial ties with the State.

Section 92 of the Government Accounts Act 2001 empowers the Court to audit organisations which have been awarded an EU grant.<sup>6</sup>

Finally, section 90 of the Government Accounts Act 2001 gives the States General the power to request the Court to carry out certain specific audits.

### **2.2.2 Powers of the Netherlands Court of Audit**

The Netherlands Court of Audit has wide-ranging powers to perform audits of government departments. Under section 87 of the Government Accounts Act 2001, the Netherlands Court of Audit 'may, in so far as it regards this as being necessary for the performance of its duties, inspect all goods, records, documents and other information carriers in such manner as it may determine'. Ministers are also obliged to supply the Netherlands Court of Audit with any information which it regards as being necessary for the performance of its duties. The Court also has certain powers to audit bodies affiliated with the State. If the government pays money to a particular organisation for a certain purpose, the Court is empowered to establish what exactly the money is used for. In the case of legal persons with statutory tasks, the Court's powers are restricted to examining the performance of the statutory tasks in question (which

---

the explanatory notes on sections 20 and 21. These sections refer to the 'efficiency of policy' and the 'efficiency of operational management'. The explanatory memorandum defines the efficiency of operational management as being the relationship between outputs and input. However, the explanatory memorandum is not entirely consistent in its use of terminology, as it also takes the term as covering the relationship between the (cost of) the deployment of management and control tools and the risks run by the organisation (p. 33). The Netherlands Court of Audit has already pointed out that that this definition of 'efficiency' is inaccurate and inconsistent with the way in which the term is used in other parts of the memorandum. The memorandum also uses the term 'efficiency of administrative expenditure', which it defines in exactly the same way as the 'efficiency of operational management', despite the fact that they refer to different concepts. This manual uses the terms 'efficiency of policy' and 'efficiency of operational management' (see chapter 1). For a more detailed description of various problems relating to the use of the term 'efficiency' in Dutch legislation, see the Netherlands Court of Audit's report on efficiency data in budgets and annual reports.

<sup>6</sup> More specifically, the Court has powers 'with respect to legal persons, limited partnerships, general partnerships and natural persons practising an occupation or carrying on a business to which or to whom the Council of the European Union, the European Parliament and the Council of the European Union jointly, or the Commission of the European Communities has awarded a grant, either directly or indirectly, on the basis of an established programme.'



means that it is not entitled to look at any other activities performed by the same organisation).

The Netherlands Court of Audit does not publish a judgement every year on the efficiency and effectiveness of the policy pursued by the government. A lack of efficiency or effectiveness in the pursuit of policy does not in itself form grounds for the lodging of a formal objection, as it is entitled to do when performing a regularity audit, if the Court's auditors object to the financial or material management pursued, or to the associated statement (see sections 88 and 89 of the Government Accounts Act 2001).

### **2.2.3 Administrative responsibility**

The principle underlying the audits performed by the Netherlands Court of Audit is that ministers are responsible for 'acts of government' (see article 42 of the Constitution) and, in the case of audits of legal persons with statutory tasks, that the governing bodies of such entities are responsible for their actions.

A number of sections in the Government Accounts Act 2001 are devoted specifically to the responsibility of ministers (and state secretaries) for the efficiency and effectiveness of government policy. Under section 20 (1) of the Act, ministers are responsible for the effectiveness and efficiency of the policy underlying their budgets. The way in which this general ministerial responsibility translates into practical action depends on the extent to which operational matters have been devolved, either to legal persons with statutory tasks<sup>7</sup> or to local authorities. This translation of a general into a specific responsibility is an important guiding element in the Court's efficiency and effectiveness audits (see chapter 3 for further information on this aspect). Section 20 (2) of the Act states that ministers are responsible for conducting regular audits of the effectiveness and efficiency of their policy. The Financial and Economic Affairs Department plays a coordinating role in this respect (see the Financial and Economic Affairs Department (Tasks) Decree). Under section 20 (3) of the Act, ministers are required to inform the

---

<sup>7</sup> More specifically, it depends on whether autonomous administrative authorities have been given responsibility for operational matters. See the Intranet site containing information on audits of legal persons with statutory tasks for detailed information on the precise difference between legal persons with statutory tasks and autonomous administrative authorities.





Netherlands Court of Audit in good time of any efficiency and effectiveness audits they order, and of their findings.<sup>8</sup>

The extent of the responsibility borne by the management boards of legal persons with statutory tasks for the application of government money disbursed to them or for the discharge of their statutory duties (in so far as these are funded from statutory levies) that determines the scope of the audits performed on such entities (which are also referred to as 'section 91 audits').

#### **2.2.4 From policy budgets to policy accountability**

A government policy document entitled *From policy budgets to policy accountability* (VBTB) was published in the late 1990s with the aim of strengthening the policy aspect of budget statements. The management model underlying the VBTB operation is result-based: the government should take action in response to its ministries' performance and the impact their performance has.<sup>9</sup>

Of particular relevance to this manual on efficiency and effectiveness audits are the policy sections in the new budgets and annual reports. The policy articles, which form part of the policy sections, are intended to answer the following questions:

- What do we want to achieve (expressed in terms of the outcomes that policy is designed to have)? Have we achieved that which we set out to achieve?
- What are we going to do to achieve these goals? Have we done what we should have done?
- How much may it cost? Has it cost what we thought it would cost?

#### **2.2.5 Dutch Central Government Performance Data and Evaluative Studies Regulations**

The *Dutch Central Government Performance Data and Evaluative Studies Regulations* were published in February 2001.<sup>10</sup> It is explicitly linked to

---

<sup>8</sup> Comparable responsibilities in relation to operational management are set out in section 21 of the Government Accounts Act 2001.

<sup>9</sup> Van der Knaap (2000).

<sup>10</sup> The order, which was formulated in part in the wake of the Netherlands Court of Audit's report entitled *Organisation of policy evaluations* (March 2000), replaces the manual on indicators published in 1994 (section A2.6 of the Central Government Financial Information and Records Manual) and the reference framework for evaluation tools published in 1998 (section A5.5 of the Central Government Financial Information and Records Manual). The new order has also been incorporated into the Manual.



the policy document on the VBTB operation. Emanating from the Ministry of Finance, it provides more detailed information on a number of provisions in the Government Accounts Act, which also have a bearing on the Court's efficiency and effectiveness audits. Page 9 of the order contains instructions for:

- the integrated use of evaluation tools;
- the arguments that should be taken into account when instituting an *ex-ante* evaluation;
- the extent to which and the frequency with which policies should be subjected to regular *ex post facto* evaluations (i.e. completeness and periodicity);
- the methodological quality of evaluation tools and the way in which policy information is prepared;
- the way in which ministers, state secretaries and senior civil servants are informed of the findings of regular evaluations; and
- each ministry's responsibility for ensuring that this order is properly implemented.

The order takes the term 'evaluation tools' as covering both systems that provide information on standard performance data and regular evaluations. Both supply information on the efficiency and effectiveness of a minister's policy (and on the efficiency of a ministry's operational management), which is why they are a vital source of information for the Court in assessing policy information and other types of information.



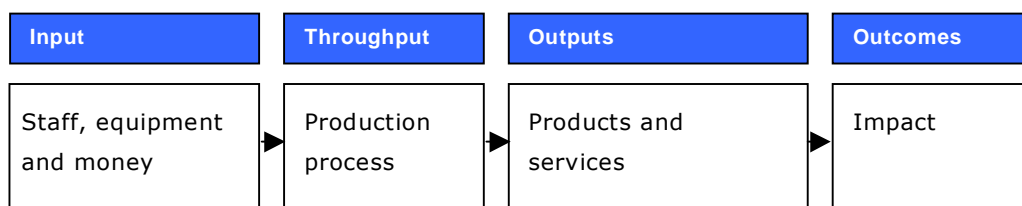
## 3 Audit subjects and types of audit

### 3.1 Introduction

The efficiency and effectiveness audits performed by the Netherlands Court of Audit are designed to answer the following questions: have the goals been achieved (in terms of outputs and/or outcomes), have the goals been achieved thanks to the policy pursued by the ministry in question, and is the policy efficient? Based on the public-sector production process, this chapter explains which aspects should be taken into account when designing audits to examine outputs and/or outcomes.

### 3.2 The public-sector production process

In its efficiency and effectiveness audits, the Court regards the work of the public sector (i.e. central government together with legal persons with statutory tasks) primarily as a production process. It is a process consisting of four aspects: input, throughput, output and outcomes. By auditing different combinations of these, you can draw conclusions about the achievement of objectives and the reaching of target groups, and on the efficiency and effectiveness of policy and the efficiency of the operational management pursued. Audits of the achievement of objectives and the reaching of target groups and of public-sector effectiveness are a matter of comparing the relevant combinations with the government's policy objectives. This has already been described in Chapter 1.



**Figure 3.1 The public-sector production process** (source: Ministry of Finance, 2001, amended; see also Bouckaert, 1999)



It is not always possible to make a clear distinction between outputs and outcomes. In many instances, there is a whole chain of outputs and outcomes, and where exactly the dividing line is drawn is a matter of agreement. If you need to make such a distinction in an audit, you should in any event come to a clear agreement with the ministry in question as to where precisely the responsibilities of the minister or the legal person's management board end.

#### *Output or outcome?*

One of the tasks (or policy objectives) of the Centres for Work and Income is to find jobs for people who are out of work. To this end, the Centres notify jobseekers of vacancies. Is this action in itself an output, or does it become an output only once an unemployed person signs a contract of employment? Or is the latter action an outcome? Also, what type of action may be construed as constituting an output or outcome? Is this when a jobseeker accepts a job, or only once he or she has been working for a specified period of time?

### **3.2.1 Outputs**

Outputs in general, and policy outputs in particular, are the results of the operating processes used by an organisation in seeking to achieve its policy objectives.

#### *Central government outputs*

Central government outputs are the results of operating processes used by a central government organisation or department that are intended for the 'outside world'; these results are related to the organisation's or department's policy objectives. Central government outputs encompass a wide range of products and services. Some are publicly 'visible' activities, i.e. activities that are perceived by citizens, civil society organisations and firms, e.g. grants, levies and public information campaigns. Others are support services provided by ministries to non-central government bodies such as local authorities, legal persons with statutory tasks and autonomous administrative authorities. Such services include enacting legislation, signing covenants, exercising supervision (for example, of the performance of local authorities or legal persons with statutory tasks), designing a computer system or collecting relevant policy information.



Some of these outputs are more internal than external and for this reason would appear to be better suited to a management review. At the same time, this type of output, such as designing a computer system for recording manure surpluses, may well be the only type of activity central government may reasonably be expected to perform in a given policy field. In which case it is also a relevant output for achieving a certain external outcome.

Support services are provided primarily in connection with policy that has been decentralised in either geographic or functional terms. In this case, the real external outputs, i.e. the results delivered to citizens, civil society organisations and firms – which form the prime targets of efficiency and effectiveness audits – are performed by non-central government bodies such as local authorities, legal persons with statutory tasks and autonomous administrative authorities.<sup>11</sup>

#### *Third-party outputs*

Third-party outputs are the results of operating processes used by non-government organisations or departments that are intended for the 'outside world'; these results are related to the central government's policy objectives. These outputs are publicly 'visible', i.e. they are perceived by citizens, civil society organisations and firms. In the case of a policy aimed at reducing the volume of waste produced by firms in the Netherlands, for example, the central government's policy outputs would include formulating legislation and putting licensing arrangements in place for local authorities. Here, the ultimate outcome, i.e. a decline in the volume of waste, would depend, *inter alia*, on the licences granted by local authorities (i.e. the outputs). Other examples of third-party outputs are the number of new police officers (i.e. an output in support of the government's security policy) and the number of places available on citizenship courses for immigrants (i.e. an output in support of the government's citizenship policy).

In many cases, a local or provincial authority also acts as an intermediary for executive agencies. This was the case, for example, with the policy on childcare facilities.<sup>12</sup> Central government distributes the money to local authorities, which pass it on to childcare centres (which are both non-

---

<sup>11</sup> See also De Groot & Goudriaan (1991) for information on the distinction between intermediate and end products.

<sup>12</sup> This ceased to be the case when the new Basic Childcare Provision Act came into effect. Funding is now provided directly through the parents.



profit-making and commercial organisations that do not have any ties with the government). The childcare centres deliver the actual output, the outcome of which is, for example, an increase in the number of mothers who remain in paid work after giving birth to children.<sup>13</sup>

Under the powers granted to it, the Court is entitled to undertake separate audits of the performance of legal persons with statutory tasks. In principle, such audits are subject to the same rules as apply to performance audits of central government bodies. What is peculiar to audits of legal persons with statutory tasks, however, is the fact that, as opposed to audits of other non-central government bodies that work in support of government policy, the Court is also entitled to audit their operational management.<sup>14</sup>

#### *Heterogeneous outputs*

The outputs delivered by both central government and non-central government bodies may be heterogeneous, which means, for example, that they may differ in both complexity and quality. Where the outputs are heterogeneous, it is important to differentiate accordingly, for example by dividing the outputs over a number of homogeneous categories (for further information on this point, see section 4.3.2.1).<sup>15</sup>

It is important for auditors to take account of the quality and complexity of the outputs they are auditing. It is possible, for example, that a rise in the volume of outputs may have an undesirable side effect in the form of a decline in their quality.

#### *Auditability*

It should generally be easier to measure outputs than outcomes. The point is that information on outputs is generally either already available or fairly easy to collect. After all, ministries may be expected to know what outputs they have delivered in implementing their own policies. The same may be expected of non-central government bodies involved in the implementation of government policy, such as legal persons with statutory tasks and local and provincial authorities. Central government will need to make arrangements with these bodies so as to ensure that it

---

<sup>13</sup> See Turksema (2000).

<sup>14</sup> More information on legal persons with statutory tasks (for example, on definitions, standards and audit proposals) may be found on the relevant Intranet site.

<sup>15</sup> See Court of Audit (2001a) for an example.



is supplied with the necessary information, which it can then process as it wishes.

### 3.2.2 Outcomes

#### *Outcomes*

*Policy outcomes* are the publicly perceptible results of policies pursued by central government or a legal entity with statutory tasks. Examples of outcomes are a reduction in CO<sub>2</sub> emissions (i.e. the desired result of the government's policy on climate change), an increase in the number of people who feel safe on the streets (i.e. the desired result of the government's policy on law and order) and an improvement in the integration of ethnic minorities (i.e. the desired results of the government's civic integration policy).

#### *Auditability*

Outcomes are hard to audit, partly because they are difficult to quantify (e.g. 'good public health in the Netherlands') and partly because it is often difficult to know whether the outcomes in question are due to the policy pursued either by central government or by a legal entity with statutory tasks. After all, other factors apart from the minister (or any other party involved in implementing the policy) may be at work in bringing about a given outcome. These include demographic trends and the situation on the labour market. Let's assume, for example, that central government wants the Centres for Work and Income to find jobs for the unemployed. Whether the Centres can find jobs for them depends not only on the activities performed by the Centres themselves, but also on the situation on the labour market (in terms of the number of vacancies) and, in tandem with this, the economic climate in general (i.e. the better the state of the economy, the more jobs there will be on offer). In order to measure the effectiveness of policy, however, you need to eliminate the influence of external factors. This is a tough task, which is why it is sometimes easier to measure outputs. This is a pity, nonetheless, which is why we would recommend that, wherever feasible, you should measure both outputs and the outcomes produced by them.<sup>16</sup> After all, outputs do not necessarily produce the desired social impact. At best, they are a necessary condition, but not a sufficient condition, for attaining the desired end result.

---

<sup>16</sup> An example of a case in which maximum attention was paid to the entire chain of outputs and outcomes is the Court's 1996 audit of subsidised labour (Court of Audit, 1996). This audit looked at a range of tools for combating unemployment and measured both outputs and outcomes, such as the creation of additional jobs, the number of people moving out of additional jobs and the number of people moving into regular jobs.



An audit opinion on outcomes is also important because it raises the public appeal and relevance of the audit in question. After all, citizens are not interested so much in outputs as in the degree to which certain social problems have been addressed or resolved. Indeed, ministers can be held to account for that.

A third reason for measuring outcomes is because of the risk that performance-driven management may divert attention away from the ultimate objective of policy. This is known as the 'performance paradox'. For example, if university funding is based on the number of graduates, this may produce a tendency for academic standards to fall. An audit of the entire chain of outputs and outcomes should identify this trend, for example because the generation of graduates in question has more difficulty finding a job after graduation (all other things remaining equal).

#### *Side effects*

Whilst you are looking at outputs and outcomes, it may be worth also taking a close look at any *side effects*, both desirable and undesirable. This is particularly relevant in the case of undesirable side effects.

An examination of side effects is complicated by its theoretically open-ended nature, i.e. it is not limited by policy objectives. One possible way of delimiting the audit is by focusing on those side effects which other government policies seek to combat (such as certain unfavourable environmental impacts caused by economic policy).

### **3.2.3 Throughput**

The term 'throughput' (i.e. activities) refers to all action taken by either a ministry or a non-central government body in converting 'input' (i.e. resources) into certain policy-related outputs. In other words, throughput relates to the *implementation* of policy rather than to the results of policy (which are expressed as outputs and/or outcomes). Examples of throughput are assessing licensing applications, attending training courses and recruiting staff.

Whilst you are examining outputs and outcomes, it may be worth also taking a close look at throughput. After all, a poor choice of activities may provide an explanation for poor outputs and/or outcomes (see section 3.3).





### 3.2.4 Input

Input consists of resources that are used for achieving outcomes or outputs.<sup>17</sup> Such resources include staff, equipment, outsourced services and expenditure on both financial and non-financial policy instruments (such as grants, loans and public-information campaigns). This applies both to central government and to non-central government bodies.

In order to measure the efficiency with which resources have been used, all the various resources need to be expressed in the same way. The most obvious way of expressing them is in monetary terms. There are two ways of doing this: in terms of costs and in terms of expenditure. Expenditure is defined as payments that are triggered by obligations. Costs, on the other hand, are the value of the resources needed for the purpose of operational management.

### 3.3 Explaining the causes

When auditing outputs and/or outcomes, you need to gain a picture of the potential *causes* of poor performance or inefficiencies. You may need information about the causes in order to make relevant recommendations to ministers, legal persons with statutory tasks and/or the House of Representatives. The causes of poor performance or inefficiencies may be either internal or external.

*Internal causes* are causes rooted in the ministry itself and/or at the non-central government body involved in implementing the government policy in question. Such causes include formulating policy objectives that are unclear or overambitious, using the wrong policy instruments, performing the wrong activities and creating insufficient conditions for effectiveness. However, *organisational characteristics* such as staffing levels and staff quality, are also regarded as internal causes.

*External factors*, i.e. factors that are beyond the control of central government or non-central government bodies and are difficult to influence, may also form the cause of poor performance or inefficiencies. Examples of external factors are insufficient support for the policy in question, administrative complexity (e.g. the complex social security system, policy that has been decentralised in both functional and geographic terms, market forces and privatisation), inadequate

---

<sup>17</sup> See also Hoogerwerf (1980): all those resources used by or available to central government or legal persons with statutory tasks for the purpose of attaining one or more objectives.



opportunities for boosting the outcome, other trends and new policies that neutralise any positive impact made by the policy in question, and economic trends.

We have compiled a matrix to present the wide range of possible explanations. The vertical axis is divided into two sections, depending on the *amount of influence central government is capable of exerting*. In fact, what we are talking about here is ministerial responsibility. The horizontal axis is also divided into two sections, one for poor performance in terms of inadequate *outputs* and the other for poor performance in terms of inadequate *outcomes*. We have placed a broad interpretation on the word 'outputs'. After all, it does not matter to the average citizen whether it is a minister or an executive agency who is not doing their job properly.

**Table 3.1: Explanations matrix**

	<b>Explanations for unsatisfactory outputs on the part of central government and associated bodies and local authorities (co-administrators)</b>	<b>Explanations for unsatisfactory social effects</b>
<b>Within the Minister's immediate sphere of influence</b>	<p><b>ORGANISATION AND IMPLEMENTATION OF POLICY</b></p> <ul style="list-style-type: none"> <li>• Mismatch between output targets, time, manpower and resources (i.e. overambitious)</li> <li>• Targets do not provide sufficient guidance</li> <li>• Targets are not SMART-C (i.e. specific, measurable, agreed upon, realistic, time-related and consistent with related targets)</li> <li>• Over-production of policy</li> <li>• Conflicting policy on the part of other ministries</li> <li>• Inadequate legal framework</li> <li>• Inadequate enforcement</li> <li>• Problems in relation to operational management: inadequate financial and material management, inadequate IT facilities, irregularities</li> <li>• Unclear tasks and powers</li> <li>• Lack of congruence in geographical divisions</li> </ul>	<p><b>QUALITY OF POLICY</b></p> <ul style="list-style-type: none"> <li>• Inadequate theoretical grounding (inadequate choice of instruments or no account taken of side effects)</li> <li>• Lack of adaptability in responding to changing circumstances</li> <li>• Unsatisfactory central government outputs</li> </ul>
<b>Outside the Minister's immediate sphere of influence</b>	<p><b>INVOLVEMENT OF AND IMPLEMENTATION BY NON-CENTRAL GOVERNMENT BODIES</b></p> <ul style="list-style-type: none"> <li>• Lack of support among stakeholders (inside/outside the sphere of influence—needs to be judged on a case-by-case basis)</li> <li>• Conflicting policy on the part of the EU/provinces/municipalities</li> <li>• Executive agency is willing but not able</li> </ul>	<p><b>FORCE MAJEURE</b></p> <ul style="list-style-type: none"> <li>• Dramatic events such as war, accidents and natural disasters</li> <li>• Economic, demographic or climatic developments</li> <li>• Conflicting policy on the part of the EU/provinces/municipalities</li> </ul>



The four categories in the matrix are not mutually exclusive: there is some degree of overlap, for example with regard to the presence of sufficient stakeholder support (i.e. is the minister responsible for this, or are the interests of the various stakeholders so disparate that it is not reasonable to expect the minister to be able to exert any direct influence over them?). The implication continues to be that a failure to deliver outputs may also explain why certain outcomes were not achieved.

### **3.4 Audit scope and methods**

Based on a number of arguments, i.e. technical, strategic and managerial, the team will have to decide which particular type of audit to undertake. This section discusses a number of aspects that may play a role in taking this decision.

#### **3.4.1 What steps should the audit involve?**

An audit starts with a set of policy objectives, i.e. what does the minister in question wish to achieve (i.e. the *desired* outputs and outcomes) and has the policy objective been formulated in SMART-C terms (i.e. specific, measurable, agreed upon, realistic, time-related and consistent with related objectives)<sup>18</sup> The next step involves acquiring information on the *actual* outputs and outcomes, preferably by using data already available, either at the ministry or elsewhere.

The data available at the ministry (i.e. policy information) will not necessarily always be suitable for reaching a judgment on the nature of the outputs and outcomes delivered by the ministry. In that case, the question is whether the Court should try and collect data itself. It is important to bear in mind in this connection that collecting and analyzing data on outputs and outcomes is extremely *time-consuming*. At the time when an audit proposal is being compiled, first the team and subsequently the management and/or the Board need to decide whether the audit is important enough (for example, in terms of its added value) to merit the deployment of a large number of staff.

It may also be sufficient, based on the findings made at the time when the audit proposal was compiled, to urge the minister in question to improve the quality of the policy information, and to target the audit at this particular aspect.

---

<sup>18</sup> See Chapter 5.



If information is available and is also of good quality, you can use it to make judgments, for example, about the outputs and/or outcomes achieved by central government or non-central government bodies.

### **3.4.2 How do you decide which type of audit to perform?**

When measuring outputs and outcomes, you need to choose from the four different types of audit discussed above, i.e. audits of whether the objective has been achieved, audits of whether the target group has been reached, effectiveness audits and efficiency audits.

Which of the four types you choose depends on the following factors, among others:

- the political context;
- the quality of the policy objective that is the subject of the audit;
- the availability of audit data and policy information;
- the complexity of the field or the length of the policy chain;
- the nature of the audit (is it designed to stimulate the auditee?);
- the availability of sufficient numbers of audit staff with the right qualifications.

*When should you simply seek to establish whether the objective has been achieved?*

An audit to establish whether or not certain objectives have indeed been achieved may be particularly relevant if there are doubts about this and if this is a matter that crops up in the debate on government policy, either past or future. Did the crime rate go up or down in a given period? Was the air cleaner or dirtier in 2004 than it was in 2000? Did more or fewer children leave school without qualifications in 2003 compared with the previous year? Apart from in such broad social terms, the question of whether the government has achieved its objectives may also be important in a situation in which the focus lies on the products or services that have been delivered (i.e. the outputs). The political context may be such as to make it worth finding out how many kilometres of dykes have been inspected in a given period of time, how many hours teachers have taught, how many hours police officers have spent on patrol and how many incapacity benefits have been claimed.

It frequently proves extremely difficult to find answers to such apparently simple questions, even though policymakers and/or members of



parliament often base their pronouncements on certain assumptions in this respect.

If the relevant information is not traceable in the policy information in the form of carefully processed data, this is a finding in itself. If you wish to take things a step further, you can then try and build up a clear picture using information from various sub-files, archives and previous audit reports (see section 4.3.2.1).

If, again, this is not possible and the Court is nevertheless keen to do more than simply signal the inadequacy of the policy information, you will have to collect the information yourself. In most cases, you will need relatively large quantities of data in order to answer these questions, particularly if you are interested in identifying wider social trends. A survey is a useful tool in this connection (see section 4.3.2.2).

If your audit is restricted to establishing whether or not the intended objectives have been achieved, you cannot of course say anything about the degree to which the policy played a role in this. At the same time, this is a question which many readers of your report are bound to ask and it will prove difficult to prevent them from reading more into your words than you actually intended. For this reason, you may find it useful to ask yourself, particularly if you are collecting data anyway, whether it is not worth going one step further by performing either an effectiveness or an efficiency audit.

*When should you simply seek to establish whether the target group has been reached?*

The same principle applies to an audit aimed at establishing whether the target group has been reached: this is particularly important if there are serious doubts as to whether major public information campaigns and news on government schemes have reached the relevant target group. Even if the target group has been reached, there is still no guarantee that the campaign or scheme in question is actually having an effect. However, there can be no doubt that the campaign or scheme is not working if the target group has not been reached.

If you find to your surprise that the target group has indeed been reached, this is the point at which you should think again. Is this a sufficient finding, or do you want to take a closer look at the outcomes achieved, in terms of changes in knowledge, attitudes and behaviour?



Audits that are aimed solely at establishing whether the target group has been reached will generally use some form of survey. If you wish to go further and measure outcomes, you will need to perform an effectiveness audit. As described below, what you need to do in this sort of situation (difficult though this may be) is to obtain comparable data on the situation before and after the campaign (or, as it were, with and without the campaign).

*When should you perform an effectiveness audit?*

The Court performs effectiveness audits that involve more than simply processing existing policy information only if:

- the relevant answers cannot be deduced from existing policy information;
- the relevant ministry or other parties are not likely to perform or commission a comprehensive study along these lines in the near future;
- there are pressing social and/or financial reasons for establishing whether the policy in question is having an impact; and
- the Court is capable of performing the audit in question.

Obviously, these are all points that you should take into account when writing your audit proposal.

All effectiveness audits should seek to answer the following two questions:

1. Has the policy in question actually had the effect it was intended to have? (This is the same question as we already discussed in relation to audits aimed at establishing whether the government's objective has been achieved.)
2. Is the fact that the objectives have or have not been achieved due to the policy in question?

In order to answer these two questions, you need to obtain information on the situation or behaviour which the policy is designed to influence, preferably at different points of time (for example, the number of young people attending training courses before and some time after the introduction of government incentives). You also need to build up a clear picture of any other factors that can affect the situation (such as the state of the labour market) and find a way of distinguishing between the



influence exerted by the policy in question and the influence exerted by other factors.

Making *comparisons* between different points of time (i.e. before and after the implementation of the new policy) and/or different groups of people (i.e. whom the policy has affected to differing degrees) is a vital factor in answering the above questions. We shall be returning to this point in more detail in section 7.4.

The best way of making comparisons is by performing an experiment, i.e. by making comparisons both between different points of time and between a range of similar groups (see section 4.3.2.2).

This is generally not feasible in practice, however, in which case you will have to fall back on less elegant solutions, such as a survey conducted at a specific point in time and a comparison with other information on the situation before the introduction of the new policy. There is, however, a risk that the choices made by your predecessors in obtaining their information, i.e. in terms of the type of questions asked and the respondents they chose to use, are less relevant to the questions you yourself are seeking to answer.

What you can also do is to *ask* people about the situation prior to the introduction of the new policy. In which case, of course, you run the risk of memory loss and of people giving you the answers they believe you want to hear.

Whilst there is no catch-all solution to the problem, there are various ways in which you can limit the risks. Basically, it's a question of knowing what you are comparing with what. In other words, you should pay close attention to the composition of the groups you wish to compare, and find ways and means of ironing out any material differences between them. You should ask the Performance Audit helpdesk at the Bureau for European and Government-wide Performance Audits (ERDMO) for advice on how best to tackle this type of audit.

There are various techniques for finding out whether certain changes are the result of a new policy and not of other factors, and some of these are complementary. First of all, you can design the audit in such a way that you limit the role played by other factors to a calculable probability. You can do this by comparing two groups that are taken at random from a larger group (for example, by tossing a coin). One group is then 'exposed'





to the policy, and the other is not. This is a genuine form of experimentation.

Quasi-experiments and natural experiments are a second technique. These differ from genuine experiments in that respondents are not allocated at random to the experimental and control groups. The advantage of this type of experiment is that it can be performed as soon as the new policy has been put into effect, which enables information on the impact of the policy to be obtained relatively cheaply.<sup>19</sup>

There is a third possibility, and this involves first thinking very carefully, for example by reading as many audit reports as possible, about other factors that might play a role and then seeking to establish systematically whether they do indeed play a role. Appendix 6 contains an example of this type of systematic approach. Because it is not possible to conduct a genuine experiment when auditing a policy that has already been introduced, you will generally find yourself using the second technique in practice, i.e. quasi-experiments and natural experiments.

#### *When should you perform an efficiency audit?*

The same sort of conditions apply to a decision to conduct an efficiency audit as apply to effectiveness audits. The Court undertakes efficiency audits only if:

- a particular question about potential efficiency gains is of crucial importance;
- sufficient relevant information is available on efficiency;
- no one else is planning to undertake an efficiency audit on the same subject in the near future;
- the Court is capable of performing the audit in question.

Efficiency audits are even more advanced than effectiveness audits. An efficiency audit is designed not simply to measure the outcomes achieved by the policy in question or the outputs it has produced; it also seeks to establish a link with the *resources* (or input) used for this purpose.

Here too, the key ingredient of the audit is *comparison*. This is because efficiency is relative in two senses. First, it's a matter of the mutual relationship between the resources used, the volume of outputs or

---

<sup>19</sup> A report published by the Netherlands Bureau for Economic Policy Analysis entitled *Lerend beleid: het versterken van beleid door experimenteren en evalueren* ('Towards evidence based policy', 2004) contains further information on this type of experiment.



outcomes and their quality. In order to generate information on efficiency, all these various buildings blocks must be related to each other. Second, efficiency can be measured only in relation to a point of reference, i.e. an organisation can only be more or less efficient when compared with another organisation or a point in the past.

As will be described in detail in Chapter 8, you can sometimes perform this type of audit with the aid of indicators. However, there are also more complex econometric techniques. A complicating factor in this respect is the fact that you need to have at least 50 comparable observations as a basis.

### **3.4.3 Ministerial responsibility**

One of the basic issues in connection with efficiency and effectiveness audits is the scope of the minister's responsibility. Where policy has been decentralised, in either geographical or functional terms, the minister is not fully responsible for the results of government policy. In such cases, the minister is said to bear an *indirect* responsibility.

According to the Ministry of Finance, a minister is indirectly responsible for the results of government policy if he or she is responsible only for putting in place the conditions that are needed to facilitate the desired results, and if other actors bear a direct responsibility for the results themselves. Where, for example, the policy article in the budget does not explicitly state that the minister is not responsible for all the desired results of government policy, the assumption is that the minister bears both a direct and an indirect responsibility.

At the same time, a minister who bears an indirect responsibility is at all times jointly responsible for assessing the results which the general policy objective seeks to bring about and which are produced by the system. Changes in policy problems and the pattern of results are the subject of monitoring and evaluation, followed where necessary by policy changes agreed in government consultative groups.

### **3.4.4 How do you deal with decentralised policy?**

The Court's brief is to assess *central government* outputs and outcomes. In order to do so, however, it may be both desirable and necessary to look at *local* outputs and outcomes. Where the implementation of central government policy is either incomplete or poor, for example, this may be due to local causes. The message is directed at the minister, however, as



it is the minister's job to ensure that local authorities deliver certain outputs or achieve certain outcomes. The question is: is the minister doing enough to bring this about?

What the Court in fact does is to inform the minister about the local situation. It provides the minister with relevant policy information on a certain aspect of the overall policy chain. The Court can use this information to urge the minister, as part of his or her indirect responsibility, to encourage local actors to make certain improvements.

The explanation the Court finds for poor performance by central government, such as ineffective central government outputs, leads it to form a judgement on the minister. This is commensurate with the Court's role as an assessor of the performance of central government (and legal persons with statutory tasks). The explanation the Court finds for poor performance by local authorities may lead it to urge the minister to adopt a policy aimed at removing the impediments in question.

#### **3.4.5 How should you audit outputs and outcomes for which non-central government bodies are responsible?**

In order to audit the outputs of geographically decentralised policy and its outcomes, the Court may decide to obtain information directly from the relevant local actors. In this case, you should proceed as follows:

- If local authorities are responsible for implementing a certain government policy, you first need to establish where the minister's own responsibilities precisely lie. The audit should tie in with the minister's responsibility. You then need to establish what the minister has done (i.e. what outputs he or she has generated) in order to ensure that the local authorities can effectively implement the policy. You should find out at the same time whether the minister has collected sufficient policy information on these 'central government outputs'.
- If the implementation of government policy has been decentralised (in either functional or geographical terms), you should find out whether the minister has access to high-quality information on how the implementation of the policy is progressing (including data on outputs and outcomes). If the minister does not have such information, the Court may decide to collect the information itself so that it can draw well-founded conclusions on the degree to which



non-central government bodies are achieving certain outputs or outcomes and on whether the policy is effective. Among the audits in which this has been done are those on the following subjects: preventing and combating juvenile crime, preparing disaster plans, and counselling and reintegrating school dropouts.

- One possible explanation for a failure on the part of local authorities, for example, to achieve certain outputs or for an absence of social outcomes, may be that the policy pursued by central government (i.e. the central government output) has not had any impact on local authorities. When seeking to identify possible causes of inadequate outputs and outcomes, you should start by examining this possible explanation. After all, it is the minister who is the prime target of the audit.

In conducting performance audits of local and provincial authorities, you are dependent on their voluntary cooperation, given that the Court does not have any mandatory powers of audit in relation to such authorities. To date, there have not been any problems in obtaining this cooperation. This is because the audit teams are able to guarantee the anonymity of individual actors (as, for example, in the case of the audits carried out on the following subjects: government policy on the big cities, counselling and reintegrating school dropouts, and preparing disaster plans).

It may well become more difficult to obtain the voluntary cooperation of local and provincial authorities if the Court finds itself stepping up the audit frequency. For this reason, it is important to bear in mind the possibility of working together with local audit offices, to proceed with great care when asking people and organisations for their assistance, and to honour any pledges of anonymity.

In more general terms, any performance audit that includes an assessment of the performance of a local or provincial authority should be aimed at forming an opinion on the relevant minister. In order to do so, you need to obtain information on the role played by central government in helping the relevant local and provincial actors deliver the outputs in question.

Other points on which the minister may be held accountable are:

- the quality of the information available to him or her on whether or not the preset objectives have been achieved (including the



- information provided by local and provincial authorities, or which such authorities are supposed to provide);
- the extent to which, in the light of the preset objectives, he or she has been encouraging the local and provincial authorities to produce certain outputs;
  - in more general terms, the extent to which he or she is working to create a system that can ensure both that the relevant outputs and outcomes can be delivered (i.e. the degree to which he or she is putting the necessary conditions in place) and that information on these outputs and outcomes is collected.



## 4 Audit strategies

### 4.1 Introduction

Section 4.2 briefly discusses how to describe the objective of your audit and define the problem you are seeking to address. The success of an audit depends critically on the care with which you define the audit objective and problem. Depending on the way in which the problem is defined and on the type of audit you have chosen to perform (i.e. an audit of whether the objective of the body in question has been achieved, whether the target group has been reached, an effectiveness audit or an efficiency audit), there are various methods – some exploratory, others geared more towards testing – that you can use for making pronouncements about the auditee’s success in delivering certain outputs and outcomes, and about the reasons for this. These methods are discussed in section 4.3. In section 4.4, finally, we return to the ‘explanations matrix’ in section 3.3, setting out which audit methods are suited, to greater or lesser degrees, to each type of explanation.<sup>20</sup>

### 4.2 Objective, problem and audit questions

The nature of the audit, and the findings it generates, depend partly on the way in which the problem is defined. In other words, this is a vital aspect of the audit process and deserves to be given your full attention. At the same time, it is clear from past experience that many auditors do not take the trouble to clearly define either the problem or the objective. In fact, this may well be one of the most frequently overlooked aspects of the audit process.

The absence of clearly defined problems and objectives is one of the main causes of problems in the audit process. For example, the absence of clear definitions often leads to delays, as auditors tend to widen the scope of their audit to cover more and more issues in an attempt to perform a ‘comprehensive’ audit.

---

<sup>20</sup> Part of the text of this chapter is taken from the course entitled ‘Audit methods and techniques’.



#### 4.2.1 Objective

The objective of an audit says what the auditor in question wishes the audit to achieve. The definition of the objective explains why the audit is being performed in the first place. In general terms, the objective of a Court audit describes the social *impact* the Court wishes to bring about. In many cases, the initial objective is to ensure that the responsible minister and/or the organisation in question takes certain *action*. Objectives formulated along these lines are referred to as 'action-driven objectives'.

Examples of typical objectives for Court audits include 'raising the quality of the financial management pursued by a legal person with statutory tasks' and 'helping to improve the efficiency of a particular department'.

#### 4.2.2 Problem

The problem is the question to which the audit seeks to provide an answer. The definition of the problem is a general description of the points covered by the audit. The problem at the centre of the audit is defined in the form of a question, e.g. does the policy sufficiently encourage those locally responsible for implementing the policy to achieve the desired results? Is the policy capable of solving the problems that have been identified and what sort of hindrances are likely to be encountered? (Court of Audit, 2003b.) In other words, the problem definition may be said to be 'knowledge-driven'.

The purpose of the problem definition is to make the audit less complex by restricting it to certain issues without oversimplifying it. The problem definition helps to guide the auditor during the audit.

The problem definition cannot be divorced from the objective of an audit. The answer to the question or questions making up the problem should, together with the resultant recommendations, help the auditor to achieve the objective. The link with the objective is more than just a formality, but is a matter to which proper attention needs to be paid. If an audit is intended to help raise the efficiency of a particular government policy, both the problem definition and the nature of the audit should be geared to achieving this end.

Obviously, the objective and the problem definition should also tie in with each other in the sense that both should relate to the same subject.



Logical though this may sound, we know from experience that this is not always the case.

#### 4.2.3 Audit questions

Audit questions follow from the problem definition, with which they have a logical relationship. They should seek to produce a solution to the problem. Detailed audit questions can help prevent the audit from becoming broader in scope than the problem itself. Although it is tempting to add a number of interesting audit questions, on the off chance that they may generate some interesting answers, there is no real need to broaden the audit remit in this sort of underhand manner if the problem definition itself covers all relevant aspects.

There is no need for all audit questions to be formulated along the same lines as the problem definition. Descriptive questions and questions about the relationship between variables are required in order to establish causal relationships.

Below follow a number of tips on how to formulate audit questions:

*Structure* the audit questions, for example by restricting each question to just a single aspect of the problem. This should allow each aspect to be investigated separately. The findings on each individual aspect can then be combined to provide a solution to the problem as defined.

Check whether the various audit questions, when combined, add up to an *overall answer* to the problem as defined.

Do not ask *too many* audit questions. Most problems can be solved by asking between two and six questions. If you find that you have more, you should ask yourself either whether the audit questions do not extend beyond the problem as defined, or whether your definition of the problem is not too wide-ranging and requires an audit that is correspondingly too wide-ranging.

Your audit questions should *not be too detailed*. Have you formulated two separate questions, for example, that should in fact be combined to form a single question?

#### 4.2.4 Hypotheses

The next step is to translate your audit questions into hypotheses that are **open to testing**. The basic principle here is that you should choose between a zero hypothesis (H0) and an alternative hypothesis (H1). Both follow from certain suppositions about the occurrence of a given phenomenon within the overall population. When choosing between H0





and H1, you should bear in mind that H0 usually stands for a negation of the supposition, whilst H1 corresponds with the supposition itself.

For example, take a supposition that the average amount of time customers spend waiting at a government counter on Mondays is not the same as the average waiting time on other weekdays. H0 is then the negation of this supposition, i.e. 'the average waiting time on Mondays is the same as on other weekdays'. H1 is the supposition itself, i.e. 'the average waiting time on Monday is not the same as on other weekdays'.

Please note that the hypotheses do not contain any pronouncement about the nature of the difference. If you suspect that the average waiting time on Mondays is longer than on other weekdays, this should be reflected by H0 and H1. H0 might then read as follows: 'The average waiting time on Mondays is no longer than on other weekdays'. H1 would read: 'The average waiting time on Mondays is longer than on other weekdays'.

#### **4.2.5 Other details on the nature of the audit**

Various other details are required in order to make clear what exactly it is that you intend to audit. These are:

1. definitions of any terms you intend to use;
2. the units of analysis;
3. the variables;
4. further limits placed on the scope of the audit (i.e. in time and in geographical terms);
5. the type of pronouncements the audit is expected to generate;
6. the main standards applied.

You will have to decide on a case-by-case basis which of these details you include in your problem definition, audit questions and hypotheses, and which you include in the form of a separate section.

### **4.3 Audit methods**

The following sections discuss the various audit methods, which are characterised as exploratory techniques (section 4.3.2.1), causality assessment techniques (section 4.3.2.2) or a combination of the two (section 4.3.2.3). First, however, we wish to examine two points that have a vital bearing on the choice of audit method, i.e. causality conditions and measurement scales.



### 4.3.1 Choice of audit methods

#### 4.3.1.1 Causality conditions

Causality conditions come into play if you are intending to audit the effectiveness of government policy. There are three causality conditions that need to be fulfilled in order to perform an effectiveness audit (Swanborn 1987, pp. 294-295):

1. the presence of a statistical link;
2. the time sequence (i.e. the explanatory factor must precede the factor that requires explaining);
3. the presence of other causes.

#### *Is there a statistical link?*

The first condition is easy to test. There are various ways of confirming the presence of statistical linkages, three of which are explained in brief below:

- *Comparing two different groups.* In this case, the average value (or another relevant quantity) of variable Y (for example, the possession of information on AIDS) in relation to a group that has been exposed to a given instrument (such as a public information campaign on television) is compared with its average value in a group that has *not* been exposed to the same instrument. If people who have seen the spots of the campaign are significantly better informed than those who have not seen them, this suggests that there is a statistical link between the public information campaign and people's knowledge of AIDS.
- *Comparing within a group.* The idea in this case is to see how the same group scores in relation to variable Y (e.g. an increase in their knowledge of AIDS) before and after the deployment of a given instrument, such as a TV commercial.
- *Comparing scores in relation to two variables.* Here, the idea is to calculate the value of a correlation measure between variables X (such as the subsidy paid on a particular product) and Y (the tendency of consumers to buy the product in question). If the correlation coefficient between X and Y is high, there is a statistical link between them.

#### *Does the explanatory factor precede the factor that requires explaining?*

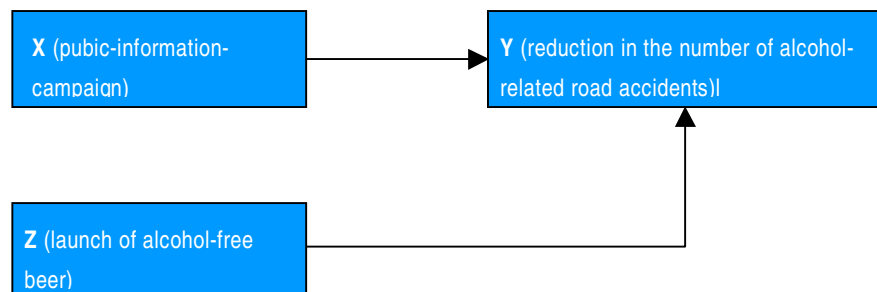
The *second* condition is rather more difficult to test, particularly in the case of surveys. The point is that it is not always clear whether explanatory factor X (such as the volume of arms purchases in a



particular region) indeed precedes factor Y, which is the factor that needs explaining (such as the frequency of armed conflicts in the same region). Without accurate information on dates, it is impossible to know which factor affects the other. The problem does not arise if you use explanatory factors that are never exposed to social influences, such as age or gender. But these do not usually lie at the core of policy audits. In which case you need to use logical arguments to prove that factor X affects factor Y rather than vice versa.

*Are there any other causes?*

The *third* condition, i.e. that the linkage established may not be caused by other factors, is also difficult to test. For example, it is possible that a link found between policy aim Y (e.g. a reduction in the number of alcohol-related road accidents) and policy tool X (e.g. a public-information campaign designed to reduce the consumption of alcohol) may have been caused by the interfering factor Z (e.g. the launch of a brand of alcohol-free beer during the period in which the campaign was running).



In such an event, it might be necessary to adjust policy tool X. For example, the public-information campaign could be altered to stress the benefits of drinking non-alcoholic beer for those who are planning on driving home by car.

There may also be a factor that has a more direct bearing on the link, for example if motorists start drinking less alcoholic beer and more non-alcoholic beer. If this contextual factor Z therefore affects both policy tool X and policy aim Y, there is said to be a correlation between the interfering terms of the comparisons (i.e. *multicollinearity*), thus making it impossible to make any pronouncements about the link between X and Y.



Analysing the score in relation to Y within the various combinations of X and Z (i.e. those who have and have not seen the campaign, and those who do and do not drink alcohol-free beer) may shed more light on the issue. If such an analysis indicates that there is nevertheless a negative correlation within the group of alcohol-free beer drinkers between their degree of exposure to the public-information campaign and their involvement in road accidents, this means that the effect of the campaign cannot be explained away by the launch of alcohol-free beer (and vice versa).

#### 4.3.1.2 *Measurement scales*

Which audit method you choose also depends on the measurement scale of the variables you use. The outcomes and outputs you wish to explain, plus the associated explanatory factors and control factors,<sup>21</sup> all need to be converted into *measurable*, empirical quantities. Some factors are relatively easy to operationalise (i.e. 'the deployment of human resources' = 'number of staff'), whilst others need to be measured by indirect means.

Outcomes in particular are often hard to measure directly. In such cases, you will need to use indicators to place an approximate value on them. An outcome indicator of road safety, for example, is the figure for the reduction in the number of road fatalities. Similarly, the number of highly qualified new business owners who devote a large amount of time and energy to R&D may be regarded as an indicator of the number of innovative entrepreneurs.<sup>22</sup> This type of indicator is referred to sometimes as a 'proxy variable'.

Whilst outputs are generally easier to measure than outcomes, they are not always open to direct measurement. In such cases, you will also need to use an indicator or proxy variable.

#### **Some outputs are difficult to measure**

An example of an output that is hard to measure is the supervision exercised by inspectorates. Indirect outputs (or proxies) are often defined in such cases; these might include the number and duration of the inspections performed, the number and duration of the contacts made by telephone, email and mail, and the number of audit reports published about supervisory activities.

<sup>21</sup> Control factors: factors that may be assumed to have a link with the factor that you are seeking to explain, but in which you are not fundamentally interested. Ignoring such factors tends to distort the analysis of the outcomes you are interested in measuring.

<sup>22</sup> Court of Audit (2002c).



Measuring requires attaching values to the variables you have constructed. Variables are classified as nominal, ordinal, interval and ratio variables. These measurement scales are relevant to the way in which variables are used in descriptive and explanatory analyses: the fact is that the type of audit method you can use depends largely on the type of measurement scale you use.

The *measurement scales* are:

- **Nominal:** if two scores for a nominal variable vary, this means that one score *differs* from the other score. Examples of nominal variables are gender, the target group at which a particular government scheme is aimed (i.e. individual citizens, firms, non-profit-making organisations, etc.). In other words, the variables can be measured only in terms of whether the individual items belong to certain distinctive categories. Even if values or codes are attached to the categories, they still cannot be quantified.
- **Ordinal:** if two scores for an ordinal variable vary, this means that one score is *higher* than the other. An example of an ordinal variable is the opinion held by citizens on the government's customer-friendliness. Here too, the differences cannot be translated into figures. A higher score suggests only that one item has 'more' of the quality represented by the variable.
- **Interval:** if the scores for an interval variable vary, this means that the *difference* (i.e. the interval) between scores a and b is *x times as great as* that between scores y and z. For example, the difference between a temperature of 40 degrees Celsius and 20 degrees Celsius is twice as great as the difference between 15 and 5 degrees. However, a temperature of 40 degrees Celsius is not 40 times as hot as a temperature of 1 degree Celsius; and it is completely impossible to say how much hotter it is than a temperature of -20 degrees Celsius (unless you convert to the Kelvin scale). In other words, using interval variables means that the only ordinary figure you can work with is the *difference* between the scores.
- **Ratio:** If two scores for a ratio variable vary, this means that one score is *x times higher than* the other score. For example, one firm's monthly sales may be 5,000 times higher than the figure posted by a smaller competitor. Typically, a ratio variable should



have an identifiable *absolute* zero point, whereas an interval variable has a *randomly* chosen zero point. In other words, you can treat the score as an ordinary figure.

Whereas a score on the ratio or interval scale can be converted into a score on the ordinal or nominal scale (although some information is lost in the process), it is not possible to do the opposite.

Your aim should be to measure at the highest possible level, because a higher level score is more revealing than a lower level score. For example, if you record the speed at which people read an information brochure in terms of a number of pages per minute (i.e. using a ratio scale), this gives you much more information than simply classifying the reading speed in terms of 'fast, medium or slow' (i.e. using an ordinal scale). This also means that you can use more analytical techniques on a higher measurement scale than you can on a lower measurement scale.

#### **4.3.2 Methods<sup>23</sup>**

##### *4.3.2.1 Exploratory techniques*

- *Literature searches and meta-research*

A literature search is a valuable way of obtaining information about a specific subject, of demarcating the audit subject and of formulating and demarcating the audit question. For this reason, it is particularly useful during the start-up stage of an audit, i.e. when you are preparing the audit proposal and performing in-depth monitoring.

Although a literature search is particularly suited for refining an audit question, it is important to start off the search with a carefully defined search query, in terms of both the subject matter and the document type. You will find it useful to find a point of departure in the form of a publication mentioning the subject, relevant authors and/or other interested parties. Be careful that you don't end up reading everything you come across as you search blindly for literature. Instead, make good use of subject-matter specialists and documentalists and take the trouble to refine your query if necessary. A literature search is a vital ingredient in the process of formulating the audit subject and the relevant audit questions. Given that, by definition, written sources provide a dated and

---

<sup>23</sup> See Annexe 2 for further information on surveys as a data collection technique.



subjective account of events, trends and phenomena, it is always worth supplementing a literature search with other types of research.

Although literature searches are sometimes referred to as 'meta-research', the two terms are not entirely synonymous. Meta-research may comprise a meta-evaluation and an evaluation synthesis. A meta-evaluation is an evaluation of already existing evaluations and is intended to answer the following questions: what sort of variations are there in the results of the evaluations, and what is the reason for these variations? An evaluation synthesis, on the other hand, seeks to aggregate (i.e. summarise) the results of the various evaluations and to use this information as a basis for making a pronouncement on a particular subject. Before you can perform an evaluation synthesis, however, you first need to have performed a meta-evaluation so as to be sure of the quality of the primary sources you are using.

Primary aim: orientation. Obtaining information on a particular subject, when demarcating the audit subject and formulating and demarcating the audit question.
---

For further information: P.G. Swanborn (1999). <i>Evalueren</i> . Meppel: Boom; <i>Handleiding metaonderzoek</i> , Netherlands Court of Audit, 2003 (see Intranet, Performance Audits Information Page).
--

o *Focus groups*

Research using focus groups stems from practices in market research in the 1920s. It is based on a group interview on a specific subject. The interview is open-ended, and is conducted by a researcher. The number of participants ranges from 8 to 12, and the group in question may be either homogeneous or heterogeneous in composition. Both types of group have their own pros and cons. As an alternative, you can also work with a number of different focus groups. It goes without saying that the use of focus groups is associated with certain methodological hazards (i.e. group dynamics, reporting techniques and the quality of the facilitator). The fact that focus groups have been around for such a long time, however, means that there has been time to develop a large number of guidelines for their effective use.

Primary aim: generating ideas for possible explanations; orientation
--

For further information: Colin Robson (2002). <i>Real world research</i> . Blackwell Publishing. P.G. Swanborn (1999). <i>Evalueren</i> . Meppel: Boom. <a href="http://www.evaled.info/frame_downloads.asp">http://www.evaled.info/frame_downloads.asp</a>
--



- o *The Delphi method*

The Delphi method is a means of amassing expert opinion on complex problems, arriving at a consensus and identifying any points on which opinions differ. Those attending Delphi 'meetings' do not actually meet in the flesh: they answer researchers' questions in writing and then read the (anonymous) summaries made of all the answers received. This is followed by a second (and generally final) round of questions.

This method is often used not just to establish the current trends in a particular field, but also to arrive at a consensus on the conditions that need to be put in place and the activities that need to be performed in order to ensure that certain trends can continue, such as the impact of the IT revolution, with applications such as teleworking and distance learning. The Delphi method is also used to understand terms with normative connotations (such as environmental quality and spatial quality), the underlying values and the associated policies. The Delphi method always involves a systematic process of questioning, data analysis, feedback, discussion and decision-making.

Among the 'qualitative' characteristics of the Delphi method are the emphasis on naming terms and categories, and the fact that it generally includes a lengthy problem-analysis stage. At the same time, both the questions asked and the analysis of the answers given to them need to meet the customary criteria applying to surveys and the analysis of survey findings (such as the need for avoiding socially desirable answers, for arriving at a well-distributed response and for careful coding).

A number of years ago, the Netherlands Court of Audit trialled the Delphi method during an audit performed by the Court's Health, Welfare and Sport Bureau.<sup>24</sup>

Primary aim can be used for both testing and orientation purposes.
--

For further information: M.C. Groenberg (1988). <i>Kwaliteit in beeld, een delphi-onderzoek naar milieukwaliteit en ruimtelijke kwaliteit in het landelijk gebied</i> . Nijmegen: Catholic University of Nijmegen.
--

P.G. Swanborn (1999). <i>Evalueren, het ontwerpen, begeleiden en evalueren van interventies: een methodische basis voor evaluatieonderzoek</i> . Meppel: Boom.
--

<a href="http://www.evaled.info/frame_downloads.asp">http://www.evaled.info/frame_downloads.asp</a>
---

<sup>24</sup> *Verantwoorde cijfers in de zorg* ('Accurate reporting in the healthcare sector') (2000).





- *Case studies*

A case study is a research technique that enables the researcher to gain an in-depth understanding of one or more objects or processes that are limited in time and/or space. A case study is characterised by a small number of research units, a labour-intensive research technique, a preference for depth rather than breadth and a well-argued, well-informed choice of research units. An important and commonly used type of case study is that based on triangulation: this involves using a series of sources (such as semi-open interviews, literature searches, observations, file research and so forth) in order to arrive at reliable observations that are independent of each other. The effect is that the researchers acquire a greater understanding of the research object.

Within the field of policy studies, case studies have provided a popular framework for studying policy processes in relation to their institutional context. Basically, there are two types of case study: single case studies and comparative case studies. In the former instance, the research object is a particular scheme, country or group of items. The aim may be to undertake an in-depth study of the object, but equally the researcher may be interested in the way in which the case in question is illustrative of a wider field.

Police forces in the Netherlands have been using performance-based contracts for some time now. Under the terms of these contracts, the police are required to meet certain targets in a number of areas, many of which are quantified. The contracts are based on the National Police Covenant and the various regional agreements reached on the basis of this Covenant. The thinking behind them is not only that there is a need for the police to be more accountable, but also that the introduction of certain incentives may help to raise the standard of police performance (the police qualify for certain financial rewards if they meet the targets set out in the contracts). The overriding aim is to create a greater sense of public safety by raising the visibility of police officers.

An example: a study of performance-based contracts

Let's assume that the Netherlands Court of Audit wishes to audit these performance-based contracts and is interested in finding out what effect they have on the work of regional police forces. In designing a single case study, the auditors may choose to focus on either a critical or an extreme case. In the former case, the study focuses on a particular case (for example, a regional police force) that allows the auditors to make plausible pronouncements on the research object as a whole, i.e. the Dutch police force. For example, the auditors could choose the police force that has the most experience in working with performance-based contracts. If this force has encountered certain problems that are associated with the use of performance-based contracts, this would suggest that the situation is probably similar among forces that have less experience with such contracts. In the latter case, the auditors select a regional



police force that, in one way or another, differs very sharply from the other police forces. For example, there may be frequent reports in the media suggesting that a certain regional police force is highly dissatisfied with performance-based contracts. A case study would generate valuable information on the *potential* problems affecting performance-based contracts, without necessarily enabling conclusions to be drawn about the situation among other police forces. It is important to note that, in both cases, the auditors need to possess a certain amount of prior knowledge, which can be obtained by means of a literature search or by conducting interviews with experts.

Primary aim: orientation.

For further information: Flyvbjerg, B. (2000). *Making social science matter: why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.

P. Verschuren (2003). Case study as a research strategy: some ambiguities and opportunities, in: *Journal of Research Methodology*, vol. 6. no.2, 121-139.

P.G. Swanborn (1996). *Case-study's, Wat, wanneer en hoe?* Meppel: Boom.

C. Robson (2002). *Real World Research*. Blackwell Publishing.

R. K. Yin (2002). *Case Study Research, Design and Methods*, 3rd ed. Newbury Park: Sage Publications.

As an alternative to a single case study, you may decide to perform a comparative case study. See section 4.3.2.3.

Analysing the way in which an organisation has defined its processes and how these processes are performed is a means of conducting exploratory research into explanations. The fact is that, although the type of data collection performed during an operational audit or an operational management audit consists of a number of assessment-based activities, the audit does not set out to test a hypothesis. The output of the audit is the knowledge that an organisation either does or does not observe certain defined procedures. Obviously, whilst this may be indicative of the reason why the organisation has failed to generate certain outputs, no guarantees can be given about this. After all, the audit has not tested whether the outputs are delivered by following the prescribed procedures. No link is established between procedures and results.

In essence, the two types of audit described below are procedural audits that are often performed by the Court.

- *Operational audits*

An operational audit is a check performed on the spot to establish whether certain previously identified processes are performed in the manner prescribed by the procedural descriptions (i.e. manuals and protocols) used by the organisation in question. This technique can be



used only if the organisation has made a written record of how a particular aspect of its policy should be implemented. The main steps involve selecting the processes that are to be audited, compiling an audit programme (i.e. a checklist) and making a note of whether procedures are indeed performed in accordance with written instructions.

Primary aim: orientation
--------------------------

For further information: A.J.G. Driesen en J.W. van der Klerk (1997). Operational Auditing. Een managementkundige benadering. Kluwer.
---

o *Operational management audits (performed on non-government organisations)*

Operational management is defined as the management and control of the operating processes used by a ministry, for example, in order to achieve its stated policy objectives. The processes concerned are both primary and support processes. Under the terms of the Government Accounts Act 2001, the operational management audits performed every year by the Netherlands Court of Audit for the purpose of the central government annual report consist of four components:

- an audit of the ministries' financial management, material management and the records kept for these purposes;
- an audit of the way in which the information on the policy pursued by the ministers and on the operational management conducted by the ministries has been prepared;
- an audit of the way in which the information in the ministerial annual reports on the policy and the operational management has been prepared;
- an audit of the quality of the information in the ministerial annual reports on the policy and the operational management.

Primary aim: orientation
--------------------------

For further information: Regularity audit manual (Intranet).
--

o *Policy theory analysis*

A policy theory consists of all the normative, causal and final elements (i.e. the relationship between the means and the end) underlying a particular policy. Clearly, the Netherlands Court of Audit is not in a position to assess the normative aspects. What we can do is to make pronouncements about, for example, the logic and consistency of a policy, the arguments underlying it and its realism. An assessment of a policy theory is in fact an evaluation that is performed without any knowledge of the actual outcome of the policy in question. By assessing the policy



theory, we can make a judgement about the plausibility and robustness of the policy.

You can assess the quality of a policy theory by following a number of lines of enquiry, of which the following are common examples:

- Is there a logical relationship between the means and the end?
- Are the assumed causal mechanisms plausible?
- Have cogent arguments been made for the policy?
- Are the objectives consistent with each other?
- Are the relevant resources present?
- Have all the relevant actors been involved in the formulation of the policy; what is expected of them; are these expectations realistic (i.e. a stakeholder analysis)?
- Are any other outcomes likely to be generated in addition to those which have been specifically mentioned (i.e. side effects; undesirable effects; will the policy still have any point by the time that it starts to have an impact)?

Some of these questions have already been asked in previous audits, such as the audit of the policy information on the Betuwe Line (more specifically, of the arguments underlying the policy). Others, such as those relating to the stakeholder analysis, constitute new elements in the Court's work. What is in any event true is that we have not previously assessed the policy theory in order to make pronouncements as to why the policy is working well or not.

Primary aim: assessing the plausibility and robustness of policy.
For further information: A. Hoogerwerf and M. Herweijer (1998). <i>Overheidsbeleid: een inleiding in de beleidswetenschap</i> . Samson.
<a href="http://www.rekenkamer.nl">Handreiking Verantwoorden over beleid</a> (www.rekenkamer.nl)
Werkbaarheidsanalyse beleidsvoornemens ( <a href="http://www.iustitie.nl/Images/11_39935.doc">http://www.iustitie.nl/Images/11_39935.doc</a> )

o *Descriptive* analyses (uni-variate)

Descriptive analyses are used to give an initial impression of the outputs (and/or outcomes), the factors that are capable of influencing them and the relationship between them.

*Frequency tables* and *histograms* list, for each variable, the number of observations for each category into which the variable may be divided. For example, how many of the staff at the Netherlands Court of Audit have read a flyer that was recently distributed on RSI? Clearly, this question can be answered only if the variable can be divided into a number of clearly identifiable categories, i.e. have read it, have not read



it, can't remember. Frequency tables and histograms cannot provide much information on continuous variables such as interval and ratio variables.

In the case of interval or ratio variables, data on the central tendency and the spread are more informative. The values for the *central tendency* indicate the score that should be taken as the 'average' or middle of the distribution. Depending on the type of variable used, the mode (i.e. the value that occurs most frequently in the sample), the median (i.e. the mid-point value) or the average (i.e. the sum of the scores divided by the number of scores) is the most suited for this purpose.

*Spread measures* can give you an idea of the differences between the scores or, in other words, the degree of variation between the scores in the distribution. If, for example, three quarters of the pupils in a given class have fully assimilated the contents of a particular lesson and score eight out of ten in a subsequent test, whilst the remaining quarter score no higher than four out of ten, the teacher will have to adopt a different tactic in the following lesson than would be the case if all pupils scored seven out of ten.

As far as nominal variables are concerned, a frequency distribution provides the most information on how people in the various categories have scored. In the case of ordinal and interval variables, you can indicate the *spread width* i.e. the range of the scores (in terms of the difference between the highest and the lowest value) as a rough indication of the spread. Measurement scales based on interval and ratio variables offer better measures of spread, i.e. the *variance* and the *standard deviation* derived from it. The standard deviation indicates the degree to which the scores are distributed around the average.

o *Heterogeneous outputs*

In many instances, a given actor produces not one, but a range of outputs<sup>25</sup> (the Netherlands Court of Audit is a case in point, as it often uses a combination of activities to publicise its reports, such as the publication of a press release, a press conference and press briefings). In such cases, the various outputs need to be set off against each other, a situation known as a '*multi-criteria problem*'.<sup>26</sup> There are two possible solutions to this problem: retaining the sub-scores, or combining the sub-scores to form an aggregate score (i.e. adjusting for heterogeneity).

---

<sup>25</sup> This is referred to as 'heterogeneity in production' or 'product differentiation'.

<sup>26</sup> Swanborn (1999), p. 176.



Sub-scores can be retained if it is either unnecessary, impossible or undesirable to combine them into an aggregate score. It is *unnecessary* to combine the various sub-scores if they all point in the same direction. It is *impossible* to combine them if you cannot calculate weighting factors for the sub-scores or if the measurement scale does not go any higher than ordinal variables (see section 4.3.1.2 for information on measurement scales). It becomes *undesirable* to combine the sub-scores if interest groups or persons use different weighting systems (Swanborn, 1999).

If none of the above applies (these are therefore the conditions that need to be met in order for it to be worth calculating the aggregate score), the sub-scores can be combined to form an aggregate score. There are various ways of doing this, and these are all explained in Swanborn (1999).

Interested in finding out more?

For further information on multiple scores, see P.G. Swanborn (1999). *Evalueren*. Meppel: Boom.

- o *Descriptive* analyses (bi-variate)  
You can use cross-tables, *Chi*-square tests and correlation in order to examine *relationships between a limited number of variables*. You should, however, be cautious when drawing conclusions, as you will often be unable to take account of all sorts of interfering factors that you should really check. To that purpose you need to use multivariate analyses.

You can illustrate the relationship between two nominal variables with the aid of a *cross-table*. In fact, a cross-table (or multi-way table) is nothing more than a frequency table (i.e. one-way table) containing simultaneous observations on at least two *nominal variables*. Although variables from a higher measurement scale can also be shown, you need to limit the number of categories as otherwise there will be too many. The aim of performing a *Chi*-square test is to determine whether an assumed correlation between two variables is statistically significant.<sup>27</sup> For example, when the Netherlands Court of Audit audited the scheme for bringing school drop-outs back into mainstream education, the auditors found that there were major differences in the three regions covered by the audit in terms of the services they provided for pupils playing truant

<sup>27</sup> 'Statistically significant' means that there is only a slight chance that the correlation you have found is the result of chance.



from school: job advice, a full-time job or counselling. However, only the differences in relation to counselling actually proved to be significant (Netherlands Court of Audit, 2001c: 7).

*Correlation coefficients* give information on the relationship between two variables measured at least on the ordinal scale. This technique is used so frequently by researchers that the term 'correlation' has become virtually synonymous with the terms 'relationship' and 'association'. Although it is a powerful research tool, it also has its drawbacks. Basically, correlation coefficients do not provide any information on the underlying causal relationships. The existence of a correlation between two phenomena is no guarantee whatsoever of a causal relationship between them.

#### 4.3.2.2 *Causality assessment techniques*

A difficult element in effectivity and efficiency audits is, how to prove the existence of any causal links. Depending on the audit subject (which often consists of a complex reality), the methodological quality of the audit and the analytical technique used, any conclusions drawn about causal links may be open to debate. You will always need to support any conclusions about causal links with methodological arguments. This section now discusses a number of the (multivariate) techniques used for this purpose by the Netherlands Court of Audit.

- *Elaboration*  
The assumption here is that an assumed causal link between two variables could be the result of a third variable. A good example would be the launch of an alcohol-free beer, as described in section 4.3.1.1.
- *Regression analysis*  
Regression analysis can be used to perform a multivariate test on assumed causal links and effects. By taking simultaneous account of the effect of a range of independent variables on a dependent variable, a multiple regression analysis provides information on the relative importance of the individual variables and also highlights any apparent links that are not in fact links at all. There are various types of regression analysis, including linear, panel (or time series), random effects (multilevel) and fixed effects (which is a combination of cross-section and time-series analysis). The advantage of regression analysis is that these techniques allow you to investigate more than one explanation at once.



- *T-tests and variance analysis*

This technique is used for obtaining information on the differences between groups. It is often used in an experimental or quasi-experimental setting that is designed to identify differences between a control group and an experimental group, with the aim of drawing conclusions about causal links.

A t-test is a useful means of finding out whether there is any difference between *two* group averages in the population. It is, however, not suited as a means of establishing whether there is any difference between the averages for *more than two* groups. This is the point where variance analysis comes into the picture.

Primary aim: testing
For further (accessible) information on statistics: see text part 1 of the <i>Methods and techniques</i> course (accessible on the DDO domain on the Intranet). You could also read: J. Stevens (1992). <i>Applied multivariate statistics for the social sciences</i> ; A. van Knippenberg & F. Siero (1980). <i>Multivariate analyse. Beknopte inleiding en toepassingen</i> .

- *Experiments (natural and otherwise)*

*Pure experimental research* is a good way of establishing the existence of a causal link. You should start by composing two comparable groups. One of the groups is exposed to a stimulus such as a public information campaign. You take measurements among both groups, both before and after the stimulus has been applied, to ascertain the situation in relation to the field affected by the stimulus, e.g. their attitude towards government, or their knowledge of the risks associated with unsafe sex.

This technique is generally not fully suited to the day-to-day practice of policy research, given the strict requirements attached to it, notably the *random* allocation of research units to the experimental or control group. In many cases, therefore, researchers decide for good reasons not to aim for the ideal situation and instead to make concessions either to the comparability of the groups or to the need for making measurements before and after the application of the stimulus. This type of test, based as it is on an experimental model, is known as a '*quasi-experiment*'. There exist a wide variety of quasi-experiments, the most common of which is that in which *measurements are taken among existing groups*, such as departments at a ministry, *before and after the application of a stimulus*.





There are all sorts of ethical and practical objections to *policy* experiments. For example, the government cannot deny one group of market gardeners certain tax breaks in order to ascertain whether they change their energy consumption, whilst at the same time affording the same tax breaks to another, comparable group. At the same time, experiments offer more opportunities than might appear at first sight. The fact is that opportunities regularly arise for 'natural' policy experiments. The policy on schools in deprived areas is a good example of this: school where pupils from ethnic minorities make up at least 70% of the student population receive extra funding, whereas schools with a lower proportion of pupils from ethnic minorities receive either less additional funding or no extra funding at all.

A 'natural' experiment makes use of the fact that certain groups may resemble each other in terms of this particular variable (e.g. a school where pupils from ethnic minorities make up 69% of the population and a school where they make up 70% of the population), and that one group will then receive a stimulus (in this case, extra funding) whilst the other group does not. This allows researchers to study the effects of the application of a particular policy tool (or, as the case may be, its non-application) on groups that are virtually identical. One testable explanation could be, for example, that an extra subsidy of 6,102.50 guilders per teacher per annum (this was the amount involved in this particular instance) actually leads to pupils from ethnic minorities achieving a higher score in their primary school leavers attainment test.

Primary aim: testing
For further (accessible) information on statistics: see text box 1 in the <i>Methods and techniques</i> course (accessible on the DDO domain on the Intranet). You could also read Baarda & De Goede (1995). <i>Basisboek Methoden en Technieken</i> ; Swanborn (1987). <i>Methoden van sociaal-wetenschappelijk onderzoek</i> . E. Leuven, M. Lindahl, H. Oosterbeek en D. Webbink (2003). <i>Zin en onzin in evaluatieonderzoek; over de effecten van extra subsidies voor achterstandsleerlingen</i> . Universiteit van Amsterdam (SCHOLAR). <a href="http://www1.fee.uva.nl/scholar/oosterbeek/achterstandsleerlingen.pdf">http://www1.fee.uva.nl/scholar/oosterbeek/achterstandsleerlingen.pdf</a>

- *Time-series analysis*

A time series is a chronological series of quantitative data, such as absenteeism rates, unemployment figures, crime figures, reoffence rates and figures on school drop-outs. You can use a time-series analysis to find explanations for a pattern of a time series in terms of other variables and for relating observations to certain structural rules of conduct. When the Netherlands Court of Audit audited covenants on combating violence



in the night-time economy, the auditors took measurements for a number of years in order to establish whether the fact that certain local authorities had signed such covenants had indeed led to a decline in the incidence of violent crime there. This involved studying both the possible environmental variables and the characteristics of the covenants.

The disadvantage of this method is that it requires a large number of research units and that the data collected must be of high quality.

Primary aim: testing
For further information: C. Chatfield (2004). <i>The Analysis of Time Series: An Introduction</i> . Sixth edition, Chapman & Hall. <a href="http://www1.fee.uva.nl/scholar/oosterbeek/achterstandsleerlingen.pdf">http://www1.fee.uva.nl/scholar/oosterbeek/achterstandsleerlingen.pdf</a>

#### 4.3.2.3 *Techniques involving a combination of exploration and assessment*

- o *Comparative case studies*

Performing a comparative case study is not the same as performing a number of case studies. The difference lies in the basis on which the cases in question are selected. For example, the cases in a comparative case study may be selected on the basis of one or more independent variables. The researchers can then opt for either minimum or maximum variation with respect to these variables. Cases may also be selected on the basis of one or more dependent variables. This is usually done by allowing for maximum variation (Swanborn, 1999: 60-66).<sup>28</sup> In other words, a good case study requires a prior knowledge of the cases in question.

A common distinction made in relation to case study design is that between a 'most similar systems design' (MSSD) and a 'most different systems design' (MDSD). It is used, for example, in political sciences to explain a phenomenon that occurs in different contexts, such as the emergence of economic tigers, coups d'état and revolutions. It is used to find explanatory factors that are common to all such situations, despite the fact that the countries involved differ widely from each other in all sorts of respects. MSSD, on the other hand, is used for comparing

---

<sup>28</sup> Another possibility is a combination of a single with a multiple case study, known as the 'embedded case study' approach. This involves differentiating a number of sub-cases within the same case, thereby enabling the researchers both to benefit from the advantages of a single case study and the greater depth resulting from savings in time, and also to study a number of cases at a lower level of abstraction. The various cases also provide a basis for making more generalised pronouncements about the audit subject.



countries that bear all sorts of similarities to each other, but differ in one respect.

Example: local authorities and specific-purpose grants

Let's assume that we are interested in finding out how local authorities use the freedom they have been given (at least on paper) to decide how to spend specific-purpose grants. Let's also assume that we want to find out why some local authorities make use of this freedom, whilst others do not. If we use a meta-evaluation to identify the success and failure factors for decentralisation processes, we can then use the results to make a very careful selection of cases and to aim our research efforts at highly specific targets. This turns the case study into an in-depth assessment.

Primary aim: both orientation and testing

For further information: Flyvbjerg, B. (2000). Making social science matter: why social inquiry fails and how it can succeed again. Cambridge: Cambridge University Press.

P. Verschuren (2003). Case study as a research strategy: some ambiguities and opportunities, in: Journal of Research Methodology, vol. 6. no.2, 121-139.

P.G. Swanborn (1996). Case-study's, Wat, wanneer en hoe? Meppel: Boom.

C. Robson (2002). Real World Research. Blackwell Publishing.

R. K. Yin (2002). *Case Study Research, Design and Methods*, 3rd ed. Newbury Park: Sage Publications.

- o *Benchmarking*

Benchmarking is a method that enables far more organisational units to be compared with each other than does a comparative case study. However, the results it generates are more superficial than those produced by a case study. The Dutch Public Administration Council defines benchmarking as follows: '*Systematically investigating the performance (and social impact) of and the underlying processes and operating methods used by one or more leading reference organisations in a particular field, and comparing your own organisation's performance and operating methods with these best practices, with the aim of arriving at a ranking and improving your organisation's performance.*' (Camp, 1992)

Example: differences between Centres for Work and Income

Central government wished to put an end to the fragmented manner in which claims for unemployment benefits, incapacity benefits and social assistance benefits were handled and assessed. On 1 January 2002, all the former job centres and municipal social services were amalgamated to form 130 Centres for Work and Income. Whilst all these Centres operate on more or less the same basis, some of them achieve a higher 'reintegration score' than others. Why is this?



By comparing all the various Centres for Work and Income and then identifying any discrepancies between their operating methods (and establishing the impact this has on their reintegration scores), we can explain the differences in their performances. Among the variables we can take into account are the size of the Centre, the use made of forms, waiting times, job placement rates, etc.

Primary aim: both orientation and assessment.

For further information: R. Camp (1992). *Benchmarking – Het zoeken naar de beste werkmethode die leiden tot superieure prestaties*. Deventer.

Raad voor het openbaar bestuur (1992). *Presteren door leren*.

Ministerie van BZK & VNG (2004). *De Gids. Prestatievergelijkingen en benchmarks binnen de Openbare Sector*. (<http://www.vng.nl/Documenten/Extranet/Bjz/Bb/degidsdefinitief2.pdf>)

Once you have defined and collected the necessary data, the next step is to calculate the differences in the efficiency of the policy outputs or outcomes on which you wish to form a judgement. There are various ways of doing this. The main ones are the use of simple indicators and the following analytical techniques: free disposable hull (FDH), data envelopment analysis (DEA) and stochastic frontier analysis (SFA). Once you have used indicators to measure the efficiency of certain outputs, you can then test the potential explanations for the differences in efficiency. Among the methods available for this purpose are tobit and probit analysis, as well as simple regression analysis.

- o **Group Model Building**

As the name suggests, Group Model Building (GMB) is a means of building models in groups that reflect social reality. Group Model Building helps teams to assess their situation from a systemic perspective and hence to amalgamate individual sub-opinions into an aggregate opinion. The model-builder acts as a group facilitator, i.e. someone who helps the group to develop this type of system dynamics model. As a result, better use is made of the pool of knowledge available in the group, but which is often underutilised due to poor communication. The ultimate aim is to find a solution that is not only of high quality, but which also has the support of the group in question.

There are two types of Group Model Building: a more qualitative type and a more quantitative type. The qualitative type (which is also known as 'mapping') is used to analyse a problem with the aid of 'causal loop diagramming'. A causal loop diagram can help to structure cause-and-effect arguments in relation to complex problems. The diagram is then used to analyse the problem and identify intervention points in the system that can be used to design strategies for solving the problem.



When the quantitative type of GMB is used, the causal loop diagram is converted into a 'stocks and flows' diagram that is then quantified, after which simulations can be performed with the model and the effectiveness and robustness of potential solutions tested.

The Netherlands Court of Audit might use GMB, for example, if it was looking for explanations for persistent problems in a complex policy field such as health care. GMB would involve bringing together all parties involved to set out their views on the effectiveness of policy. What factors and actors play a role? For which factors and actors does the policy cater? And does it do so effectively?

It is important to bear in mind that GMB requires a high degree of commitment on the part of the actors involved, and also a considerable amount of specialist knowledge, which the Netherlands Court of Audit does not possess. This means hiring external consultants.

Primary aim: both orientation and assessment.
For further information: J. Vennix (1996). <i>Group Model Building</i> . Wiley. J. Vennix (1999). Group Model-building: tackling messy problems. <i>System Dynamics Review</i> , 15 (4), 379-401. <a href="http://www.par-groep.nl/index.php?m=1&amp;s=5">http://www.par-groep.nl/index.php?m=1&amp;s=5</a>

o *Network and stakeholder analysis*

Generally speaking, ministers implement their policies on particular fields with the aid of local authorities, private firms, civil-society organisations and/or groups of citizens. A network analysis involves looking at the role played by the government as a 'co-player', i.e. an actor seeking to exercise control amid other actors seeking to do the same.

You can use network and stakeholder analyses to analyse hierarchic relationships, patterns of interaction, interdependencies and problem definitions. The key questions with this type of analysis are:

1. Who is involved in the implementation of policy?
2. What is the relationship between the various actors?
3. In what way can individual actors either foster or impede the implementation of the policy in question?
4. Do all the actors share the same view on the policy problem and its potential solutions?

The answers to the above questions can be used to find out why there have been delays in the implementation of the policy in question, for



example because the hierarchic relationships have not been clearly defined, or because the actors who need to implement the policy have other priorities.

Example: division of responsibilities between the Ministry of Education, Culture and Science and the Primary Education Process Management Team.

Organisations for process management are defined as semi-independent organisations of a temporary nature that operate at a distance from the ministry in question and that are responsible for supervising major, complex renewal processes. The Primary Education Process Management Team (PEPMT) is one such organisation.

An audit (Court of Audit, 1999c) revealed a lack of clarity about the role of the PEPMT. Was it an independent organisation, part of the ministry or a consultant to the state secretary? Officials at the ministry acknowledged that, with hindsight, insufficient thought had been given to the question of what the PEPMT was and should do. A study of possible explanations could look explicitly at the relationship between the PEPMT and the Ministry of Education.

Primary aim: both orientation and assessment.

For further information:

Handleiding Normen voor netwerksturing (Court of Audit Intranet)

Marcoen Roelofs. *Criteria for the evaluation of public action taking place within networks* (paper made for the 6th conference of the European Evaluation Society, sept/octobre 2004).

Gepubliceerd op <http://www.mande.co.uk/networks>, november 2004.

J.A. de Bruijn, E.F. ten Heuvelhof (1995). *Netwerkmanagement: strategieën, instrumenten en normen*. Utrecht, Lemma.

G. Teisman (1992). *Complexe besluitvorming, een pluricentrisch perspectief op besluitvorming over ruimtelijke investeringen*. VUGA.

[http://www.stormmcs.nl/denieuwedorpen/downloads/DND\\_stakeholders\\_analyse.pdf](http://www.stormmcs.nl/denieuwedorpen/downloads/DND_stakeholders_analyse.pdf)

<http://www.ketens-netwerken.nl/>

#### 4.4 Methods matrix

In section 3.3, we used the explanations matrix to discuss the possible explanations for poor outputs and outcomes or inefficiencies. The following 'methods matrix' is a means of indicating, for a number of the methods described in the previous section, which of them are (more or less) suited for a study of explanations within one of the four quadrants. The methods in the box in the middle are suited for a wide range of applications.



	Outputs	Outcomes
Internal	<ul style="list-style-type: none"> <li>Operational audits</li> <li>Operational management audits</li> </ul>	<ul style="list-style-type: none"> <li>Policy theory analysis</li> <li>Experiments (natural and otherwise)</li> </ul>
External	<ul style="list-style-type: none"> <li>Operational management audits performed on non-governmental organisations</li> <li>Network- and stakeholder analyses</li> </ul>	<ul style="list-style-type: none"> <li>Time-series analyses</li> <li>Event analyses</li> </ul>

Literature searches, case studies (single and comparative), benchmarking studies, focus groups, Delphi method, regression analyses, Group Model Building



## 5 Analysing policy aims and policy information

### 5.1 Introduction

This chapter discusses the analysis of policy aims and policy information. Section 5.2 explains how to analyse policy aims, including the standards for and reconstruction of policy aims. Section 5.3 looks at techniques for analysing policy information, and discusses the various types of policy information, sources of policy information, standards for policy information and how to assess policy information, before going on to make a series of recommendations.

### 5.2 Policy aims

#### 5.2.1 General

All performance audits are based on the policy aims formulated either by a minister or by a legal person with statutory tasks. These days, a minister's general policy aims are set out in the VBTB budgets.

Following the publication of a government policy document entitled *From policy budgets to policy accountability* (VBTB), the various ministerial budgets have been transformed from budgets containing a mass of financial information into budgets containing specific policy objectives. Ministers are required to make clear in their VBTB budgets:

- (1) what they wish to achieve;
- (2) how they intend to set about this;
- (3) how much this is going to cost.

As a result, these budgets form an ideal starting point for the Court's performance audits (see Annexe 3 for more detailed information on VBTB budgets).

In practice, the objectives formulated in the policy articles in budgets are generally too abstract to be used for the purpose of analysing policy aims. The fact is that policy documents and parliamentary papers contain more information on specific, operational objectives.





The aims of legal persons with statutory tasks may be found either in legislation (they are, after all, organisations with statutory duties) or else in ministerial budgets, deeds of incorporation or annual reports published by the entities themselves.

### **5.2.2 Reconstruction and operationalisation of policy aims**

If, despite being required to publish a budget that meets the demands of the VBTB operation, a minister fails to formulate clear policy aims, the Court will try to reconstruct or operationalise these aims. In doing so, it is important to ask the minister to give his or her approval at each stage of the reconstruction or operationalisation process.

#### *Reconstructing policy aims*

In order to identify policy aims, it makes sense not just to examine official policy documents, but also to consult the auditee. Interviews tend to show that the official policy aims do not always paint an accurate or reliable picture of the auditee's actual objectives. The official policy aims may prove to be vague, overambitious, underambitious or out of date. Despite this, it is always worth basing the audit itself on the official policy aims.

Even if the auditee is seeking to achieve an objective that is not consistent with its official policy aim, the latter forms the main source of legitimacy for the policy in the democratic process. From the viewpoint of democratic control, therefore, the official policy aims to provide a point of reference. This is not to say, incidentally, that findings and conclusions cannot be formulated about any discrepancies found between the official and actual policy aims.

One way of *reconstructing* policy aims is by making use of a 'policy tree'. A policy tree is used to link tools with secondary and final policy aims. Building a policy tree is no sinecure. As the building blocks used to create a policy tree are often taken from a range of policy documents, it is worth consulting the auditee about the tree.

#### *Operationalising policy aims*

It is fairly easy to operationalise policy aims yourself if these have been formulated in immediately observable terms, such as 'lowering the average speed on motorways, i.e. to x kilometres an hour'. In many



cases, however, policy aims are formulated in terms that bear no relation to any immediately observable features. One such policy aim would be 'the prevention of waste'.

What you need to do, therefore, is to try and design indicators that paint a picture of the aims formulated by policymakers that is as accurate and complete as possible, and yet are fully quantifiable. In doing so, you should ask yourself the following questions:

*Do the indicators cover the full scope of the policy aim?*

For example, do the indicators relate solely to the prevention of absolute quantities of waste (i.e. quantitative prevention) or do they also measure attempts to reduce the harmfulness of waste substances to the environment (i.e. qualitative prevention)?

*Is there a sufficiently robust relationship between the policy aim you are trying to operationalise and the indicators you have selected for the purpose of your audit?*

For example, an indicator defined as 'a reduction in the quantity of waste produced by companies' provides only a limited indication of the extent to which waste emissions have been 'prevented'. This is because the volume of waste production does not depend solely on preventive measures. It also depends on autonomous factors such as the economic growth rate.

*Are there any factors that distort the picture painted by the indicator of the policy aim?*

This is what happens, for example, when a policy aim defined as 'preventing hazardous waste' is operationalised as an indicator defined as 'changing the amount of hazardous waste reported by companies'. This indicator ignores any changes in companies' reporting behaviour. In other words, if companies become less scrupulous in reporting hazardous waste emissions, the result will be a decline in the volume of hazardous waste that is actually reported without this leading to the prevention of hazardous waste.

### 5.2.3 Standards for policy aims

Policy aims must be SMART,<sup>29</sup> i.e. specific, measurable, agreed-upon, realistic and time-related. More specifically, this means that:<sup>30</sup>

---

<sup>29</sup> Bouckaert et al. (1998). The NCA decided to use the term 'agreed-upon' instead of the more currently used 'achievable'.

<sup>30</sup> Based partly on Court of Audit (2002a).



SMART	Description
Specific	Aims must be specific: which outputs and outcomes is the policy intended to produce? Which target groups (assuming these exist) does it seek to reach?
Measurable	It must be possible to measure whether or not the policy has succeeded in achieving its aims. This requires an unambiguous definition of terms, as well as (in many cases) some sort of baseline survey.
Agreed-upon	At the very least, the relevant actors (such as the House of Representatives or the relevant executive agency) must have been consulted about the policy aims.
Realistic	The policy aims must be formulated in such a way that they are achievable in all plausible circumstances (i.e. exogenous variables and policy measures).
Time-related	A date must have been fixed by which the final aim must have been achieved; intermediate aims and intermediate deadlines may also be formulated for long-term projects.

Moreover, the policy aims also need to be consistent:

Consistency of policy aims	
Consistent	Mutual consistency between aims: one policy aim may not exclude another.
	Consistency between aims and basic data: if the basic data change, the aims should be amended accordingly.

For this reason, the Court refers to 'SMART-C' policy aims. More detailed information on the above standards is given in Annexe 4. See also the standards database (which you can consult by choosing the 'Court of Audit Applications' option in the 'Start' menu), as well as the explanatory notes in Annexe 4, for examples of the use of the standards.<sup>31</sup>

#### 5.2.4 Assessing the formulation of policy aims

Check lists may be used for the purpose of assessing the formulation of policy aims. See section 5.3.7 for further information.

#### 5.2.5 Recommendations

An examination of policy aims generally results in a series of recommendations, such as that:

- the policy aims should be reformulated to make them easier to assess;

<sup>31</sup> Under the *Central Government Budget and Accounting Regulations 2001*, general policy objectives (or at least one operational objective) must be expressed as far as possible in terms of measurable targets that are capable of being influenced. These should also contain a description of the target group or groups, and should state a time horizon.



- any mutual inconsistencies should be removed or an indication given of the degree to which they are allowed to conflict with each other (e.g. policy aim A may not impede policy aim B to such an extent that policy aim B falls below a given minimum limit or exceeds a given maximum limit);
- better arguments should be provided in support of the policy aims (i.e. by assessing their feasibility);
- changes in policy aims or evaluation criteria should be defined more clearly and better evidence should be provided in support of them.

### 5.3 Policy information

#### 5.3.1 General

Policy information plays a key role in the management of public services. The importance of such information has only increased since the publication in 1999 of a government policy document entitled *From policy budgets to policy accountability* (VBTB) and the publication of the first VBTB budgets in 2002. The *availability* of policy information (and its assessment by the Netherlands Court of Audit) is not an aim in itself, but is a condition that needs to be met in order to monitor and foster the efficiency and effectiveness of government policy.

#### 5.3.2 What policy information?

##### *Purposes of policy information*

Policy information can be used for a variety of purposes and at various stages of the policymaking process:

- during the preparatory stage: as a basis for the formulation of policy and/or the preparation of the relevant *budget*;
- during the implementation stage: for *managing and adjusting* policy (including supervising the implementation of policy);
- during the evaluation stage: for the purpose of *reporting* on the policy that has been pursued.<sup>32</sup>

The different purposes are associated with different target groups and different types of use in the various cycles. When used for *management*

---

<sup>32</sup> Management and reporting do not relate solely to policy, but are also geared towards operational management and funding. See the report entitled *Jaarverslag in the politieke arena, naar een nieuwe stijl van verantwoorden* ('Annual reports in the political arena: towards a new reporting style', May 1998).



*purposes*, for example, policy information is particularly relevant to the policy department and the minister in question; certain policy information will also be needed by the House of Representatives for the same purpose. The need for using policy information for *budgeting and reporting purposes* is based on the House of Representatives' right to amend government budgets. The ministers supply this information to the House of Representatives, either to gain approval for the policy they are planning to pursue (in the case of budgets) or to report on the progress made in implementing existing policy (in the case of annual reports). This information is provided as part of the annual budgetary cycle.

The role played by information during the preparatory stage also has a bearing on the Court's effectiveness and efficiency audits. The information in question consists of a problem analysis ideally incorporating a baseline survey and a policy theory, plus *ex-ante* evaluations for a range of scenarios and alternative policies. Information on the arguments behind the policy aims is also relevant.

#### *Policy information on outputs and outcomes*

The type of information that is required on outputs and outcomes depends largely on the aims that the government has formulated for its policy, for example in budgets and policy papers (this applies, for example, to information on the *type* of impacts that should be examined and the nature of the *criteria* that need to be formulated in order to treat certain impacts as adequate).

The ***Dutch Central Government Performance Data and Evaluative Studies Regulations*** describe the nature of the policy information that needs to be supplied on outputs and outcomes. Three different *types* of performance data are distinguished:

1. policy objectives, which should be operationalised as far as possible in terms of desired and actual *outcomes*, linked to certain target groups and to a specified time horizon;
2. information on the cost price and quality of *outputs* delivered or to be delivered by the government (i.e. products and services);
3. programme expenditure, supported as much as possible by *volume and price data* (such as the expected number of recipients of a given grant and the value of the grants paid).

In other words, the term 'performance data' has been broadly defined as encompassing not only *outputs*, but also *outcomes* and *costs*. Under current rules, budgets must also include information, for example on the



cost price and quality of the outputs delivered or to be delivered by the government.

See also chapter 3 for detailed definitions of the terms used here.

### 5.3.3 Sources of information on outputs and outcomes

Policy information on outputs and outcomes may be obtained from a wide variety of sources. A great deal of policy information may be found in budgets and annual reports, which are supposed to contain – alongside data on the *projected and actual* impacts of government policy – information on past or future activities performed with the purpose of achieving the impacts in question (including outputs delivered or to be delivered) and the associated cost. This information is required in order to make pronouncements about whether the government has achieved its aims, the government's policy and operational management have been efficient and the policy has been effective.

The *Dutch Central Government Performance Data and Evaluative Studies Regulations* referred to above cited systems for generating standard performance data and regular evaluations (both *ex ante* and *ex post facto*) as being sources from which policy information may be obtained. Alongside these sources, however, there are also non-standard sources of information, such as non-government organisations.

#### *Standard performance data*

According to p.19 of the *Dutch Central Government Performance Data and Evaluative Studies Regulations*, systems for generating standard performance data are 'information systems that are used to generate information at regular intervals on:

1. the degree to which the government has achieved its objectives or delivered the intended outcomes;
2. the progress made by the government in implementing its policy; and
3. the number, cost and quality of the outcomes delivered.'

Although the word 'system' appears to imply something complex, electronic and expensive, there is no reason why this should be the case. An organisation that each year publishes a set of indicators that paint a clear picture of the number and quality of the outputs delivered, may also be said to be using a performance information system.



### *Evaluations*

It goes without saying that evaluations (both *ex ante*<sup>33</sup> and *ex post facto*<sup>34</sup>) performed by or on behalf of the ministries also generate policy information. Because of their in-depth nature, evaluations are ideally suited as tools in conducting audits of the *efficiency* and *effectiveness* of government *policy*. As a result, evaluations performed by or on behalf of ministries are a vital source of policy information on efficiency and effectiveness.

The ***Dutch Central Government Performance Data and Evaluative Studies Regulations*** state that the general and/or operationalised policy objectives included in budgets and annual reports generally need to be evaluated at least *once every five years* with the aid of an *ex post facto evaluation*.

### *Other sources*

In addition to standard sources, policy information may also be found in inspection reports, (non-ministerial) non-standard progress reports, policy documents, parliamentary papers, and so forth. Databases may also be regarded as constituting a source of raw policy information, which means that they can be used for the purpose of conducting performance audits.

A relevant question in this connection is whether information on policy must have been collected by or on behalf of a ministry in order for it to qualify as policy information. After all, non-government organisations such as universities and trade associations also collect information on government policies and their effects. It is also conceivable that certain raw data may be available at a ministry that is not suitable for immediate use.

The Court regards such types of information as policy information, even if the ministry in question is unaware of its existence or has no direct access to it. Ministers are regarded as being at fault if they are unaware

---

<sup>33</sup> An *ex-ante* evaluation is a systematic means of comparing the cost and anticipated effects and side effects of two or more policy alternatives.

<sup>34</sup> The term '*ex post facto* evaluation' is defined as follows: 'a systematic study of the effects of existing government policy, the way in which policy is implemented and/or the cost and quality of the products and services supplied' (*Ministry of Finance* (2001), p.21). There are various forms of *ex post facto* evaluation (*Ministry of Finance* (2001), p.21): studies of the extent to which the government has achieved its policy aims, studies of the effectiveness of government policy, studies of the efficiency of government policy and studies of the efficiency of the government's operational management.



of the existence of relevant information or if they decide to make no use of it without having any good reasons for not doing so.

#### **5.3.4 Quality of policy information (as a product)**

Assessing the quality of policy information is a possible aspect of performance audits. A large amount of documentation is available on the standards with which such information is expected to comply.

In order to assess the quality (and usefulness) of policy information, the team of auditors makes use of detailed standards formulated by the Court for policy information, and also of the applications described in the standards database (which you can consult by choosing the 'Court of Audit Applications' option in the 'Start' menu). These standards apply both to policy information generated by systems for producing standard performance data, and to policy information thrown up by evaluations. It goes without saying that the standards also apply to information from other sources as described in section 5.3.3.

All audits performed by the Netherlands Court of Audit are based on the *same* basic standards for the assessment of information. There are two categories of basic standards<sup>35</sup> (see also Annexe 5a):

- *Basic technical standards*: relevance, reliability and conformity with statutory regulations and authoritative guidelines for the contents of information.
- *Basic standards on form and presentation*: comprehensibility, comparability and conformity with statutory regulations and authoritative guidelines for the form of information.

If the policy information in question complies with the six basic standards, the Court labels it as being 'useful'.

When conducting effectiveness and efficiency audits, you will in any event also need to consider the following secondary standards for the relevance of data. The information should be:

- complete;
- up to date in its calculations and assumptions ;
- available within a reasonable lapse of time.

---

<sup>35</sup> Court of Audit (2004): *Guide to standards and terms* (part of the Netherlands Court of Audit Manual (HANDAR)).





You will in any event also need to consider the following secondary standard when measuring the auditee's performance in relation to the basic standard for *comparability*:

- consistency (in time; between different parts of the policy chain; and between different sources).

These basic and secondary standards together form what is known as a 'basic package' of standards. The way in which secondary standards are applied will vary from one case to another.

Detailed information on 'completeness' as a secondary standard is given in Annexe 5b. This annexe describes in general terms the type of information required to manage policy at the various stages of the policy process. The term 'optional' is used to describe aspects that are *not* regarded as forming part of the basic package of requirements for completeness as a secondary standard. Examples of applications of the standards are given in Annexe 5c.

### **5.3.5 Quality of policy information (as a process)**

#### *Preparation of information*

In order to establish whether policy information – whether supplied by systems for generating standard performance data or evaluations – complies with the six basic standards outlined above, you will generally need to look carefully at the three stages for the preparation of information. Information is prepared in the following three stages:<sup>36</sup>

1. Collection, including initial recording. The product generated at this stage is 'basic information';
2. Processing. The product generated at this stage is 'refined information';
3. Distribution. The product generated at this stage is 'distributed information' or 'reporting'.

These three stages are inextricably bound up with each other. After all, if the government or any other organisation wishes to distribute information that complies with the requirements (stage 3), both the information collection (stage 1) and the information processing stages (stage 2) must come up to standard.

For example, in order to be able to distribute *relevant* information, ministries and other organisations should adopt a rational, analytical

---

<sup>36</sup> From: *Guide to standards and terms* (2004).



approach to the *collection* of data (particularly on policy results). This means that they must have formulated and operationalised objectives, and defined performance data (including data on outcomes).

In addition, when collecting data, ministries should take account of the users of the information in question. This means that they:

- (1) need to have a clear picture of the users;
- (2) should be aware of the users' information needs (in terms of quantities and units); and
- (3) should design the data collection processes to suit the type of information required by the users.

When a ministry or other organisation *processes* (i.e. aggregates, synthesises, etc.) basic information, it is important to retain its integrity. This means that processing methods also need to comply with the requirements in terms of their reliability, validity, orderliness and auditability. If you assess the reliability and validity of the information provided, you must be able to rely on the guarantees provided by the ministries and organisations in question. These are, for example, that the right type of expertise has been deployed, that job segregation has been used, that clear procedures have been used and that all data have been properly recorded (so that there is a proper audit trail). If the ministries and other organisations do not process the data themselves, they must satisfy themselves that the other parties processing the information do so properly.

#### *Requirements relating to regulations*

Clause 5.1 of the *Dutch Central Government Performance Data and Evaluative Studies Regulations* state that the information systems (for standard performance data) and evaluation tools used for budgeting and reporting purposes must comply with the requirements in relation to validity, reliability, accuracy and usability.

These requirements are fleshed out in more detail on pp. 34-37 of the *Dutch Central Government Performance Data and Evaluative Studies Regulations*. Clause 5.2 states that the policy information included in budgets and annual reports should be prepared in an orderly and auditable fashion. It should be clear how the relevant information has been compiled: 'Clear information should be given on the choices underlying the preparation of the information. Systems for standard performance data and the procedures for processing such data must be recorded in such a way as to enable their quality to be assessed. It must



be possible to ascertain the quality and impartiality of evaluations with the aid of an audit file.<sup>37</sup>

### *Assessing systems for measuring performance*

The first step to take in evaluating a performance information system is to determine what organisational purpose the system serves. In other words, is it acting as a neutral thermometer, simply recording things from an objective distance without having any impact on the organisation in question? Or does it play a critical role in the organisation's planning and control cycle? Ideally, of course, data collection and analysis should form part and parcel of the planning and control cycle, as this guarantees that the information is used for revising policies and for reporting purposes. After all, there is no point in collecting data if you're not going to do anything with it.

The second question you need to ask is whether, in the light of the policy's strategic and operational aims and of the role performed by the performance information system in this respect, the performance information system *contains the right performance indicators*. This is because the strategic and operational aims are not always translated into the appropriate indicators, which makes the information generated by the performance information system unsuitable for use either for revising policies or for reporting purposes. For example, if the objective of the government's policy on waste emissions is 'to prevent the production of waste', a class of performance data defined as 'decrease in the quantity of waste' does not actually say much about the prevention of waste production, which means it is not a good indicator of the outcome of the policy in question. However, if the objective is defined as 'reducing the quantity of waste', the performance data says a great deal about the government's success in attaining its aim, and is hence a good indicator of the outcome of the policy.

The third step involves establishing *whether the performance indicators are appropriate from a measuring viewpoint*. In other words, this is a question of ascertaining their validity and reliability.

The Netherlands Court of Audit sets the following requirements with regard to systems that measure the achievement of objectives and the effectiveness and efficiency of certain policies or organisations:

---

<sup>37</sup> Ministry of Finance (2001).



- the performance information system should contain information on inputs, throughputs, outputs and outcomes;
- the information generated by the performance information system should satisfy the product criteria listed earlier on in this section;
- the information generated by the performance information system should allow an external auditor to reach a judgement on effectiveness and/or efficiency;
- the indicators should be linked to specific strategic or operational objectives set for the policy or organisation in question;
- the indicators should not encourage short-termism at the expense of a long-term perspective;
- the benefits gained from data collection should be in reasonable proportion to its cost;
- the information generated by the performance information system should help management to reach decisions;
- the information generated by the performance information system should be used, for example, in the budgetary process, in regularly recurring decisions or in evaluations;
- old information should be stored, so that it is possible to analyse chronological changes.

#### **UK criteria for performance information systems**

A manual produced by the HM Treasury and the UK's National Audit Office (HM Treasury, 2001), entitled *Choosing the right FABRIC. A framework for performance information*, includes a chapter setting out the properties of a good system of performance information.

These properties are as follows:

*Focused* on the organisation's aims and objectives;

*Appropriate* to, and useful for, the stakeholders who are likely to use it;

*Balanced*, giving a picture of what the organisation is doing, covering all significant areas of work;

*Robust* in order to withstand organisational changes or individuals leaving;

*Integrated* into the organisation, being part of the business planning and management processes; and

*Cost effective*, balancing the benefits of the information against the costs.

There is a copy of the manual in the Netherlands Court of Audit's library.

### **5.3.6 Assessing the use made of policy information**

Once you have assessed the quality (i.e. the fitness for purpose) of policy information, the next logical step is to examine whether the information is actually *used* for the purpose of developing and revising policy. An



examination of the use made of policy information is absolutely crucial, particularly if you are investigating the reasons for pursuing certain policies (i.e. if you are performing an *ex-ante* evaluation such as that of the policy information on the Betuwe Line, a new rail link from Rotterdam to Germany).

Even if the emphasis in a particular audit is not on the arguments underlying a given policy, it is worth thinking about including this aspect. It is important to ensure that ministries do not simply collect information without finding out whether it is actually used in practice. Moreover, the fact that certain policy information is not used (or that use is made of poor-quality information) may in itself explain why the government has not been able to achieve certain outputs or outcomes.

Clause 6.1 of the *Dutch Central Government Performance Data and Evaluative Studies Regulations* cites a number of aspects that have a bearing on an examination of the use that has been made of policy information and performance data:

- senior officials and ministers must be informed, in accordance with a standard procedure devised by the ministries themselves, of the results of regular evaluations;
- performance data should be used for the purpose of (a) the national budget, (b) the ministries' budgetary cycles, (c) the ministries' policy processes, and (d) operational management.

The following requirements are also relevant in this connection (see also Court of Audit, 1991c):

- the right people (e.g. members of parliament, the minister and officials) should receive the results of the audit *in good time*;
- the competent authorities should decide (*in good time*) on the type of action to be taken on the basis of the audit findings (e.g. a change in or support for the policy or the organisational structure in which the policy is made);
- those concerned should be notified of any changes (in terms of aims, organisation or tools).

For reasons of auditability, it is important that documentary evidence is available of the use made of policy information.

### **5.3.7 Checklists for assessing policy aims and policy information**

You can use *checklists* for assessing the formulation of policy aims and/or policy information (on outputs and outcomes, among other aspects).



Checklists may be used, for example, for systematically ascertaining whether each report in a series of audit reports meets a number of preset requirements. It's not simply a matter of *assessing* the requirements for the formulation of policy aims or policy information. A checklist may also be used for *listing* certain general and/or methodological aspects of the report in question.

Examples of checklists used by the Netherlands Court of Audit are available on the Court's Intranet.

### **5.3.8 Provision of information to the House of Representatives**

The Court has drawn up guidelines for the provision of information by ministers. These guidelines, which are available for inspection in HANDAR (in the folder entitled 'primary processes / audit process'), should be observed when examining the way in which ministers discharge their obligation to inform the States General. The guidelines not only contain standards for the provision of information, but also discuss the terminology used by the Court in its reports on this delicate matter. The guidelines also contain a checklist.

You should be alert to the fact that additional information requirements apply to projects that are subject to the procedural regulations for major projects.

### **5.3.9 Other points of interest**

You should pay special attention to the following points when assessing policy information. First of all, you need to know the level at which information needs to be available. We have already made clear that, in principle, the minister should have access to this type of policy information. In practice, however, this depends on the responsibility borne by the minister for the policy field in question (see also chapter 3). Where a regulatory authority has been made responsible for overseeing the field in question, for example, the type of information available to the minister may well differ from the information available where the minister bears full responsibility for the policy. This will need to be defined for each individual policy field.

#### **Assessing policy information at local level**

When the Netherlands Court of Audit seeks to assess the systems put in place to guarantee the quality of local information, it is moving in an area in which it has no real authority. This means it has to rely on voluntary cooperation. This was the case, for example, with the audit on the subject



of counselling and reintegrating school drop-outs, where the Court sought to assess the quality of local policy information. The same was the case when the Court audited government policy on the big cities, which involved looking at the quality of the government's tools for monitoring and auditing its policy. In both cases, however, the Court explicitly decided to target the minister, whose systemic responsibility for policy included a responsibility for ensuring the effective provision of information.

In its audits on the subjects of preventing and combating juvenile crime, and preparing disaster plans, the Court also looked at the amount of information provided to certain local actors. Its aim in doing so was to assess both the coordinating role played by local authorities and the supervisory role played by provincial councils. Although the minister is not directly accountable for the actions of local actors, he or she is in a position to take them to task for any apparent shortcomings.

Second, you need to know how to assess policy information produced by professional suppliers of information, such as the National Institute for Public Health and the Environment, Statistics Netherlands, the Netherlands Bureau for Economic Policy Analysis and the Social and Cultural Planning Office of the Netherlands.<sup>38</sup> In theory, these organisations should be treated in the same way as other suppliers of information, and there is no reason why the Court's auditors should not be able to get beyond their front doors. In practice, however, we will need to see how far we can and wish to go in assessing this type of information and what level of detail we wish to pursue. The same applies to organisations such as regulatory authorities, inspectorates and other supervisory bodies.

### 5.3.10 Recommendations

In making recommendations, it is important to suggest *specific* means by which structural improvements could be made to the quality of policy information.

For example, in its report on preventing and combating juvenile crime,<sup>39</sup> the Court discusses all sorts of technical details that need to be included in the information flows (such as a feature showing whether juveniles

---

<sup>38</sup> See Ed Broeze's memorandum of 12 February 1999 to the consultative committee of directors and secretary-general for information on the pros and cons of this type of audit. Relevant questions include: are we capable of assessing the value of the information (notably the models) published by this type of specialist organisation? Do we know enough about the practical processes such models are intended to represent? Do we have enough expertise to be able to assess the methodological aspects? Do we need to call in external experts to help us? If so, are there any independent experts available, given that knowledge tends to be concentrated in monopolistic organisations?

<sup>39</sup> Court of Audit (2002e).



have actually received any help and, if so, with what effect; a clear picture of turnaround times).

Apart from their contents, the format of systems for measuring the quality of information is also a critical factor. For example, the report on the subject of counselling and reintegrating school drop-outs<sup>40</sup> discusses in detail the need for linking school records with the records kept by bodies responsible for counselling and supervision (such as juvenile welfare institutions and school attendance officers), as well as the information sent to the Ministry of Education to support its policy on school drop-outs. After all, it is far more efficient for all parties concerned to compile the data required for reporting on national policy by computer-processing the data the various parties need in order to perform their work, than to undertake a series of separate operations. In order for this to be possible, however, guarantees need to be given on the quality of the 'local' data, and the scope of such guarantees is limited. It goes without saying that the privacy laws should also be strictly observed.

Recommendations along the lines of 'the quality of policy information should be improved' are too vague. The following are examples of specific recommendations:

- Investigate why 44% of the juveniles classified as suspects in the Juvenile Crime Tracking System are not recorded as having been prosecuted.
- Review the policy on the big cities more systematically, so as to gain a picture of the government's success or otherwise in achieving its objectives, the effectiveness of the policy, and the possible explanations for these.
- Design an information system containing relevant information on inputs, throughputs, outputs and outcomes. The performance information system should mesh in with the planning and control cycle, and with the policy cycle.

---

<sup>40</sup> Court of Audit (2001b).





## 6 Audits of the achievement of objectives (and the reaching of target groups)

### 6.1 Introduction

We explained in Chapter 3 that there are three basic types of performance audits: those geared towards assessing the achievement of objectives or the reaching of a target group, effectiveness audits and efficiency audits. The difference between them lies in the fact that they each come with their own problem definitions, audit subjects and audit methods. This chapter looks in more detail at the first type: audits geared towards assessing the achievement of objectives or the reaching of target groups.

### 6.2 Audits of the achievement of objectives

#### 6.2.1 Problem definition and audit questions

Audits of the achievement of policy objectives are designed to assess the degree to which a general and/or a more specifically operationalised policy objective has been achieved. In accordance with the VBTB operation<sup>41</sup>, these objectives may be formulated in terms of target groups (see section 6.3), target values and deadlines. It is generally fairly easy to produce a problem definition for such audits. A good example would be:

- Have the aims of the policy been achieved?

The precise aims of the policy must already have been identified (see Chapter 5). The policy aims themselves may be formulated in terms either of outputs or of outcomes (i.e. changes in society). The problem definition itself can then be fleshed out in greater detail, so as to include, for example, specific audit questions on each policy field, ministry or legal person with statutory tasks, administrative level, policy tool and/or

---

<sup>41</sup> An operation aiming at formulated the national budget in such a way that it is clear what the state wants to achieve, what it is planning to do to that end, and how much this will cost.



period. You may also wish to examine the role played by the minister in relation to the achievement of the policy aims. The following are examples of audit questions for this type of audit:

- How many subsidies were awarded in the past year, and does this number correspond with the government's stated aims? (This is a notional example.)
- What outputs have been delivered in practice by the Pesticides Approval Committee (Netherlands Court of Audit 2002f, p 29)?
- How good are the records kept on the collections held in state museums, and how well are these collections conserved (Netherlands Court of, 2000b, pp 8-9)?

### 6.2.2 Criteria

There are no standard criteria for the achievement of outputs and outcomes, as do exist for the formulation of aims and policy information (see Chapter 5). The point is that, when auditing outputs, you will generally wish to compare the *desired* outputs and outcomes with the *actual* outputs and outcomes.

Information on the desired outputs and outcomes (i.e. the reference points) may be obtained from a range of different sources. Such information may be obtained, for example, from arrangements made in the past on the volume of outputs and outcomes, as in the case of arrangements made between central government on the one hand and agencies and/or autonomous administrative authorities on the other (see also Chapter 5 on the formulation of aims).

### 6.2.3 Audit design

Audits geared towards establishing whether a policy objective has been achieved are not designed to ascertain whether the fact that the objective in question has been achieved is due to the use of certain policy tools. This means that there is no need for a complex audit design.

All you need is either a system for measuring the various outputs and/or outcomes after the deployment of the relevant policy tools, or a combination of measurements before and after the use of the policy tools. The former (i.e. only a form of measurement after the event) is required if the official policy aim is formulated as follows: to achieve a given value in a particular variable (for example, a noise level not exceeding X decibels). A combination of measurements before and after the event is required if the official policy aim is as follows: to achieve an increase or



decrease in the value of a given variable (for example, a reduction of X decibels in the noise level).

In the absence of any specific arrangements in this connection, you can make pronouncements about outputs and/or outcomes on the basis of a comparison.

- *Simultaneous comparison*: this technique is particularly suited to situations in which there is more than one executive agency with either the same or *comparable* production processes and products (such as one type of inspectorate operating in a number of different regions); comparing the performance of the target organisation with that of other organisations is often referred to as 'benchmarking'. The 'best practice' or 'average practice' is the benchmark in this respect.
- *Sequential comparison*: this involves comparing the organisation's current performance with its past performance. The criterion generally applied in such cases is that current performance should be better than past performance. When making comparisons over time, you should bear in mind that changes in outputs are not simply due to the performance of the organisation itself, but also to other factors (such as an increase or decrease in the workload, e.g. the number of applications for licences).

#### 6.2.4 Performance indicators

Various indicators can be used for making pronouncements about the outputs delivered by a ministry, legal person with statutory tasks or local authority. The Social and Cultural Planning Office of the Netherlands (2001, p. 10) divides performance measures into the following two categories:

- *performance indicators*: these relate to the end products produced (such as the number of licences issued or the number of covenants signed);
- *process indicators*: these relate to the amount of work performed or the volume of intermediate products produced.

When interpreting indicators, you should be critical both of their quality and of the way in which they have been produced. They should in any event comply with the following three conditions:<sup>42</sup>

- they must be formulated as quantifiable variables;
- they must be reproducible;

---

<sup>42</sup> Court of Audit, 2002c.



- they must be able to be produced over a prolonged period of time, so as to enable chronological comparisons to be made.

You may find it useful, when using performance indicators, to distinguish between outputs *that are visible to the outside world* and *facilitatory* outputs (see also Chapter 3).

### 6.2.5 Outcome indicators<sup>43</sup>

The outputs delivered by firms, households and institutions may provide an indication of the social impact of a given policy (e.g. the number of high-tech start-ups<sup>44</sup> as an indication of the number of innovational firms). There are also *direct outcome indicators* (De Groot & Goudriaan, 1991) that provide an indication of the final outcome in society (such as the number of road fatalities, which may be used as an indicator of traffic safety in general). These indicators are designed to measure the end outcome, in terms of the impact either on those citizens for whom the product is intended or on society in general. *Utilisation indicators* (i.e. the degree to which the target group has been reached) may also be used as proxies (i.e. indirect indicators) for measuring certain outcomes (see next section).<sup>45</sup> An example of such an indicator is the number of market gardeners using energy-saving measures.

At the same time, data on throughputs and outputs delivered by central government can also serve as outcome indicators. This applies specifically in those cases in which such throughputs or outputs *contribute directly* to the achievement of the desired outcome (for example, the number of military sites where the soil has been decontaminated as an indicator of soil remediation on military sites). However, by no means all throughputs and outputs delivered by central government may be seen as indicators of the desired final outcome.

Whether performance data actually say anything about the final outcome depends also on the way in which the aim has been formulated. If we take the previous example of the government's policy on waste production, where the objective was formulated as 'preventing the production of waste', the figure for the 'decrease in the volume of waste'

---

<sup>43</sup> If you are planning to perform an audit of intermediate outcomes (i.e. outputs delivered by local authorities), you may find *Maten voor gemeenten* (Social and Cultural Planning Office, 2000) a useful tool.

<sup>44</sup> Defined as prospective business owners who wish to market a technically innovative product.

<sup>45</sup> You need to make sure you have enough of them, though. Otherwise, you may find yourself measuring the prescribed use rather than the intended use.



does not actually say much about the prevention of waste production, which means that it is not a good indicator of the outcome. If, however, the objective is reformulated to read 'reducing the volume of waste', this figure then says a great deal about the extent to which the objective has been achieved, and is therefore a good indicator of the outcome.

### 6.2.6 Recommendations

Possible recommendations include the following:

- Adjust the policy tools, so as to ensure that you do achieve your policy aims in the future.
- Adjust the date by which the policy aims are supposed to have been achieved.
- Adjust the policy aims themselves. This is a useful recommendation if the audit shows, for example, that the policy aims are overambitious. If this is the case, there is no point in retaining the same policy aims. The audit may also find that the policy aims are technically achievable, but only at an extremely high cost. Again, there is no point in retaining the policy aims in such a situation. It would be advisable to formulate less ambitious aims that come with an acceptable level of cost.

Recommendations may also be intended to remove the *causes* of failures to achieve certain policy aims (see section 3.3 for details of possible causes).

## 6.3 Audits of the reaching of target groups

### 6.3.1 Problem definition

There are two basic types of audit of an auditee's success in reaching the target group: target-group audits (in the strict sense) and facilitation audits.<sup>46</sup>

The former type of audit is designed to ascertain whether and how the groups targeted by the policy (i.e. members of the public, institutions and

---

<sup>46</sup> This term is used, *inter alia*, by the National Land Use Planning Agency in a publication entitled *Evaluatie van indicatieve plannen; Handleiding en begrippen* ('Evaluation of indicative plans; manual and glossary'), January 1990. Page 10 of this publication contains the following passage: 'A number of minimum conditions need to have been fulfilled before we can say that a perspective impacts on the decisions taken or the plans made by policymakers further down the line. The policy-maker needs to be familiar with the contents of the perspective, to have considered the perspective and to have agreed with it. If these conditions are met, the perspective may be said to have had an impact.'



firms) have been reached. The same question may be posed in respect of executive bodies that are not part of central government (such as provincial authorities, local authorities, legal persons with statutory tasks and autonomous administrative authorities); these are vital links in the policy chain, with central government at one end and the ultimate target group at the other.

A typical problem definition in relation to target-group audits is:

- *Has the target group been reached, in the desired manner and to the desired extent, during the period under review?*

A typical problem definition in relation to facilitation audits is:

- *Do the aims formulated by central government impact on other actors in the policy chain?*

### 6.3.2 Audit questions

Reaching the target group

The problem as defined in relation to target-group audits may be translated into the following four audit questions:

1. How has the target group been defined?
2. Has the right target group been chosen?
3. How does the auditee wish to reach the target group and what proportion of the target group does the auditee wish to reach?
4. Has the desired proportion of the target group actually been reached in the desired manner?

#### **1. How has the target group been defined?**

Although every policy aim needs to be achieved by human intervention, the target groups of government policy are by no means consistently defined in terms of individual people. Policy is often geared towards organisations: firms are encouraged to reduce their energy consumption, non-commercial institutions are expected to do their best to reduce absenteeism due to illness, and sports clubs need to attract more young people. This means that the first audit question should seek to ascertain how the target group has been defined.

#### **2. Has the right target group been chosen?**

When assessing whether the right target group has been chosen, you should take the word 'right' as referring to the arguments underlying the policy in question rather than as having a political connotation. For example, in the case of a scheme designed for low-income earners, you



might wish to ascertain whether the scheme itself specifies the same target group.

### **3. How does the auditee wish to reach the target group and what proportion of the target group does the auditee wish to reach?**

Before you can say anything about the degree to which the target group has been reached in practice, you first need to know how the auditee wishes to reach its target group. This can be defined in all sorts of different ways: 'coming into contact with', 'taking cognisance of', 'understanding', 'accepting',<sup>47</sup> 'taking part in', and so on. The usual requirement in relation to the proportion of the target group that the auditee wishes to reach is that a given percentage of the target group must have been reached by a particular time.

### **4. Has the desired proportion of the target group actually been reached in the desired manner?**

Only now can you set about answering the vital question as to whether the target group has actually been reached. Two types of audit question are relevant here. The first relates to the *state of affairs at a given point in time*. Typical audit questions would include:

- How large a percentage of people are aware of the existence of the housing benefit scheme?
- How large a percentage of people know where and how to apply for housing benefit?
- How large a percentage of those who are entitled to housing benefit are aware of their entitlement?
- How large a percentage of those who are entitled to housing benefit actually apply for one?

The second type of question relates to *changes in the proportion of the target group that is reached after a specific output has been delivered*. In most cases, this is a matter of looking at the use that is made of public-information material in informing people about their rights (such as the publication of a leaflet entitled 'You may be entitled to housing benefit') or encouraging them to display a particular type of behaviour (such as the publication of a leaflet entitled 'Environmental protection starts at home').<sup>48</sup> In these cases, a target-group audit involves identifying changes in the proportion of the target group that is reached after an

---

<sup>47</sup> See Netherlands Court of Audit, 1991b.

<sup>48</sup> NB: Generally speaking, audits of the impact of public information campaigns should contain the following four questions: what is the *reach* (i.e. level of awareness), in what way has public *knowledge* changed, in what way have public *attitudes* changed, and in what way has public *behaviour* changed?



information campaign has been mounted. The following are examples of relevant audit questions:

- Has there been an increase in the percentage of under-fours regularly visiting a post-natal clinic in the wake of the public information campaign mounted to this end?
- How many farmers have applied for a subsidy for land they leave fallow since the change in the subsidy regulations was announced?

#### Facilitation

In the case of facilitation audits, the problem definition may be translated into the following audit questions:<sup>49</sup>

- Which actors, apart from central government, are involved in implementing policy X? In other words, what is the nature of the policy chain in respect of policy X?
- How does central government expect the actors concerned to contribute to the implementation of policy X?
- Are the actors concerned familiar with central government's aims in relation to the implementation of policy X?
- Have the actors concerned taken these aims into account?
- Do the actors concerned agree with these aims? In other words, is there a shared set of problem definitions, viewpoints, values and ideas about potential solutions?
- Do the actors concerned act in accordance with central government's aims in relation to the implementation of policy X? In other words, does their behaviour correspond with these aims?

#### 6.3.3 Criteria

In order to make a pronouncement about the proportion of *either the target group or certain intermediaries that have been reached*, or about the change in the proportion of the target group that has been reached, you need to have detailed information on the following criteria:

1. the size of the target group that the auditee wishes to reach, or the change the auditee wishes to achieve in the size of the target group reached by the policy;
2. the way in which the auditee wishes to reach the target group;

---

<sup>49</sup> Similar audit questions were used in the audit of the government's policy on waste prevention (Netherlands Court of Audit, 1997). In this particular audit, the term 'facilitation' was used in a broader sense than is the case above: it referred not only to those responsible for implementing policy, but also to the ultimate target groups. The auditors sought to assess the policy's impact on firms, for example, by studying a number of corporate environmental plans.





3. the date by which the auditee wishes the above to have been achieved.

**1. The size of the target group that the auditee wishes to reach, or the change the auditee wishes to achieve in the size of the target group reached by the policy.**

The first step is to specify, in the form of a specific target number (such as a percentage or an absolute figure for the size of a target group or a group of intermediaries), or in terms of a desired trend (such as 'more' or 'less'), *the size of the target group* or the extent of the change the auditee wishes to bring about. This is not simply a matter of quoting numbers, but also of specifying the precise nature of the target group or intermediaries. After all, there is a considerable difference between a target group defined as comprising 'all residents of the Netherlands' and one defined as comprising 'all residents of the Netherlands aged between 18 and 24'.

**2. The way in which the auditee wishes to reach the target group. Obviously, the auditee must also have indicated *how* it wishes to reach its target group. In most cases, this will be defined in terms of knowledge and familiarity.**

**3. The date by which the auditee wishes the above to have been achieved.**

Finally, the auditee should have a clear idea of the *date* by which it wishes to have reached the target group or intermediaries. This may involve listing a series of dates making clear what proportion of the target group or intermediaries should have been reached by certain intermediate dates. If you then measure whether the desired proportion of the target group has been reached on the specified dates, no one can accuse you to starting to measure at too early a juncture.

In principle, detailed information should be available, in relation to the policy under review, stating how and by which date the auditee is planning to reach how much of which target group. In certain cases, however, you may be able to reach a judgement based on more general information. This applies, for example, to public information campaigns: the Government Information Service has stated that, broadly speaking, it is satisfied if a television commercial reaches at least 80% of the relevant target group (Netherlands Court of Audit, 1991).



#### 6.3.4 Audit design

If the desired target group and/or the desired impact has been formulated in terms of a criterion (such as 80% or X number of people) that must be met by a given date, a single measurement will suffice. If, however, you are interested in measuring a trend, you will need to perform two measurements, i.e. before and after the implementation of the policy in question.

#### 6.3.5 Target group indicators

You can use a *utilisation indicator* to indicate the degree to which the target group has been reached. A *utilisation indicator* says something about the users of a service (Social and Cultural Planning Office of the Netherlands, 2001, p. 10) and the degree to which the service in question is used.

#### 6.3.6 Recommendations

If the way in which the target group has been reached does not meet the criterion formulated, the auditors will usually urge the auditee to try and improve the way in which the target group is reached. This recommendation may take a number of different forms, depending on the situation:

- In its simplest form, the recommendation is geared towards the *tools used*. The auditors may recommend, for example, that the auditee should use other media or intermediaries to reach its target group (depending on the nature of the audit findings).
- Another possible recommendation is to make a *clearer definition of the ultimate target group or intermediaries*. The audit may suggest, for example, that either a tighter or a looser definition should be used.
- The recommendation may also relate to *the size of the target group and/or the way in which it is reached*. If problems are identified in these areas, you may wish to recommend either that the auditee should seek to reach a smaller target group or that the auditee should be less ambitious in the way it wishes to reach the target group. This kind of recommendations is particularly relevant if such lower ambitions can be realised at a substantially lower cost.



## 7 Effectiveness audits

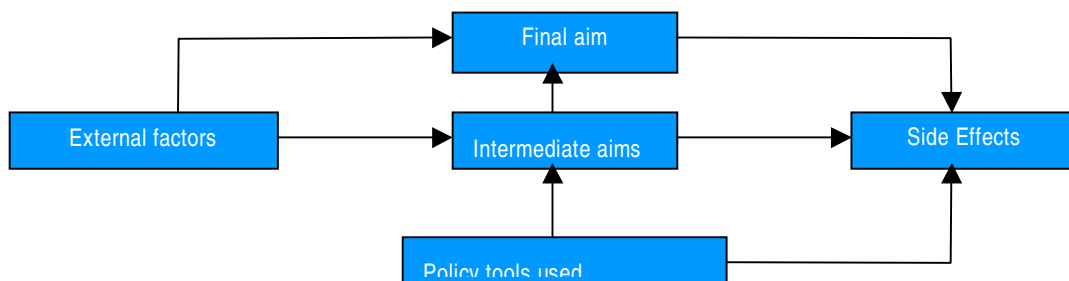
### 7.1 Introduction

We explained in Chapter 3 that there are three basic types of performance audits: those geared towards assessing the achievement of objectives or the reaching of a target group, effectiveness audits and efficiency audits. This chapter takes a closer look at the second of these: effectiveness audits. This is a complex type of audit, primarily because it seeks to establish a link between the outputs delivered by a minister and the subsequent social outcomes. Because the *policy chain may well be long*, and because *external factors* can easily affect the ultimate aims of policy, it is often difficult to prove the existence of a causal link between outputs and outcomes.

### 7.2 Problem definition and audit questions

Effectiveness audits seek to answer the following two (interrelated) questions (see Figure 7.1):

1. Have the policy aims been achieved thanks to the tools and inputs that have been used?
2. What possible explanations are there for the effectiveness (or ineffectiveness) of the policy in question?



**Figure 7.1: Model for analysing policy effectiveness**

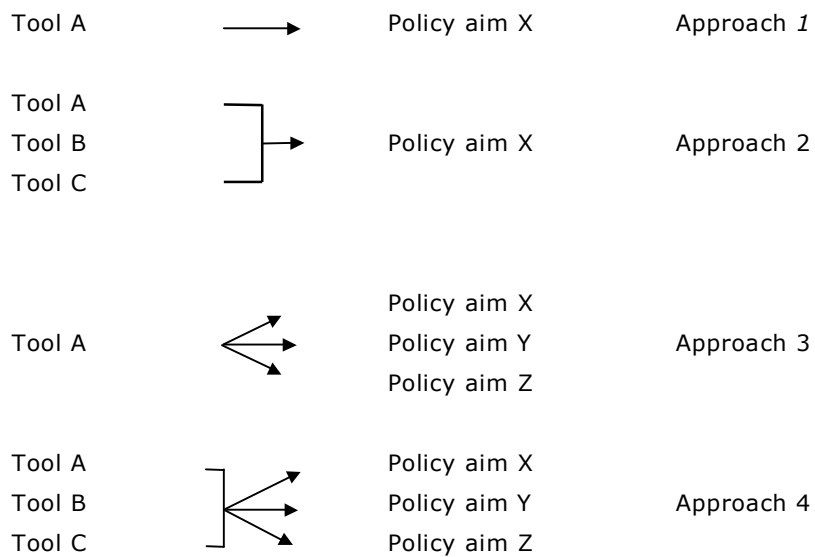
The first question is important in order to establish whether it was indeed the policy pursued by the minister or the legal person with statutory tasks



that produced the outcomes in question, or whether these were produced – either in full or in part – by external factors. If the outcomes are due solely to external factors, the minister clearly based his or her policy on the wrong assumptions: either the wrong tools were used or the tools used were not actually needed in order to achieve the minister’s policy aims.

The second question (which is in fact complementary to the first) needs to be answered in order to make recommendations. In other words, what action does the minister need to take in order to make his or her policy more effective?

Effectiveness audits involve examining any of a number of different relationships:



It is worth adopting the first approach, for example, if the government wishes to achieve a given aim with the aid of a single tool (rather than a set of tools) and if there is estimated to be a low risk of undesirable side effects. An example might be the use of a remigration scheme to encourage certain categories of immigrants to return to their native countries. The first approach is also relevant where a set of policy tools has been used to achieve a given policy aim, but where you wish to focus your audit on a single tool in order to make the audit more effective. The other tools will provide a context in such a situation.



You may wish to adopt the second approach if you are planning to examine a set of policy tools that have been used in relation to a single policy aim. In some cases, the tools are interwoven so closely as to make it very difficult to assess the role played by individual tools in the achievement of a policy aim. What you then do is, in effect, to assess the effectiveness of a mixture of policy tools.

Sometimes, however, it is possible to assess the effectiveness of individual policy tools. This was the case, for example, with an experiment in relation to the government's policy on energy-saving.

The effectiveness of tools used for the government's energy-saving policy: an experiment

An experiment performed by an energy supplier called EDON in 1993 and 1994 (supervised by B&A-group working on the instructions of the Ministry of Economic Affairs) sought to assess the effectiveness of two tools that could be used for implementing the government's energy-saving policy: public information and financial incentives (of either NLG 150 or NLG 75). The existence of six distinct groups of energy-users made it possible to compare the effects of the tools on energy consumption:

- a group that received public information only;
- a group that received a financial incentive of NLG 150 only;
- a group that received a financial incentive of NLG 75 only;
- a group that received a combination of public information and a financial incentive of NLG 150;
- a group that received a combination of public information and a financial incentive of NLG 75;
- a control group (see Hufen's article in the 8th information pack produced by the Focus Group, 1997).

When performing an audit along these lines, the auditor should preferably compose the groups on a randomised basis before the policy tools are actually used, so that they are as similar as possible in terms of their other characteristics (apart from the influence exerted on them by the policy tools, of course). If this is not possible, you can also compare the groups once they have formed spontaneously. There is a potential drawback here, in that they may also differ from each other in terms of other characteristics, which makes it less easy to attribute any differences between groups to differences in the degree to which they have been exposed to the policy tools in question.



As a matter of fact, EDON found that a combination of public information with the larger of the two financial incentives was the most effective tool.

You may wish to adopt the third approach if you expect the use of a given policy tool to be associated with certain effectiveness risks. This was the case, for example, with the Court's audit of the amalgamation of the National Centre for the Nursing Organisations with the Home Help associations.<sup>50</sup> The audit sought to establish the impact of the merger on (1) overhead expenses, and (2) the standard of service provided (in particular the elimination of waiting lists).

The fourth approach is the most comprehensive and is suited for use in situations in which a number of policy tools are closely interwoven and are also likely to be associated with a wide range of effectiveness risks. The Court's audit of subsidised labour is a good example of the use of this approach.<sup>51</sup> The audit involved assessing the impact of a number of tools used for combating unemployment on a range of policy aims (such as the creation of jobs and a reduction in the number of benefit claimants) and side effects (such as crowding-out).

### 7.3 Criteria

Naturally, effectiveness is the main gauge by which to measure effects. At the same time, there are no set standards of effectiveness, as there are also no fixed standards for measuring the achievement of policy objectives. This means that specific standards need to be established for every individual audit, depending on the aims of the policy under review. For example, if the purpose of merging the National Centre for the Nursing Organisations with the Home Help associations is to lower overheads, the criterion is that the policy tools used should indeed lead to a reduction in overheads (preferably by a specified percentage). It would seem logical to base an assessment of effectiveness on the ultimate aim that the policy in question is seeking to achieve. After all, the achievement of this particular aim is what it's all about. At the same time, you can easily find yourself facing a dilemma if you concentrate on ultimate aims. This is because it is often the case that: the higher you climb the policy tree, the more abstract are the terms in which the aim has been formulated, and hence the more difficult it is to gauge the extent to which the aim has been achieved;

---

<sup>50</sup> Court of Audit, *December Report 1993*.

<sup>51</sup> House of Representatives, 1996-1997, 25 070, nos. 1-2.



the higher you climb the policy tree, the more difficult it becomes to establish causal relationships. This is because, the higher you get, the greater is the likelihood of distorting factors coming into play.

The presence of these barriers may induce you to focus your attention on intermediate aims (such as the outputs of local authorities, legal persons with statutory tasks, firms, members of the public and institutions). This is all the more worthwhile if you can demonstrate the link between the ultimate and intermediate aims.

If you are also looking at side effects in your audit, you could stipulate a second criterion alongside effectiveness, which is that there should not be any significant adverse side effects.

An additional criterion that you could apply is that the following effects should be either completely or virtually absent:

*Gift effect*: this effect is seen with subsidies, for example. The subsidy recipient would have behaved in the same way even if he or she had not received the subsidy, or had received a smaller amount. This was the case, for example, with the single grant paid on owner-occupied houses.<sup>52</sup>

*Crowding-out effect*: this effect is often seen when labour-market tools are used: whilst some people obtain better jobs for themselves, they do so at the expense of others, either those already in employment or other job-seekers.<sup>53</sup>

*Benefit-snatching*: suppliers of a product that is subject to a subsidy (or other suppliers in the production chain) take advantage of a subsidy scheme to *raise their prices*. For example, when PCBs were replaced, prices for PCB waste disposal quadrupled in the period from 1984 to 1986. It is not clear whether the price rise was due in its entirety to the increased cost of waste removal.<sup>54</sup>

## 7.4 Audit design

In order to assess the degree to which the achievement of a given policy objective is due to the policy itself, you need to make a *comparison*.

<sup>52</sup> Example taken from Kraan-Jetten (1991). What you can in fact say here is that, although the policy aim has been achieved, the policy itself was not effective. At the same time, you can also see it as an example of an audit of policy efficiency: couldn't the same effect have been achieved with less money?

<sup>53</sup> See the Court's report on its audit of subsidised labour (Court of Audit, 1996a) for more information on this particular effect. Depending on the purpose of the tool in question (i.e. job creation in general or the creation of jobs for a specific target group), this outcome may be seen either as ineffectiveness or as an adverse side effect.

<sup>54</sup> Which means that it is in fact a special sort of side effect. The example is taken from Vermeulen (1992).



Ideally, such a comparison should consist of measurements taken before and after the implementation of the policy, plus a measurement taken among a control group that was not exposed to the policy. In practice, however, it is often difficult to make comparisons along these lines, partly because of a lack of material.

Audit designs can be ranked according to the degree to which they comply with the causality conditions set out in section 4.2.5. These designs are described in detail in the literature.<sup>55</sup> What we should like to emphasise at this point is that it is absolutely essential to make comparisons in order to measure or to estimate effectiveness. There are various ways of doing this. For example, you can make comparisons: *over time*: what was the score before and after the deployment of the policy tool in question?<sup>56</sup>

*between groups*: how did a group that was exposed to the policy tool score in comparison with a group that was not exposed to it?<sup>57</sup>

*between different types of (implementing) policy*: what works in what circumstances? In order to find this out, you need to have a detailed description of the various types of policy, and also of the different contexts in which the policy is implemented.<sup>58</sup>

*with a hypothetical situation*: what would the score have been if the policy tool had not been used? This you can find out, for example, by undertaking a subjective assessment, i.e. by asking questions such as: 'If you had not received the subsidy, would you still have modified your behaviour along the lines envisaged by the subsidy scheme?' This method is not very suitable, as there is a risk of receiving socially desirable answers that are not valid for this reason. However, if the interviewer is able to follow up with further questions, it can still give an indication of effectiveness if used in combination with other methods.<sup>59</sup>

*by combining* the above methods.

---

<sup>55</sup> In addition to the literature previously quoted in footnotes, the following works are relevant in this respect: P.H. Rossi and H.E. Freeman, *Evaluation, a systematic approach*, 1993; T. D. Cook and D.T. Campbell, *Quasi-Experimentation, Design and Analysis Issues for Field Setting*, 1979; Veldkamp Marktonderzoek b.v., *Naar een zilveren standaard; een aanzet tot richtlijnen voor effectonderzoek naar campagnes van de Rijksoverheid* ('Towards a silver standard: proposed guidelines for auditing the impact of central-government information campaigns'), 1993.

<sup>56</sup> Described as the 'before-and-after' approach in *Heeft beleid effect?* ('Is policy effective?'), Policy Analysis Division of the Ministry of Finance, 1999.

<sup>57</sup> Described as the 'with-and-without' approach in *Heeft beleid effect?* ('Is policy effective?'), Policy Analysis Division of the Ministry of Finance, 1999.

<sup>58</sup> See in this connection the article by R. Pawson and N. Tilley, 'Whither (European) Evaluation Methodology', *Knowledge and Policy*, Fall 1995, vol. 8, no. 3, pp.20-33.

<sup>59</sup> See P. Swanborn, *Evalueren onder de loep; op weg naar een standaardvoorziening voor evaluatie in de Arbeidsvoorziening* ('Evaluation under the microscope: towards a standard form of evaluation for job centres'), O&A werkdocument 95-03, 1995, p.72.





## 7.5 Audit methods

An *ex post facto* evaluation is the best way of assessing, or at least demonstrating, the degree to which the use of certain policy tools (such as subsidies, covenants and licences) has contributed to the achievement of certain policy objectives.

It may also be worth identifying plausible disruptive factors and finding out whether these had any effect on the case in point. The more you can exclude the role of disruptive factors, the easier it is to demonstrate the degree of effectiveness of a particular policy tool.<sup>60</sup> This is the essence of the *modus operandi method* (Latin for 'way of working'). The method is explained in detail in Annexe 6.

The designs that are best from a technical viewpoint are generally hardest to work with in practice. For example, it is often impossible to perform a pre-policy measurement because an evaluator is not contacted until after the policy has been implemented. Alternatively, a control group cannot be formed as the policymakers regard it as being ethically unacceptable to deprive certain target groups of the benefits of the policy in question.

Nonetheless, there are also occasionally opportunities for devising creative solutions: for example, you can use groups of people on a waiting list for comparison purposes. You can also use more than one method – each of which has its own strengths and weaknesses – at the same time, the idea being to find out whether they all show the same type of effects.<sup>61</sup>

---

<sup>60</sup> Cf. the sample audit performed by Audit Directorate I on the remediation of fuel-contaminated soil on land owned by the Ministry of Defence. Presumably, no other factors or actors are involved in reducing the degree of contamination apart from the corrective action taken by the Ministry of Defence itself. In this case, the achievement of the objective (= the effectiveness) follows from the strategic choice of the audit design.

<sup>61</sup> See the arguments given in favour of this approach by E. Chelimsky: "This means that we think *less* today about the relative merits of one method versus another and *more* about whether and how using them in concert could result in more conclusive findings." (in: 'Where we stand in the practice of evaluation', *Knowledge and Policy*, Fall 1995, vol. 8, no. 3, pp.8-19).

A good example of the use of a combination of methods is given in a report published by the Social Insurance Supervisory Board, on the findings of an audit of the effects and side effects of the introduction of financial incentives into the Sickness Benefits Act (J. Vijgen in the 8th information pack produced by the Focus Group, 1997). A telephone survey was conducted among employers to find out whether they were aware of the changes in the Sickness Benefits Act and also what sort of action they had taken in order to reduce the rate of absenteeism due to illness (now that they bore a greater responsibility). A number of other methods were also used: an analysis of existing figures on absenteeism, a calculation of the correlation between the number of measures taken and the rate of decline in the absenteeism rate, a chronological comparison of absenteeism rates among firms that had taken action with those among firms that had not taken much action, so as to identify autonomous trends in absenteeism rates, etc.



Details on the relevant methods and techniques are given in Chapter 4.

## 7.6 Effectiveness indicators

There are no off-the-peg effectiveness indicators, as there are for the achievement of policy objectives. The point is that an effectiveness indicator should make clear how the policy in question contributed to certain social outcomes, and this can only be measured by means of an *ex post facto* evaluation. In practice, therefore, indicators used for showing whether policy objectives have been achieved are often wrongly presented as effectiveness indicators.<sup>62</sup>

## 7.7 Recommendations

If you discover that certain policies are ineffective, you will need to know *why* in order to make recommendations for improvements. You will need to identify potential explanations at an early stage of the audit so that you can decide during the course of the audit whether they have actually affected the effectiveness of policy.

A wide range of potential explanations are quoted in the literature, including inaccurate assumptions, poor implementation practice, conflicting policy aims, overambitious policy aims, inadequate organisation of policy, and government bodies that are at loggerheads with each other (see also section 3.3).<sup>63</sup> Depending on the applicability of the above explanations, you may wish to make any of the following recommendations:

- Undertake an *ex ante* evaluation (or, if you have already undertaken one, improve its quality) before selecting and using policy tools. Consider not simply the outcomes you are seeking to achieve, but also the cost of alternative policy tools.
- Reassess the use of this tool in relation to the policy aim you are seeking to achieve.
- Reassess the use made of this tool (e.g. raise or lower the amount of subsidy granted).
- Reassess the mix of policy tools.
- Adjust the operational management, so that the policy is implemented according to plan.

<sup>62</sup> See also Court of Audit, 1997b.

<sup>63</sup> See in this connection: H. Bressers and P.J. Klok, *Handleiding instrumententheorie* ('Guide to policy tool theory'), 1988; A. Kraan-Jetten, *Effectiviteit van overheidsbeleid* ('Effectiveness of government policy'), 1991; W.J.V. Vermeulen, *De vervuiler betaald: onderzoek naar de werking van subsidies op vier deel terreinen van het milieubeleid* ('Paying the polluter: a study of the effects of subsidies in four areas of environmental policy'), 1992.



- Continue to use the policy tool in question, but redefine your policy aims, as these are overambitious.
- Try and avoid adverse outcomes by adjusting the policy tool in question.



## 8 Efficiency Audits

### 8.1 Introduction

In Chapter 3 we explained that there are three basic types of performance audits: those geared towards assessing the achievement of objectives or the reaching of a target group, effectiveness audits and efficiency audits. This chapter takes a closer look at the third and final type of audit: efficiency audits.

The chapter begins by discussing the format of efficiency audits (section 8.2) and the relevant terms (section 8.3), before going on to look at the four discrete stages encompassed by an efficiency audit, i.e.:

1. defining the subject matter (section 8.4);
2. collecting data (section 8.5);
3. calculating the efficiency (section 8.6);
4. looking for factors that account for any efficiency differences (section 8.6).

### 8.2 Format of efficiency audits

#### 8.2.1 Objective

The objective of an efficiency audit performed by the Netherlands Court of Audit could be formulated as follows:

*'Helping to raise efficiency by providing information on the degree of efficiency and accounting for differences in efficiencies.'*

As an additional benefit, by showing an organisation how to perform an efficiency audit, you can help improve the way in which it monitors its own efficiency (i.e. generates efficiency data).

#### 8.2.2 Problem definition and audit questions

The following problem definition follows from the objective as formulated above:



*'What differences in efficiency can you identify between (parts of) organisations and/or over time, and what are the reasons for these differences?'*

You can translate the problem definition into the following audit questions:

- Measuring efficiency:
  - ❑ What sort of outcomes or outputs (bearing in mind any differences in quality) does the organisation deliver? What sort of inputs does it use for this purpose? What quantities are involved?
  - ❑ Are there any differences in efficiency between the organisations or over time, and if so, how great are these differences?
  
- Explaining differences in efficiency:
  - ❑ What potential explanations are there for differences in efficiency?
  - ❑ What factors account for any differences in efficiency?

### **8.2.3 Criteria**

As we already explained in Chapter 1, you may wish to measure the efficiency of policy *outcomes* or *outputs*, regardless of the nature of the organisation or department responsible for delivering them.

If you are seeking to establish whether an organisation has delivered certain outputs or outcomes efficiently, what you really want to find out is:

- whether it could have produced the same outputs or outcomes with fewer inputs; or
- whether it could have produced more outputs or outcomes with the same inputs.

As we have already mentioned, the distinction being made is between the efficiency of outcomes and the efficiency of outputs.



### *Criteria for efficiency and efficiency data*

In order to determine the efficiency of outcomes or outputs, therefore, you need to have access to data on the three ingredients included in the above definition:

- *Data on inputs.* You should look at all the inputs used to generate the outputs or outcomes in question.
- *Data on the quantity of the outputs or outcomes delivered.* Policy outputs and outcomes must be measurable, and must be a logical consequence of the policy objectives.
- *Data on the quality of the outputs or outcomes delivered.* You must take account of any differences in quality between the outputs or outcomes delivered by the various organisations or at different moments in time.<sup>64</sup>

Moreover, efficiency is relative. Therefore there are no set standards by which to measure efficiency.

Finally, in order to measure efficiency and identify the reasons for any efficiency differences, you need to use the right techniques in the right way.

These criteria are discussed in detail in the remainder of this chapter.

#### **8.2.4 Recommendations**

An efficiency audit can indicate:

- the degree to which an organisation can improve its efficiency in relation to the chosen reference point (for example, an organisation might be capable of delivering 10% more outputs with the same quantity of inputs). In other words, efficiency data is capable of explaining how the same money can be used to do more for the general public, which is the ultimate aim of every government policy. Alternatively, it can indicate whether the same ends can be achieved with fewer resources or less money, in which case the savings made can be spent on a different policy;
- how the organisation can raise its efficiency. In other words, an efficiency audit can help improve the auditee's operational management.

---

<sup>64</sup> This means using homogeneous groups of outputs or outcomes. The term 'homogeneous' is taken to mean that the workload and/or quality of the outputs or outcomes must be the same in each group.



Clear efficiency data is also an important tool for reporting on the implementation of government policy and the outputs delivered in this process, for example to the House of Representatives.

Recommendations for improving efficiency should be based on factors that have a demonstrable impact on the level of efficiency. These should preferably be factors over which the organisation in question is capable of exerting some degree of influence (i.e. internal factors). Such factors offer the organisation an opportunity to adjust its policies. See also section 3.3.

### **8.3 Terms**

#### **8.3.1 Outcomes, outputs and inputs**

As we have already mentioned, three ingredients are involved in measuring efficiency: the quantity and quality of the outcomes or outputs, and the inputs used in order to deliver the outcomes or outputs in question. See section 3.2 for a definition of outcomes, outputs and inputs, and sections 1.1.4 and 8.2.3 for a definition of efficiency.

#### **8.3.2 Comparison**

Efficiency is relative. This means that policy outcomes and outputs are achieved either more or less efficiently as compared with those achieved by another, comparable organisation or as compared with those achieved in the recent past. In other words, all you can say is that an organisation has been either more or less efficient compared with another organisation or with another period in the past. It is not possible to make any absolute judgements about efficiency. For example, if a ministry's annual report contains an indicator showing the relationship between inputs on the one hand and outputs or outcomes on the other, it does not say anything about the ministry's efficiency unless it is compared with a specified reference point. A judgement about efficiency must be based on a comparison.

A typical form of comparison is with previous years. You can also make comparisons with previous months, quarters or half-years, although this is more difficult (and hence costlier) in terms of data collection.

If you decide to make a comparison with other organisations, it is important to use organisations that are genuinely comparable. In other words, they should pursue the same policies (for example, regional police



forces or a national network of hospitals or schools) or deliver comparable outputs (for example, a state school and a private school).

In summary, there are two basic types of comparison you can perform:

- comparisons with the past;
- comparisons with one or more other, comparable organisations.

You can also combine the two.

### **8.3.3 Quality**

When measuring efficiency, you should always take account of any differences in quality between the outputs or outcomes delivered by different organisations and/or at different times (see also section 3.2.1 in this connection). For example, certain outputs that seem homogeneous at first sight may prove to be heterogeneous upon closer examination. For instance, not all tax returns are alike, nor are all grant applications. It goes without saying that more resources are required to process complex tax returns and grant applications than are needed to process standard returns and applications. This will obviously result in a difference in the nature of the outputs delivered.

Experience also shows that differences in quality often come into play even when you are comparing outputs delivered in one year with the same type of outputs delivered in another year, or when you are comparing outputs delivered by one organisation with those delivered by another. You can only conclude that an organisation has raised its efficiency if it has delivered cheaper outputs of at least the same quality. For example, if the cost of a given output is lower than in the previous year because there has been a decline in quality, this does not entitle you to conclude that there has been an improvement in efficiency. When measuring efficiency, therefore, you should allow for possible differences in quality between the outputs delivered, so as to avoid the risk of drawing the wrong conclusions (see section 8.5.4).

It is often fairly difficult to measure the quality of outcomes or outputs resulting from policies pursued by central government or a legal person with a statutory task, because most of them deliver their outputs in the form of services. A common way of operationalising the quality of outputs is by measuring customer satisfaction. A good-quality survey is a useful means of generating the necessary data. If it is not possible to perform





such a survey, you can estimate the level of customer satisfaction, for example, by counting the number of complaints.

#### **8.3.4 Explanations**

You can measure the relative efficiency of policy outcomes or outputs by relating them to the inputs used to produce them, and comparing the resultant ratio with equivalent ratios for comparable organisations and/or for other moments in time. 'Efficiency scores' show when the organisation in question was performing with maximum efficiency and/or which organisations performed well compared with others.

Where organisations are now found to be performing relatively poorly, it is important to find out what they can do to raise their efficiency. Hence the need to identify factors that are capable of accounting for differences in efficiency, so as to make clear how an organisation can improve its own efficiency (see also section 3.3 in this connection). Examples of factors that are capable of accounting for differences in efficiency are the composition of the staff complement (broken down into age categories, and into part-time and full-time staff), wastage and the quality of operating processes.

Let's take as an example a situation in which you wish to compare two organisations pursuing the same policies in terms of efficiency. You find that the efficiency score for one of them is much higher than the other's. If you then discover that the latter organisation employs twice as many managers as the former, and that there are two departments at the latter organisation that work at cross-purposes with each other, these are both factors that could explain the differences in efficiency between the two organisations.<sup>65</sup>

#### **8.4 Defining the limits of the audit subject**

If you want to generate efficiency data on a given policy field, you will first need to decide what exactly it is whose efficiency you wish to measure. Do you want to measure the efficiency of all outcomes or outputs delivered for a particular policy article or legal person with statutory tasks, or are you interested only in measuring the efficiency of some of these? Are you interested more in outcomes than in outputs, or vice versa? These questions are discussed in the following sections,

---

<sup>65</sup> Clearly, this is a highly simplified example. In reality, efficiency is affected by all sorts of internal and external factors.



together with a proposal for a step-by-step analysis of an organisation's efficiency.

#### **8.4.1 General or operational objectives?**

Obviously, an analysis begins with the policy objectives whose efficiency you wish to measure. After all, these objectives describe the outcomes or outputs that the ministry or legal person is seeking to achieve with the aid of the policy in question.

The first decision you need to take in relation to a ministry is whether you wish to measure efficiency in terms of its general objectives (i.e. covering all outcomes or outputs) or in terms of one or more operational objectives (i.e. covering some of the outcomes or outputs). If you decide to go for one or more operational objectives, you then need to decide which of these to concentrate on. If there are a number of operational objectives, and you want to use them as a basis for producing data on the efficiency of the policy defined in the ministry's general objective, you will obviously need to look at all the ministry's operational objectives when calculating its efficiency.

If the ministry's policy aims have not been clearly formulated (i.e. in terms of the outcomes or outputs the ministry intends to deliver), you will not be able to calculate the efficiency of these outcomes or outputs. After all, you need to be able to compare the outcomes or outputs with the inputs used, and you cannot do this if you do not have a clear picture of the outcomes or outputs in question. In this case, the ministry will first need to adjust its policy aims.

In order to allow the efficiency of outputs or outcomes to be measured, the objectives need to be concrete, i.e. specific and measurable. Advice on the formulation of effective policy objectives is given in section 5.2, as well as in the guidelines on the formulation of policy aims and performance data published by the Ministry of Finance, and the Court's own guidelines on policy reporting.<sup>66</sup>

Practical experience (in 2004) shows that most general objectives are not defined in concrete terms.<sup>67</sup> This means it is not clear which particular

---

<sup>66</sup> Ministry of Finance, *Handreiking doelformulering en prestatiegegevens*, December 2000; Court of Audit, *Handreiking verantwoorden over beleid*, 2000.

<sup>67</sup> This is hardly surprising, given that the *Dutch Central Government Performance Data and Evaluative Studies Regulation* states that, if a general objective cannot be quantified or does



outcomes the ministry is seeking to achieve, and hence makes it impossible to measure their efficiency. What you can do in such a situation, however, is to fall back on the outcomes or outputs listed in the operational objectives that the ministry intends to deliver in order to achieve its general objectives (see also the following section).

#### **Example of a general objective and an operational objective**

The Ministry of Justice has formulated the following general objective for its policy on legal aid: 'to guarantee that people seeking justice who require expert legal assistance but do not have sufficient financial resources, are nonetheless able to obtain access to justice'. The first operational objective quoted in this context is: 'an adequate system of subsidised legal aid for people of limited means'.

### **8.4.2 Outputs or outcomes?**

You can gear your audit towards measuring either the efficiency of outcomes or the efficiency of outputs. You will not be able to measure the efficiency of outcomes in relation to all policy objectives. In some cases, this will prove too expensive. In other cases, you will not be able to quantify the outcomes or to demonstrate the degree to which a given policy leads to a given outcome. In practice, it is easier to analyse the relationship between the inputs used and the outputs delivered. See section 3.2 for further information on this point.

The second question you need to answer is whether you wish and are able to obtain information on the relationship between inputs and outcomes or between inputs and outputs. It is best to adopt a step-by-step approach in this respect.

- Start by deciding where the most opportunities lie for producing efficiency data. Which of these are the most suitable at present, in political and social terms, and for which of these is it easiest to generate efficiency data?
- If necessary, produce a clear description of the general and/or operational objectives of the policy in question.
- Try and devote most time and resources at the outset to assessing the efficiency of outputs, given that this is relatively easy to measure.

---

not offer any useful reference points for policy management purposes, a policy article must then contain one or more operational objectives.



- First produce data on the quantity and quality of policy outputs before producing data on the efficiency of these outputs.
- In the meantime, you can start thinking about collecting data on outcomes and the efficiency of outcomes, in so far as this is feasible in both practical and financial terms.

## 8.5 Collecting data

Once you have defined the subject on which you wish or are able to compile efficiency data, you then need to start collecting the requisite data on:

- the outcomes or outputs, and the inputs used in order to deliver these outcomes or outputs (see section 8.5.2);
- the quality of the outcomes or outputs (see section 8.5.3);
- potential explanations for any differences in efficiency (see section 8.5.4).

Obviously, you need to ensure that these data are of good enough quality, i.e. relevant, reliable and comparable (see also section 5.3<sup>68</sup>). We shall start by explaining what we mean by this.

### 8.5.1 Quality of data

Data can be used to measure efficiency only if they are *relevant*, *reliable* and *comparable*.

#### *Relevance*

There is no point in collecting redundant data. The type of data you do collect depends on the nature of the policy field in question, and the question you need to answer in this respect. Data are 'relevant' if they accurately reflect the policy pursued by or the activities undertaken by the ministry or organisation in question.

#### *Reliability*

The data you collect should not contain any material inaccuracies or omissions.

---

<sup>68</sup> Three criteria set out in section 5.3 are not relevant here because the data only need to be relevant to the audit itself. These criteria are comprehensibility (i.e. to readers), conformity with the rules in terms of content, and conformity with the rules in terms of form.



### *Comparability*

The data you collect should enable comparisons to be made over time and/or between different organisations.

Depending on the way in which you make comparisons, the outcomes or outputs whose efficiency you are measuring and the way in which these are achieved should be comparable for all the organisations in question. Similarly, the efficiency indicators describing the relationship between inputs on the one hand and outcomes and outputs on the other should be defined in exactly the same way for all the organisations and/or for all the periods of time you wish to compare. Uniform definitions should also be available for the data you need in order to calculate the efficiency indicators.

#### **Examples of comparable data**

If you wish to compare the efficiency of the various Centres for Work and Income, the indicators you use should be defined in exactly the same way for each individual Centre. For example, each Centre should define, calculate and record in the same way its costs and the number of jobseekers for whom it has found jobs.

Should you wish to compare the efficiency of the Centres for Work and Income with that of another organisation, you should consider very carefully beforehand whether the latter organisation is fully comparable. You can make a relevant comparison only if their outputs and operating processes are comparable, and this will not generally be the case.

If you wish to compare the efficiency of the various Centres for Work and Income, the indicators you use should be defined in exactly the same way for each individual Centre. For example, each Centre should define, calculate and record in the same way its costs and the number of jobseekers for whom it has found jobs.

Should you wish to compare the efficiency of the Centres for Work and Income with that of another organisation, you should consider very carefully beforehand whether the latter organisation is fully comparable. You can make a relevant comparison only if their outputs and operating processes are comparable, and this will not generally be the case.

It is important to devote a great deal of time and energy at the outset to the quality of your data. This is usually a very time-consuming business. You are advised to make use of any expertise that is already available (for example, a monitoring committee).



#### Collecting data takes time

The Social and Cultural Planning Office examined the efficiency of general hospitals and university teaching hospitals in 1998. It took a whole year to sort out an already existing collection of hospital output data. During the Court's 2001 audit of the efficiency of job centres, the data collection stage also took about a year.

You always need to assess the quality of the data you collect (see also section 5.3). Depending on how sure you are about the quality of the data, the nature of this analysis can range from very general to extremely detailed. Section 8.6 describes how you can use an indicator analysis to gauge the quality of the data you are planning to use.

### 8.5.2 Data on outcomes or outputs

Outcomes are defined as the effects of policy that are perceived by the general public. Most outcomes do not lend themselves to a direct form of measurement. In such cases, you will need to estimate the outcomes by using indicators (also referred to as 'proxy variables'). If you are using outputs as outcome indicators (which is not a good idea), you can measure the efficiency of the outcomes by measuring the efficiency of the outputs.

Although it is generally easier to measure outputs, these too are not always capable of direct measurement. See sections 3.2 and 4.3.2 for definitions, examples and further information on the measurability of outcomes and outputs.

### 8.5.3 Inputs

It is important to make clear exactly which inputs have been used to deliver the outcomes or outputs in question (see also section 3.2.4). There are two points you need to bear in mind in this connection.

Firstly, you should in theory take account of all inputs that have been used to deliver the outcomes or outputs in question. Inputs may be either fixed or variable. Variable inputs include staff and equipment, whilst fixed inputs comprise resources such as buildings over which the organisation in question is unable to exercise much control in the short term. Because of the difficulty of making changes to fixed inputs, they are sometimes ignored for the purpose of calculating efficiency scores. In such situations, only variable inputs (or costs) are taken into account.



#### All relevant inputs

The Ministry of Education, Culture and Science cites the 'cost per pupil' as an indicator in a number of its education policy articles. Whilst an indicator can be a useful means of identifying trends in efficiency, this particular indicator is not capable of measuring efficiency in its present form, partly because it does not take all inputs into account. The inputs included in the indicator are the transfer payments made by the Ministry to schools. This means that the indicator does not take account of costs incurred by schools that are not covered by these transfer payments. The indicator suggests that a reduction in the level of expenditure by the Ministry can be equated with an increase in the efficiency of teaching at schools. This is not the case, however, if there has not been a corresponding decrease in the level of cost incurred by schools. This illustrates the importance of including all inputs in the calculation.

Secondly, whilst the calculation should include all relevant inputs, it should not include any more inputs than those used for delivering the outcome or output in question.

#### All inputs and *only* those used for delivering the output in question

The 2003 budget for the Ministry of Housing, Spatial Planning and the Environment contains a policy article (no. 12) entitled 'Management of environmental risks associated with chemicals, waste and radiation'. This policy article revolves around the need to control the public and environmental health risks associated with the use of chemicals, waste, radioactive substances and radiation, and genetically modified organisms, taking account of social and economic factors. The budget includes an efficiency indicator for the import, export and transit of waste, i.e. the cost price of each notice issued under the EU Waste Shipment Regulation. However, this figure does not include all costs involved in the issue of notices under the EU Waste Shipment Regulation, such as the Ministry's internal expenditure (consisting mainly of staff expenses). Equally, not all of the costs included in the cost price relate to the output in question. In other words, not only does the cost price not take account of all relevant inputs, it also includes more than just the inputs used for the purpose of delivering the output in question. Both problems affecting the composition of the cost price need first to be resolved in order to produce a useful efficiency indicator.

As we have already mentioned (see section 3.2.4), both cost and expenditure are used in practice for expressing inputs in monetary terms. Ideally, however, cost should form the basis for measuring efficiency. If you are making comparisons over time, you should really express the cost in constant prices, in order to take account of the distorting effects of inflation. As we have already said, you can also use variable costs for this purpose.



#### Difference between cost and expenditure

If you buy a number of computers in a particular year, you pay for them in full when you buy them (expenditure). However, given that computers have an economic life lasting several years, you will use up only part of their economic life during that year (cost). If there is a difference between expenditure and cost – as is often the case – any attempt to link expenditure with either outputs or outcomes will lead to the wrong conclusions about efficiency. This is because there can be wide fluctuations in expenditure from year to year. For example, expenditure will be high in the year in which the computers are bought. If the figure for expenditure in the following year were then taken as the basis for measuring efficiency, this would lead to the erroneous conclusion that there had been an improvement in efficiency in the following year.

Costs, too, do not always lend themselves to direct measurement, in which case they too will have to be measured with the aid of an indicator or a proxy variable. For example, in the current budgetary system, the only available information is often data on expenditure. The expenditure figures could be used as a second-best solution, as a proxy for the cost. However, in order to do this, you first need to establish whether there are any major items of expenditure that fluctuate widely from year to year and are hence capable of distorting the picture. If this is the case, either you should distribute the expenditure over the years in which the goods or services in question are 'consumed' or you should not take them into account (i.e. expenditure less investment).<sup>69</sup>

#### 8.5.4 Quality

You should avoid lumping together outputs and outcomes that differ from each other (see also section 3.2). If you treated all tax returns processed by the Tax and Customs Administration as being intrinsically identical, you would fail to take account of the fact that tax officials spend less time processing standard returns than they do dealing with difficult cases. If there is a difference in the proportion of 'easy' to 'difficult' outputs delivered by the organisations you wish to compare, or from year to year, you will need to differentiate between easy and difficult outputs.

<sup>69</sup> See Aarts et al. (2002) for details on the difference between cash accounting (i.e. the current budgetary system) and accrual accounting (which does generate information on outputs and cost prices).





#### Easy and difficult outputs

One of the duties of the Centres for Work and Income (i.e. one of their policy aims) is to help unemployed people find jobs. Let's assume that when an unemployed person signs a contract of employment, this is regarded as constituting an output. Finding a job for a jobseeker who is of less than average employability requires more time and energy than it does for someone of higher than average employability. In order to compose groups of homogeneous outputs, you need to distinguish between different groups of jobseekers, according to their degree of employability.

You also need to take account of quality differences if you are planning to compare outcomes or outputs delivered in one year with the same outcomes or outputs delivered in another year.

#### Quality differences

Let's assume that a given department delivers a single output, i.e. it processes tax returns. If a comparison of two branches of the same department, both of which process the same type of tax return, shows that one branch processes its returns quickly but with a large number of errors, whilst the other branch is slower but makes fewer errors, you would not be justified in looking solely at the cost per output. If you did so, you might be led to the false conclusion that the first branch is more efficient than the second.

Not only when you are collecting data, but also when you are analysing your data and drawing conclusions about the trend in efficiency, you need to take account of any differences between outcomes and outputs.

### 8.5.5 Finding and using explanations

Once you have analysed the efficiency of a ministry or another organisation (or of a department within a ministry or organisation), the next step is to identify the factors that account for any efficiency differences. This is because you cannot start working on efficiency improvements until you have identified the relevant explanatory factors. See also section 3.3.

#### *Internal and external factors*

Explanations may originate from within the ministry and/or the organisation involved in implementing the policy in question. These are known as internal or endogenous factors. However, explanations may also originate from beyond the organisations themselves, in which case they



are known as external or exogenous factors. Examples of explanatory factors are given in the following table.

Conclusion for which explanation is sought	Possible internal factors	Possible external factors
Decline in efficiency compared with previous year	<ul style="list-style-type: none"> <li>• High staff turnover</li> <li>• Recent reorganisation</li> <li>• High rate of absenteeism due to illness</li> <li>• Changes in the production process</li> <li>• More complaints</li> <li>• Inconsistent policy aims</li> </ul>	<ul style="list-style-type: none"> <li>• Production process affected by more interim requests from the House of Representatives</li> <li>• New legislation requiring changes to be made to the production process</li> <li>• Tighter labour market</li> <li>• Production process affected by poorer weather</li> <li>• Decline in state of domestic or international economy</li> <li>• More traffic congestion</li> <li>• Production process affected by deterioration in soil conditions</li> <li>• Changes in customer characteristics</li> <li>• Inconsistent policy aims</li> </ul>
Organisation A is more efficient than organisation B	<ul style="list-style-type: none"> <li>• Differences in operating methods</li> <li>• Differences in overhead expenses</li> <li>• Differences in age structure of staff</li> <li>• Differences in staff qualifications</li> <li>• Differences in average salary</li> <li>• Differences in staff experience</li> <li>• Differences in proportion of part-time staff</li> <li>• Differences in number of complaints received</li> <li>• Differences in amount of floor space per employee</li> <li>• Different number of sites</li> <li>• Differences in proportion of work outsourced to subcontractors</li> </ul>	<ul style="list-style-type: none"> <li>• Differences in local labour market</li> <li>• Differences in state of local economy</li> <li>• Differences in soil conditions</li> <li>• Differences in customer characteristics</li> <li>• Differences in age structure of local population</li> <li>• Differences in qualifications of local population</li> </ul>



Generally speaking, internal factors represent opportunities to raise efficiency, as the ministry or organisation in question can do something about them. External factors, on the other hand, are not generally capable of being influenced by the ministry or organisation in question. Whilst such factors do account for a given level of efficiency or inefficiency, they do not represent any opportunities for improvement.

#### External factors are hard to influence

A study performed by Goudriaan et al. of the efficiency of colleges of higher professional education (1998) revealed that the bulk of the efficiency differences were accounted for by changes in student numbers, which is a factor that is virtually beyond the control of individual institutions. Although explanations are the vital ingredients of recommendations for improvements, it was not possible in this case to make any recommendations in relation to this particular factor.

When collecting data on explanatory factors, it is worth making as much use as possible of any personal expertise already available at the ministry or organisation in question. You may also find it useful to consult the literature or spend a day watching the staff of the organisation at work.

#### Using the findings

If you identify certain opportunities for improvements, it makes sense that the audited organisation uses these for this purpose. If they relate to the efficiency of non-central government organisations that are involved in the implementation of government policy (such as local authorities), ministers can encourage the organisations in question, for example by means of consultation and/or regulation, to make the necessary improvements.

### 8.6 Calculating efficiency

Once you have defined and collected the necessary data, the next step is to calculate the efficiency of the outputs or outcomes under review. There are various ways of doing this, the most common of which are the use of simple indicators and the following analytical techniques: free disposable hull (FDH), data envelopment analysis (DEA) and stochastic frontier analysis (SFA).<sup>70</sup>

<sup>70</sup> See, for example, Coelli (1998), Blank (2000) and New South Wales Treasury (2001).



We will first discuss the most commonly used indicators (section 8.6.1) before describing in brief the use of more complex techniques, i.e. FDH, DEA and SFA (section 8.6.2). In theory, the latter are highly suited for calculating the efficiency of a relatively large number of comparable organisations or calculating the efficiency of a single organisation at a number of different times. The chapter concludes with a general description of various ways of identifying explanations for efficiency differences that can help the organisation in question to raise its efficiency (section 8.6.3).

### *Prologue*

When calculating efficiency, you can try and work out:

- whether the policy outputs or outcomes in question could have been delivered with fewer inputs; or
- whether the same inputs could have delivered more policy outputs or outcomes.

The first perspective is referred to as an 'input orientation', and the latter perspective as an 'output orientation'.

Whichever perspective you adopt from which to calculate the efficiency of an organisation, you first need to decide which of the two orientations is most suitable, i.e. an input or an output orientation. Generally speaking, the choice depends on the inputs or outputs<sup>71</sup> over which a manager of the organisation in question has the most control. After all, these are the factors which he or she can change. If the inputs are fixed, you should adopt an output orientation, i.e. you should try and see whether the same inputs can be used to produce more outputs.

#### Output orientation

An example of a situation in which you should adopt an output orientation is an audit of schools. Schools operate with fixed budgets (inputs).

If, however, the outputs are fixed, you should adopt an input orientation, i.e. you should seek to establish whether the same outputs can be produced with fewer inputs.<sup>72</sup>

<sup>71</sup> In order to improve the readability of this manual, we will henceforth refer only to 'outputs' rather than to 'outputs and outcomes'.

<sup>72</sup> For further details, see Coelli (1998), Blank (2000) and Netherlands Court of Audit (2001).



#### Input orientation

An example of a situation in which you should adopt an input orientation is an audit of electricity companies. Demand for electricity (i.e. the companies' output) is a fixed constant over which the companies are virtually incapable of exerting any influence.

Analyses based on indicators (and/or simple regression analyses) are always a good starting-point for measuring efficiency and identifying explanations for efficiency differences, even if indicators are not the best means of calculating efficiency. The fact is that this type of analysis generates a good initial impression of the structure of the organisations in question, as well as of a number of aspects of the relationship between inputs and outputs. See also sections 8.6.2.1 and 8.6.3.2 in this respect.

Annexe 7 contains a checklist that you can use to decide whether you can measure the efficiency of a given subject.

### 8.6.1 Efficiency indicators

#### 8.6.1.1 What are efficiency indicators?

Efficiency indicators are measures of the relationship between inputs and outputs at a given moment in time. A figure representing the 'cost per grant' is an example of an indicator. You can identify the relative degree of efficiency by comparing indicators for a number of comparable organisations and/or by comparing indicators for a single organisation with those for periods in the past.

#### Comparing indicators

	1998	1999	2000	2001	2002
Cost per grant (in euros, in prices as at 1 January 2002)	200	180	210	225	240
Quality: customer satisfaction (on a scale from 1 to 10)	8	6	6	6	8

The time series points to a rise in the cost per grant since 1999. Measured in terms of customer satisfaction, the quality remained the same during this period – with the exception of 2002, a year in which there was a slight improvement in quality. It is clear from these two trends that there was a decline in efficiency in the period from 1999 to 2001. The picture for 2002 is not immediately clear as an increase in the cost per grant was offset by an improvement in the level of quality.



The *advantages* of indicators are that:

- they are easy to calculate;
- they can be readily used to gauge efficiency and trends in efficiency;
- they can be produced at regular intervals (such as monthly or annually) without much trouble or cost, which means that they are a source of frequent information and offer opportunities for a swift response;
- in a situation involving more than one outcome or output, they can give information on each one.

#### 8.6.1.2 *When should you use indicators?*

##### **A single output or outcome**

Indicators are handy for calculating efficiency in those cases where there is just a single output or outcome for each policy article or aim, and where all inputs can be linked to the same single output or outcome. Such indicators are known as 'single indicators' (see above example).

##### **A range of outputs or outcomes**

In most cases, however, policy aims are associated with a number of different outcomes or outputs. If you know which inputs have been used to generate each individual outcome or output, you can construct a single indicator for each one.

##### *Problems and potential solutions*

If it is not possible to ascribe the inputs used to individual outcomes or outputs, you may well be able to use timekeeping data. You should bear in mind, though, that this is relatively labour-intensive work. Timekeeping data show the number of hours spent on the various outcomes and outputs by the staff involved in the implementation of the policy. By linking the number of hours worked to the salary earned by the staff in question, you can estimate the cost per outcome or output. Although this does mean disregarding non-staff expenditure, you can use allocation formulae as a rough means of allocating these to the various outcomes or outputs.

If you do not have access to timekeeping data, you will have to use other means of producing indicators, such as:



- using allocation formulae to allocate inputs directly to the various outcomes and outputs;
- weighting the various outcomes and outputs to produce a single composite outcome or output. By linking this composite figure to the inputs, you can then produce an indicator which you can use as an efficiency measure. This is known as a *composite* indicator.

### **Don't forget to account for quality**

Ideally, your indicator should take account of the quality of the outputs or outcomes delivered. In practice, however, this is unlikely to be the case. As a good alternative, you can design one indicator for the cost per output or outcome and another indicator for the quality of each output or outcome.

#### *8.6.1.3 How can you draw conclusions from indicators?*

##### *Allocation formulae or weighting*

If you use allocation formulae or a method of weighting, you can never be sure that these are exactly right. After all, such formulae or weighting factors are generally chosen on an arbitrary basis. The risk of using them is that you may unwittingly either iron over or exaggerate efficiency differences. For example, you may allocate irrelevant inputs to a particular outcome or output, resulting in the outcome or output becoming overpriced and hence wrongly being labelled as 'inefficient'.

Even if you have constructed a composite indicator with the aid of weighting factors (i.e. you have weighted the various outcomes and outputs to form a single outcome or output), you still need to be careful, as you don't know which particular part of the policy process accounts for the score recorded for the indicator. It then becomes difficult to find opportunities for raising efficiency, as you can't tell how each individual sub-process contributes to the efficiency (or inefficiency) of the policy process as a whole.

##### *Interaction between sub-processes*

Even if there are a number of different outcomes or outputs and you have succeeded in constructing efficiency indicators for each one, you still need to overcome a number of hurdles. Although this set of indicators is capable of generating information on the various parts of the 'production process', you still do not have a clear picture of the way in which the various sub-processes *interact* with each other. Let's assume that you



make a comparison over time. Although you can discern the trend in the cost of each individual output, you don't know for sure whether the policy process used to deliver the output is connected with any other processes. If the sub-processes used to deliver individual outcomes or outputs are not isolated from each other, you have to be careful as there is a risk that improvements made for the benefit of one particular outcome or output result in another outcome or output becoming less efficient.

#### Sub-processes influence each other

The outputs delivered by universities are research and teaching. However, it is not always clear how much money is spent on the production of each type of output. Moreover, the research work undertaken by university staff helps to improve the content and quality of their teaching. So how can you devise indicators for the inputs used in the production of each output? You can, of course, use allocation formulae and weighting factors to allocate the costs on a *pro rata* basis to teaching and research. Let's assume that the formulae you use are an accurate reflection of reality, and generate the following information on the trends in costs and quality:

	1998	1999	2000	2001	2002
Cost per teaching unit (in euros, in prices as at 1 January 2002)	1000	1200	1400	1500	1600
Cost per research unit (in euros, in prices as at 1 January 2002)	4950	5250	5400	5500	5600
Quality of teaching (results of peer review, on a scale from 1 to 10)	5	6	7	8	9
Quality of research (results of peer review, on a scale from 1 to 10)	4	6	7	8	9

The time series points to an increase in both the cost per teaching unit and the cost per research unit during the period from 1998 to 2002. The quality (measured in this case in terms of the score awarded each year by a peer review board) also improved during this period. This set of indicators therefore gives you a picture of the various parts of the production process.

At the same time, the various parts of the production process may not be viewed separately from each other. As we have already said, the research work undertaken by university staff helps to improve the content and quality of their teaching. Let's assume that both the rise in the cost per research unit and the improvement in the quality of teaching are caused in their entirety by an improvement in the quality of research. What happens is that, whilst the efficiency of research remains stable, the rise in the cost per teaching unit points to a decline in the level of efficiency. In this case, the opportunities for improvement are on the





teaching side.

But what if you did not know that teaching was influenced by research work? You might then have concluded that the efficiency of both research and teaching had remained the same during the period from 1998 to 2002. In other words, you have to be familiar with the underlying processes before you can draw the right conclusions and make constructive suggestions for improvements.

### *Contradictory indicators*

There is plenty of scope for drawing the wrong conclusions when interpreting a set of indicators. Moreover, different indicators can easily send out contradictory signals. In other words, you should take great care when drawing conclusions. Unequivocal conclusions may be drawn only if the trend in the 'efficiency' of all indicators is the same. For example, if you are using one indicator for the cost per outcome or output and another indicator for the quality of the same outcome or output, you can make a clear pronouncement on the trend in efficiency only if both indicators point in the same direction (i.e. both show either a falling or a rising trend).

Where you are faced with opposing trends, a pragmatic solution – albeit second best – is to weight the trends in question. This means deciding how much weighting to attach to one trend relative to the others (for example, a 2% increase in the cost per output is taken as being equivalent to a 1% decline in quality). See above for the risks attached to weighting.

### *Weighting opposing trends*

Returning to the previous example about university outputs, we have weighted the trends in relation to research. During the years between 1998 and 2002, the cost per research unit rose by approximately 6%, 3%, 3% and 3% respectively. The quality also improved during the same period, by 2%, 1%, 1% and 1% respectively. A 1% rise in quality is regarded as being equivalent to a 3% rise in cost. This means that the university's research efficiency remained the same throughout this period.

### *An example of how not to do it*

An efficiency audit performed by Bakkenist Management Consultants into home-care services provides a good example of the risks associated with the use of indicators. The study encompassed seven home-care



organisations, and involved the use of indicators.<sup>73</sup> A number of the findings are set out in the following table.<sup>74</sup>

	1	2	3	4	5	6	7
Number of hours of care provided by district nurses per FTE	966	<i>1075</i>	960	na	726	865	954
Number of hours of care provided by home helps per FTE	1421	1401	1434	na	<i>1490</i>	1449	1477
Number of hours of care per NLG 10,000 budgeted	<i>261</i>	212	202	225	225	213	237
Proportion of nursing staff (as % of total)	<i>93</i>	88	92	91	89	91	91
Proportion of productive staff (as % of nursing staff)	80	75	71	<i>81</i>	76	77	78
Managerial staff per 100 FTEs of nursing staff	<i>0,92</i>	1,06	0,67	0,85	<i>0,39</i>	0,91	0,56
Administrative staff per 100 FTEs of nursing staff	<i>1,47</i>	2,66	2,67	2,37	2,50	2,92	2,91
Computer staff per 100 FTEs of nursing staff	0,76	0,97	0,84	<i>0,52</i>	0,76	0,65	0,71
Support staff per 100 FTEs of nursing staff	<i>1,34</i>	3,61	3,52	2,67	2,07	3,19	1,99
Rate of absenteeism due to illness (as %)	11	<i>7</i>	11	12	11	9	11
Proportion of temporary staff (as %)	15	2	3	<i>19</i>	15	4	8
Proportion of part-time staff (as %)	46	57	48	44	58	<i>42</i>	61

The researchers deduced a best-practice figure for every indicator in the table, and this is the figure in each row that is printed in italics and cross-hatched. The researchers decided that the set of best-practice figures could together be regarded as forming a virtual organisation that could serve as a model for all the other organisations. For the sake of convenience, they assumed that the processes underlying the various indicators were all performed independently of each other. They did not seek to establish whether or not this was true. For example, there is no reason to assume that organisation no. 4 should be regarded as the trend-setter in terms of the proportion of temporary staff (let alone whether it is correct to assume that a higher proportion is better for overall efficiency). Whether coincidentally or not, the fact is that organisation no. 4 also has the highest rate of absenteeism due to illness. The researchers did not think of asking what the reasons were for these differences, let alone actually come up with explanations.

In other words, those organisations that decide to adjust their processes on the basis of the findings of this study could well find themselves actually worse off than before.

<sup>73</sup> For a critical discussion of the study, see *Benchmarken of de kunst van het vergelijken* ('Benchmarking, or the fine art of comparison', Blank, 1998).

<sup>74</sup> The first three rows in the table contain general information on efficiency (the outputs are the number of hours of care provided, whilst the inputs are expressed in terms of FTEs as a second-best solution). Whilst the rest of the data in the table do not contain any meaningful information on efficiency, it could explain certain differences in the level of efficiency.



*Another example of how not to do it*

The analysis by Taphoorn et al. in 1998 of the efficiency of the Custodial Institutions Service during the period from 1994 to 1996 is another example of the infelicitous use of indicators. The researchers in question devised five indicators for measuring the changes in the efficiency of the Dutch prison service:

- the cost price per day;
- the budget outturn, i.e. profit or loss;
- the occupancy rate, i.e. the average occupancy divided by the average available capacity;
- the staff ratio, i.e. the number of staff compared with the number of prisoners;
- the overhead ratio, i.e. indirect expenditure compared with aggregate expenditure.

Indicators	1994	1995	1996
Cost-price per day (1994 = 100)	100	93	91
Budget out-turn (in NLG million)	- 14,3	- 8,5	1
Occupancy rate	100,5	100,5	88
Staff ratio (as %)	96	92	98,6
Overhead ratio (as %)	n.a.	8,2	9,4

The researchers also examined the changes in quality, which they measured in terms of the number of complaints for every one hundred occupied cells. They found that the number of complaints affecting all types of institutions within the prison service either stabilised or decreased during the period under review. The researchers concluded that, on balance, the five indicators pointed to a definite improvement in efficiency in 1995 and 1996. They were unable to say by exactly how much efficiency had improved, though. At the same time, they did give a number of explanations for the efficiency improvement, claiming that the main factors were the fact that institutions had been granted agency status (leading to greater opportunities for creating reserves, opportunities to reallocate equipment and staff at will, and more freedom in relation to premises) coupled with capacity increases (leading to savings in design and building costs, and economies of scale).

Again, this example illustrates some of the risks associated with the use of indicators.<sup>75</sup> Despite the *opposing trends* in the various indicators, the

<sup>75</sup> A method such as DEA would have been a better way of measuring the efficiency of a range of outputs and inputs in both this and the previous example. See section 8.6.2 and Blank (2003).



researchers nonetheless concluded that they pointed to an improvement in the level of efficiency. This is in spite of the fact that the underlying trend is not always the same. Moreover, the researchers failed to recognise the *importance of the various indicators*. For example, are declines in the occupancy rate and the staff ratio and an increase in the overhead ratio good news or bad news? It is also unclear whether or not all the indicators are relevant. Finally, there is no clear indication as to why there has been an improvement in the budget outturn. A move away from the red could be the result of more government funding, for example.

In fact, the only indicator that can be used for measuring changes in efficiency (provided the conditions set out in section 8.5 are met) is the cost-price per prisoner per day. The remaining data are all potential explanations for differences in efficiency. In other words, the researchers jumped far too quickly to their *conclusions* about the explanatory factors. There is no real evidence to suggest they were justified in claiming that the granting of agency status and capacity increases were the main reasons for the changes in the level of efficiency. These should be seen rather as hypotheses that need to be tested.

### **8.6.2 FDH, DEA and SFA**

#### *General description of the techniques*

The three above analytical techniques, i.e. free disposable hull (FDH), data envelopment analysis (DEA) and stochastic frontier analysis (SFA), are all means of calculating efficiency by making a comparison with the best ratio of inputs to outputs in practice. All three techniques can be used to calculate the relationship between the inputs used and the outputs delivered by any organisation, to decide which organisations are capable of comparison and which individual organisation, within each group of comparable organisations, can be classified as 'top of the class'. The latter is taken as meaning the organisation<sup>76</sup> which uses the least inputs to produce a given quantity of outputs or which produces the most outputs with a given quantity of inputs. Finally, efficiency is measured by comparing the performance of all organisations that are not top of the class with that of the top of the class. In other words, the best-performing organisations are taken as the reference points, i.e. as

---

<sup>76</sup> In order to improve the readability of this manual, we refer only to 'organisations'. An audit may, however, involve examining the efficiency of a single organisation at different points in time, or comparing different organisations at different points in time.



constituting the best practice. They deliver the best output-input ratio that can be achieved in practice. These are efficient organisations.

The relative degree of efficiency is then measured by calculating the distance between each organisation and the given reference point (i.e. how does a particular organisation score in comparison with the most efficient comparable organisation?). The degree of efficiency is stated in the form of an *efficiency score*. Efficient organisations have an efficiency score of 100% (or 1), whilst inefficient organisations score lower than 100% (or 1). The lower the score, the less efficient is the organisation in question and the more opportunities there are (in theory) for it to improve its efficiency.

#### 8.6.2.1 *Free disposable hull*

The free disposable hull (FDH) technique is used to calculate the efficiency scores of a number of organisations by comparing their performance with the best practice. The model only makes comparisons with *actual* organisations. In assessing an organisation's efficiency, the idea is to try and find existing organisations that deliver the same outputs with fewer inputs.<sup>77</sup> FDH differs in this respect from DEA and SFA, as the latter two techniques allow for comparisons to be made with virtual organisations (see sections 8.6.2.2 and 8.6.2.3).

FDH is fairly easy to use. There are a number of standard, user-friendly FDH software packages on the market.<sup>78</sup>

There now follows an example of the use of FDH analysis. The Netherlands Court of Audit used the model when auditing the efficiency of job centres (now known as Centres for Work and Income) in the period between 1998 and 2000.<sup>79</sup>

#### *Job centres*

The Netherlands Court of Audit calculated, for each of the 62 job centres included in the audit, by how much the inputs could be reduced in delivering at least the same quantity of outputs. The inputs are the

---

<sup>77</sup> Although this is an input orientation, it is of course possible to adopt an output orientation.

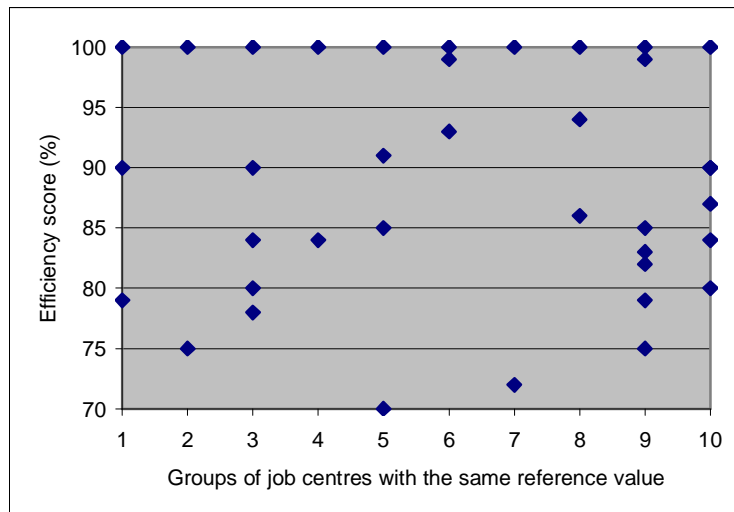
<sup>78</sup> Although the Netherlands Court of Audit does not have any FDH software at the time of writing (i.e. early 2005), you can always arrange for a licence to be bought in good time if you are planning to carry out an audit for which you wish to perform an FDH analysis.

<sup>79</sup> Netherlands Court of Audit (2001). See also Israël, Matheeuwsen, Roelofs and Roijackers (2001).



variable costs. Fixed costs, such as rent, depreciation and interest charges relating to buildings, were not included in the audit.<sup>80</sup>

The outputs delivered by the job centres are the jobs found for jobseekers (i.e. the number of jobseekers signing contracts of employment), broken down according to the relative employability of the jobseeker in question (in terms of category 1, category 2 and category 3). Category 1 jobseekers are defined as being the most employable, i.e. those most likely to find a job, etc. This breakdown according to the employability of jobseekers is a useful means of taking account of differences in the quality of the outputs (see section 8.5).



Efficient job centres are awarded an efficiency score of 100%, whereas inefficient job centres are awarded a lower score. An efficiency score of 90% means that the job centre in question can reduce its costs by 10% and still find jobs for the same number of jobseekers, broken down according to their relative employability. The above figure shows the findings for 1998.

Of the 62 job centres, 32 were awarded a 100% efficiency score, implying that there were no better comparable job centres. Ten of these efficient

<sup>80</sup> These costs are difficult to influence in the short term; including them in the audit might have led the auditors to draw the wrong conclusions. If the only difference between two job centres was that the housing costs incurred by one of them were twice as high as the housing costs incurred by the other, including housing costs in the calculation of an efficiency score would lead to the conclusion that one job centre was less efficient than the other. However, there is no need to perform an efficiency audit in order to draw this type of conclusion. Moreover, most job centres do not have much choice as regards their location as they are under a statutory obligation to serve a given catchment area.



job centres acted as reference points for inefficient job centres. When compared with comparable job centres, these ten delivered at least as many comparable outputs at the lowest level of cost. They are the ones with a 100% efficiency score in the figure. The efficiency score was calculated by comparing the inefficient job centres with their reference points. This resulted in the formation of ten groups of comparable job centres (i.e. the ten columns in the figure). The scores for the inefficient job centres ranged from 70% to 99%.

The figure shows that there are very wide differences between the efficiency scores of comparable job centres. By way of illustration, the following table shows the differences in costs among a group of job centres (group 5 in the figure) in 1998. The job entries figure is the number of placements, whilst the costs are quoted in millions of euros.

Job centre	Job entries for category 1 jobseekers	Job entries for category 2 jobseekers	# placements fase 3 Job entries for category 3 jobseekers	Costs	Efficiency score
Reference value	1.125	407	306	2,4	100%
Job centre 1	924	267	302	2,6	91%
Job centre 2	778	404	306	4,0	85%
Job centre 3	700	344	265	5,9	70%

Source: Netherlands Court of Audit

The table shows, among other things, that the reference value has recorded the largest number of job entries in all categories of jobseekers, at the lowest level of cost. Job centre 1, for example, can raise its outputs and at the same time lower its costs to the level of the reference job centre.

#### 8.6.2.2 Data envelopment analysis

Data envelopment analysis (DEA) is used to draw an efficiency boundary representing the best practical combination of outputs and inputs. The following figure illustrates this with a simple example.

The points represent the various organisations based on, in this particular case, the single input they use to deliver a single output. The line is the boundary as calculated with the aid of DEA. The figure takes an output



orientation (i.e. delivering the maximum quantity of outputs with a given quantity of inputs).<sup>81</sup> Organisations A to D deliver the largest quantity of outputs compared with the quantity of inputs required to produce them. These actual, best-practice organisations are located on the boundary line. Located on the lines between these actual organisations are virtual organisations which perform as well as the actual organisations.

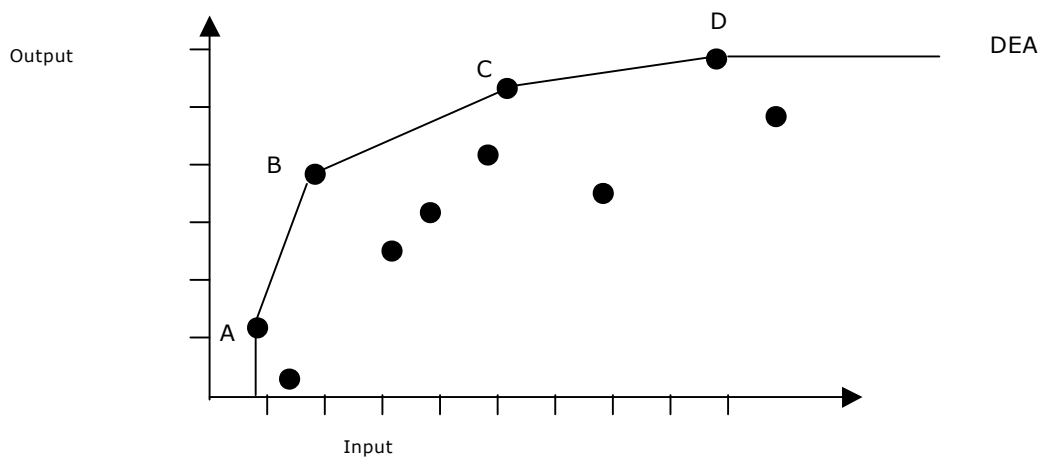


Figure 8.1 DEA: a simple example presented in graph form

In other words, DEA allows you to compare an organisation with both actual and virtual organisations. See the following section for further information on this point.

DEA is a relatively straightforward mathematical model for which various software packages are available that are capable of measuring efficiency in a relatively straightforward manner,<sup>82</sup> e.g. DEAP and DEA-solver.<sup>83</sup> DEA is a popular model, and thousands of articles and reports have already been published based on material produced by means of DEA analysis.<sup>84</sup>

An example of the use of DEA is presented below. The model was used in an audit performed by the National Institute of Public Health and the

<sup>81</sup> On a variable scale. This means that the outputs may increase either less or more than proportionately if there is a change in all the inputs (see section 8.6.2.4 for further information).

<sup>82</sup> See also, for example, Blank (2003) and Eggink and Blank (2001).

<sup>83</sup> See Coelli (1996) for information on DEAP, and Cooper (2000) for information on DEA-solver. See Blank (2003) for information on other software packages.

<sup>84</sup> Blank (2003).





Environment in 2001, the subject of which was the efficiency and quality of ambulance services in the Netherlands.<sup>85</sup>

#### *Ambulance services*

The National Institute of Public Health and the Environment used DEA to assess the efficiency of ambulance services in the Netherlands. The cost figure quoted is the figure for aggregate operating costs. The auditors assumed that the services could be broken down into four different outputs:

- the number of journeys made in which an ambulance is required to reach the patient as quickly as possible (known as category A1 journeys);
- the number of journeys made in which an ambulance is required to leave immediately, with the aim of reaching the patient as quickly as possible (category A2 journeys);
- the number of preplanned journeys (category B journeys);
- availability: the average of the availability indices for each ambulance depot, weighted according to the type of journeys made.

The DEA analysis covered 73 out of the total of 80 ambulance services in the Netherlands (i.e. 91%). The average efficiency score allotted to these 73 ambulance services was 87%.<sup>86</sup> The following table shows the scores awarded to the various services, broken down according to their size. Twenty ambulance services were allotted a 100% efficiency score. These are the reference points against which the performances of the other services are compared. The lowest efficiency score was 57%. The table also shows that average efficiency scores tend to rise in accordance with the size of the ambulance service.

---

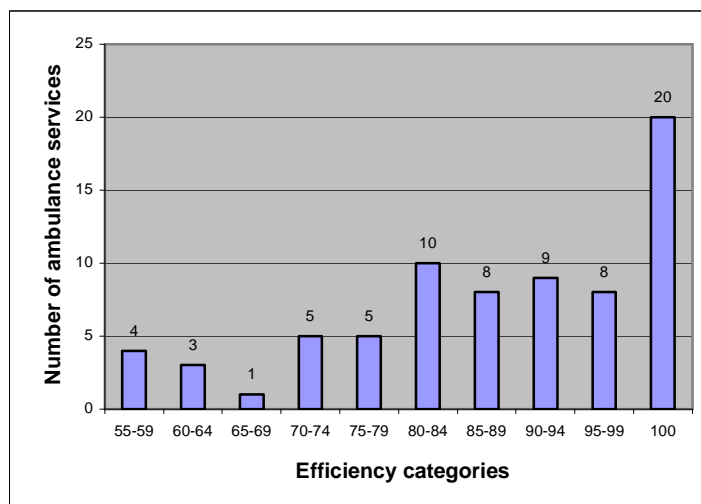
<sup>85</sup> Van der Veen, Zwakhals, Hazelnet-Crans and Van Manen (2001). A similar example is given in: Finansministeriet (2001). *Budgetredøgørelse 2000/2001* (Budget Review 2000/2001). Albertslund, Schultz Information. This describes an audit commissioned by the Danish government of the efficiency of district courts, in which the auditors made use of DEA.

<sup>86</sup> Obviously, this figure does not take account of factors that can explain the degree of efficiency, such as environmental and institutional factors.



Size category	Number of services	Average efficiency score	Number of services scoring 100
1: < 2.000	11	81	4
2: 2.000 – 5.000	16	87	4
3: 5.000 – 10.000	19	83	2
4: 10.000 – 20.000	17	90	5
5: > 20.000	10	95	5
<b>Total</b>	<b>73</b>	<b>87</b>	<b>20</b>

The following figure shows the scores awarded to the services, broken down into categories. The figure shows that the number of services in the four categories from 80-84 to 95-99 is more or less the same. Four services, or 5% of the total, scored less than 60%. These are very low (i.e. unusually low) scores for this type of audit.



### 8.6.2.3 Stochastic frontier analysis

Just like DEA, stochastic frontier analysis (SFA) involves drawing an efficiency boundary, thereby allowing the subject organisation to be compared with both actual and virtual organisations. The difference between DEA and SFA lies in the way in which the efficiency boundary is computed. In the case of SFA, the boundary is plotted with the aid of a mathematical function. This requires certain prior assumptions to be made about the relationship between inputs and outputs (i.e. it is a mathematical model). In many cases, though, the precise nature of this relationship is not clear. For this reason, it is worth starting by trying out

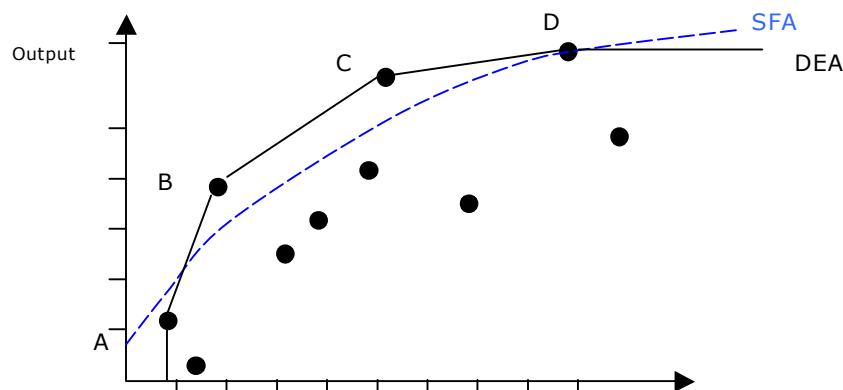


a number of simple regressions, as these can give you a sense of the potential nature of this relationship.

A number of models for the relationships between outputs and inputs are cited in the literature. The main ones to which we should like to draw your attention at this point are the production function (based on an output orientation) and the cost function (based on an input orientation).<sup>87</sup>

SFA is not easy to use, as it requires a knowledge of econometric methods, which are used for calculating (i.e. estimating) the mathematical function.<sup>88</sup> Standard software packages are also available for SFA, Frontier Version 4.1 being an example.<sup>89</sup>

Figure 8.2 uses a simple example to illustrate DEA and SFA. The figure reproduces the boundary line (plotted with the aid of DEA) in Figure 8.1, supplemented with a boundary line plotted with the aid of SFA.



**Figure 8.2 DEA and SFA: a simple example presented in graph form**

In relative terms, organisations A to D deliver the largest quantity of outputs compared with the amount of inputs used to produce them. This is why the DEA boundary line runs through these points. DEA involves solving a linear programming problem for each organisation, after which all organisations are plotted either on or under the boundary line. If they are on the line, this means that they deliver the largest quantity of

<sup>87</sup> Other relationships include output-distance functions and input-distance functions. A production function involves looking at the quantity of inputs, among other factors, as an explanation for the quantity of outputs. In the case of a cost function, the quantity of outputs delivered is taken as being one of the explanatory factors for the costs. See, for example, Coelli et al. (1998) and Blank (2000).

<sup>88</sup> See also, for example, Eggink and Blank (2001).

<sup>89</sup> Coelli (1996) and New South Wales Treasury (2001).



outputs compared with the other organisations in the population. If they are under the line, this means that they deliver fewer outputs using the same quantity of inputs. Any position away from the boundary line is assumed to be the result of inefficiency. In other words, the model does not take account of the influence of statistical distortions, such as measuring errors, inaccuracies in the underlying data and other shortcomings.

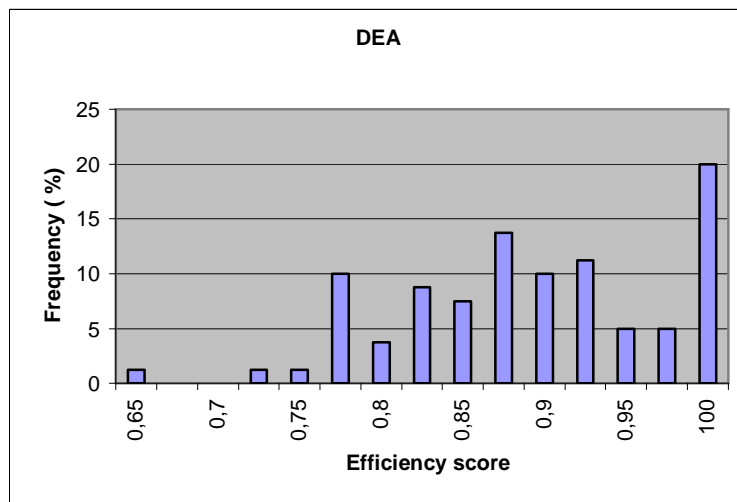
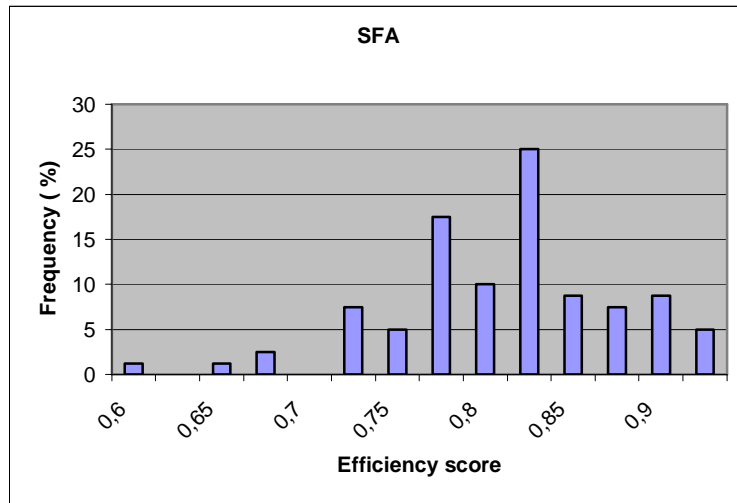
Boundary lines drawn with the aid of SFA do take account of this aspect, on the other hand. This means that you may end up plotting positions beyond the boundary line. SFA allows you to distinguish between inefficiency and statistical noise.

We should now like to give an example of the use of SFA. The model was used in a study published by the Social and Cultural Planning Office in 2001, the subject of which was institutional care for the mentally disabled.

#### *Institutional care for the mentally disabled*

The Social and Cultural Planning Office used SFA (the cost function) to assess the efficiency of institutional care services for the mentally disabled. DEA was also used. This particular example is restricted solely to general rather than specialist institutions. Observations from over one hundred general institutions covering the period from 1984 to 1998 were taken as the basis for the analysis (which involved both a comparison over time and a comparison among organisations). The cost figure used was the figure for variable costs. The outputs used were the number of patient days, broken down according to the degree of the patient's disability, i.e. minor, major and very severe. The general institutions recorded an average efficiency score of 86-89% in 1998. This means that, in theory, these institutions should be capable of cutting their costs by an average of over 10% and still delivering the same outputs.

The following figures show the percentage of institutions that were awarded the same efficiency score (on a scale from 0 to 1). The first figure shows the scores calculated with the aid of SFA, whilst the second figure shows the scores as computed with the aid of DEA. Both models point in similar directions, suggesting that firm conclusions may be drawn.



The general conclusion of the Social and Cultural Planning Office was that there was only a small gap between the efficient institutions for the mentally disabled and the remainder (i.e. the less efficient institutions). This means that the opportunities to free up resources for problem-solving purposes by changing policies or improving operating processes are fairly limited. What the institutions can do, however, is to raise their efficiency (at least marginally) by changing their input mix.

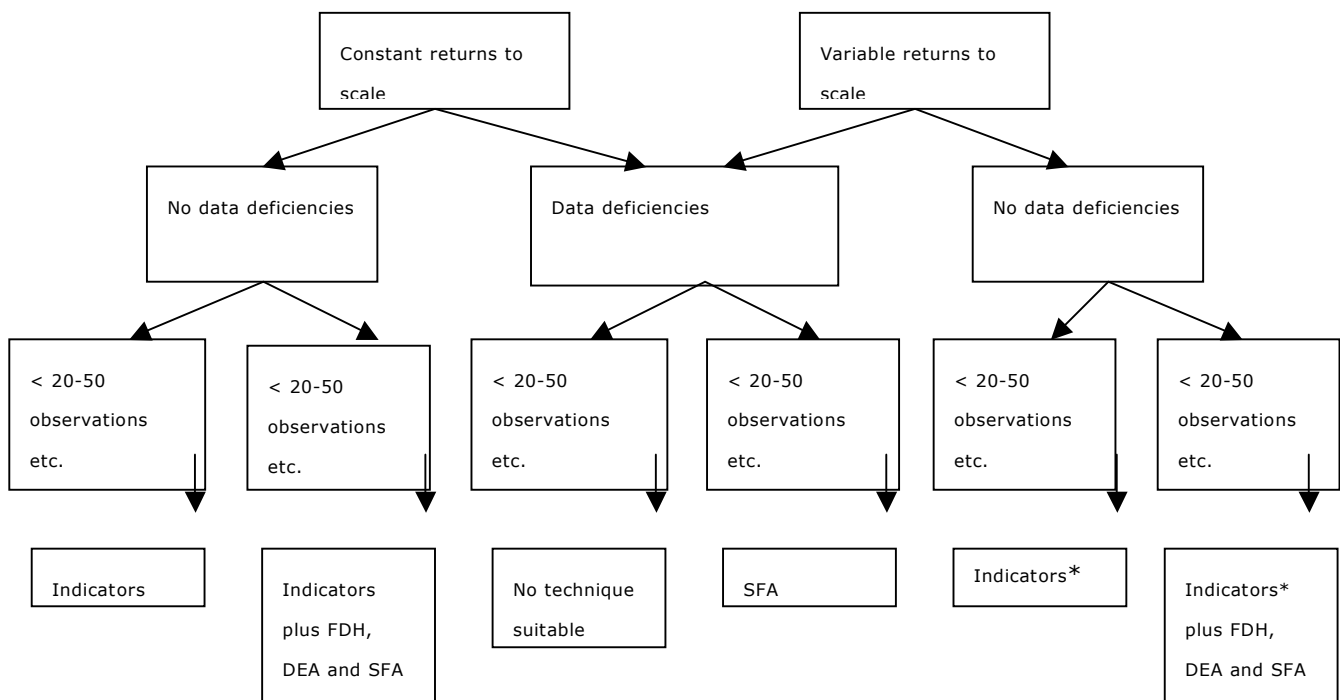
See Fried et al. (1993), Coelli (1998), Blank (2000), Cooper et al. (2000), Scheel (2000) and New South Wales Treasury (2001) for detailed descriptions of the various techniques.



8.6.2.4 Which technique should you use?

Various conditions need to be met in order to use both indicators and the FDH, DEA and SFA models. Certain conditions apply to all of these techniques, others only to some of them. Moreover, each technique has its own peculiar strengths and weaknesses.<sup>90</sup> Which technique is best suited for measuring the efficiency of a particular organisation in a given situation depends, *inter alia*, on the number of organisations and/or periods under review, and on the number of different outputs delivered by the organisations in question.

The following diagram is a *basic* flow chart suggesting which techniques should be used in which situation.



\* = additional data required

1. Constant or variable returns to scale

The first factor affecting the choice of technique for measuring efficiency is the question of whether the returns to scale are constant or variable.

<sup>90</sup> See Lovell (1998), Coelli (1998), Blank (2000) and Eggink and Blank (2001) for further information on this point.



In the case of constant returns to scale, any change in the inputs leads to a proportionate change in the outputs. For example, if the organisation in question delivers 10% more outputs (i.e. the scale is increased), this will result in a 10% increase in the level of cost. This means that the average cost per output is the same, irrespective of the size of the organisation.

In the case of variable returns to scale, on the other hand, a change in all inputs leads to a disproportionate increase or decrease in the quantity of outputs delivered. For example, if the organisation in question delivers 10% more outputs, this will result in the costs rising by either more or less than 10%. The average cost per output is usually highest among small and large organisations, while the lowest average cost is found among medium-sized organisations.<sup>91</sup>

With the exception of indicators, all the various techniques are able to take account of both constant and variable returns to scale. Efficiency indicators (such as the cost per output), on the other hand, implicitly assume constant returns to scale, which means that the cost per output is not affected by the size of the organisation in question. Despite this, there are situations in which indicators can be used for variable returns to scale. There is, however, an additional condition that needs to be fulfilled compared with constant returns to scale, which is that the analysis should include a set of indicators for different organisational sizes (e.g. small, medium-sized and large). See section 8.6.3.2 for an example.

## 2. *Data deficiencies*

A second factor affecting your choice of technique is whether there are any deficiencies in the data you intend to use for calculating the efficiency of the organisation in question. You can find this out by performing an analysis *beforehand*, either one using indicators or a simple regression analysis. The point is that, with the exception of the SFA technique, all models are susceptible to data inadequacies. As we have already mentioned, SFA is capable of distinguishing between differences in efficiency and differences triggered by statistical distortions, i.e. problems relating to the quality of data.

---

<sup>91</sup> See Annexe 7 for further information plus an example.

**Home care**

PricewaterhouseCoopers and Berenschot performed a DEA analysis in 1999 in order to assess the efficiency of home-care services in the Netherlands. Although the report states that not all the data used were equally reliable, the authors did not make use of SFA analysis, despite the fact that it is more suited to a situation in which there are doubts about the quality of the data.

### 3. *Number of observations*

A third factor affecting your decision is whether you have access to the necessary data on the outputs delivered and the inputs used to this end, *in relation to at least 20-50 organisations or points in time.*<sup>92</sup> Generally speaking, you will not need as many observations in order to use indicators as you need in order to use FDH, DEA and SFA. And you will not generally need as many observations in order to use FDH and DEA as you will need in order to use SFA. Roughly speaking, you can use indicators if you have fewer than 20-50 observations, whilst you can use FDH and DEA if you have access to around 20-50 observations. In order to use SFA, you will generally need to have at least 50 observations.

Although you can use FDH for comparing a relatively small number of organisations, you will need to divide them into a number of categories that are comparable in size. As we have already said, FDH allows you only to make comparisons with actual (rather than virtual) organisations. If no comparable organisations are available, in terms of the quantity of outputs delivered or inputs used, the model will tend to classify your subject organisation(s) as 'efficient by default'.

One of the advantages of DEA and SFA over FDH is that there is less likelihood of your subject organisation(s) being classified as efficient by default (as you can also make comparisons with virtual organisations). The problem with FDH is that it has a tendency to regard a relatively large number of organisations as being efficient, which means that it is not as good at providing useful information on efficiency differences, i.e. it is less able to differentiate.

---

<sup>92</sup> The number of 20-50 is intended only as a rough indication. Depending on the number of observations relative to the number of variables (such as types of output and input), you will need to decide whether the number of observations is sufficiently large to enable you to use the techniques in question. The phrase 'sufficiently large' is used here to mean large enough to allow you to make statistically reliable pronouncements.





### *Preference*

You should proceed with care in all situations that lend themselves to the use of indicators. This is because, for example, different indicators can easily produce conflicting signals, which makes it difficult to make any pronouncements about the level of efficiency as a whole. Indicators are not very good at explaining efficiency differences. See section 8.6.1 for further information on the use of indicators.

In the two situations in which, broadly speaking, all the various techniques can be used, one or more of them are often more suitable than the others, depending on their own individual strengths and weaknesses. We have already discussed a number of these strengths and weaknesses. The following are also important points in this respect:

- Indicators are often less suitable when you are dealing with a range of outputs and/or inputs. Unlike indicators, FDH, DEA and SFA allow you simultaneously to relate a range of outputs to the inputs used and hence to assess the way in which they affect each other. A production function in SFA is less suitable for dealing with a range of outputs, as it is relatively difficult to differentiate between more than one output.
- As we have already mentioned, the use of a mathematical function in SFA requires you to make a number of prior assumptions about the relationship between inputs and outputs, even though you often do not know much about its precise nature. If you are not very confident about the function you have chosen, you are advised to use indicators, and FDH and DEA rather than SFA (provided that these are suitable techniques, of course).<sup>93</sup>

The following table gives a rough indication of the main strengths and weaknesses of the various techniques.

	Indicators	FDH	DEA	SFA
Capable of taking account of both constant and variable returns to scale	-	+	+	+
Capable of assessing the relationship between a range of outputs and inputs	-	++	++	+
Less sensitive to data deficiencies	-	-	-	+
Can be used with a relatively small number of observations	++	+	+	-
Easy to use	++	+	+	-
Provides information on efficiency differences	--	-	-+	+
Methodologically sound	-	+	+	++

<sup>93</sup> In the case of SFA, the efficiency score is also affected by the type of function you choose.



#### 8.6.2.5 *Using more than one technique at the same time*

Each of the various techniques that can be used for measuring efficiency has its own strengths and weaknesses. You can test the robustness of your findings by using a number of methods alongside each other (i.e. indicators,<sup>94</sup> FDH, DEA and SFA). Try and use at least two at the same time.

If all the models come up with the same findings, the pronouncements you can make are more reliable than would be the case if the findings were contradictory.

A high degree of robustness is required in order to make recommendations.<sup>95</sup> This is why it is worth (if possible) always using a number of different methods in order to calculate the degree of efficiency and compare the results.

### 8.6.3 **Accounting for differences in efficiency**

If you can identify certain *explanatory factors* that account for differences in efficiency, you can make specific recommendations to the subject organisation on ways of improving its efficiency. In practice, this means using both statistical and non-statistical techniques.

#### 8.6.3.1 *Statistical techniques*

Once you have used indicators to measure the efficiency of certain outcomes or outputs, the second step is to examine the possible explanations for any discrepancies. Among the methods available for this purpose are tobit and probit analysis, as well as simple regression analysis,<sup>96</sup> provided of course that you have a sufficiently large number of observations. SFA, DEA and FDH are also suitable for this purpose, although FDH only to a limited degree. If you use SFA, DEA or FDH to measure efficiency, you can use the same techniques to identify explanations for any efficiency differences.

We should now like to give two practical examples. These are both examples of audits in which the auditors sought to measure the efficiency

---

<sup>94</sup> Incidentally, you can also use a simple regression analysis to measure efficiency.

<sup>95</sup> See, for example, Israël, Matheeuwsen, Roelofs and Roijackers (2001) and Eggink and Blank (2001).

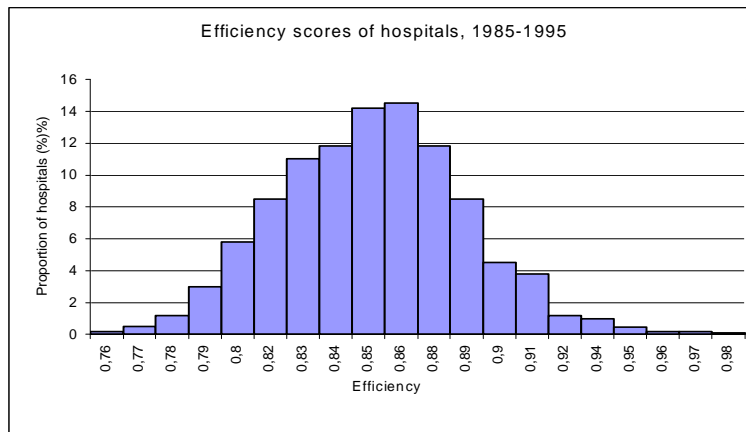
<sup>96</sup> For further information, and a discussion of the theory underlying the use of these techniques, see, for example, Stevens (1992) or Tacq (1991).



of the outputs delivered by a particular organisation, and to use statistical techniques (i.e. SFA and DEA in the first example, and probit analysis in the second example) to pinpoint explanations for efficiency differences. The first example relates to hospitals, the second to nursing homes.

### Hospitals

The Social and Cultural Planning Office measured the efficiency of Dutch hospitals in the period from 1985 to 1995 (Blank, 1998). SFA (the cost function) was used to calculate an efficiency score for each hospital on a scale from 0 to 1. The findings are shown in the following figure.



The cost function was used to examine the potential explanations for differences in efficiency. This analysis showed, *inter alia*, that efficiency:

- declines by 0.3% if the composite price of nursing staff rises by 1% (the composite price is the ratio between the price paid by a hospital and the average price; the composite price is high if the staff employed by the hospital are relatively highly qualified and experienced);
- declines by 0.17% and 0.13% respectively if the composite price of other staff and paramedical staff rises by 1%;
- declines by 0.1% if the proportion of paramedical staff rises by 1%;
- declines by 0.04% if the rate of absenteeism due to illness rises by 1%.

Other explanatory factors included in the analysis, such as the bed occupancy rate and the proportion of consultants employed by the hospital being greater than 10%, proved not to have any material effect on a hospital's efficiency.



DEA was also used to calculate the efficiency scores and to identify any factors that had a material effect on a hospital's efficiency. Broadly speaking, it generated the same conclusions.

#### *Nursing homes*

In 1996, the Social and Cultural Planning Office published a study of the efficiency of Dutch nursing homes in the years 1984 and 1987-1993.<sup>97</sup> SFA (the cost function) was used to calculate an efficiency score for each nursing home on a scale from 0 to 100% for each of the years under review. The cost figure taken was that for the variable costs of staff and equipment. The study examined performance in relation to the following outputs:

- number of patient days;
- number of day-treatment days;
- care intensity (i.e. care needed by patients and standard of service);
- extramural activities.

The most efficient nursing homes were awarded a score of 100%. The average efficiency score posted by all nursing homes during the period under review was 93%. This means that the nursing homes should be capable of reducing their costs by an average of 7%.

One of the study findings was that the bigger a nursing home is, the more efficient it tends to be. Probit analysis was used to establish which variables play a role in deciding whether or not a nursing home is efficient. These explanatory factors were divided into three categories, i.e. product features, market features and type of institution, and operational and process features. The results of the probit analysis are presented in the following table. A plus sign indicates that the variable in question tends to raise the nursing home's efficiency, whilst a minus sign indicates that it tends to lower the efficiency. A zero means that it is not possible to make a pronouncement about the effect of the variable in question.

	<i>Estimate</i>
<i>Production</i>	
- Proportion of trainee nurses	+
- Quality of qualified nurses	-
- Quality of other staff	-
- Capital per bed	0

<sup>97</sup> See Blank et al. (1996) and Institute for Research into Public Spending (1998).



- In-house pharmacy	+
- Needs index	0
- <i>Market features and type of institution</i>	
- Level of competition	+
- General non-denominational	0
- Protestant	-
- Foundation / trust	0
- Somatic	0
- Psychogeriatric	0
- <i>Operational and process features</i>	
- Occupancy rate	+
- Absenteeism due to illness	-
- Use of outsourcing	+
- Use of insourcing	+
- Size of department	-

The researchers concluded that there were only limited opportunities for raising efficiency (i.e. by reducing the level of cost). Some of the inefficiencies identified were due to external (or exogenous) factors such as market conditions, which are either difficult or impossible to influence or control. The researchers also concluded that, whilst a form of upsizing could lead to marginal efficiency gains for small homes, it would in fact lead to efficiency losses for large nursing homes.

The study also showed that, the more highly qualified the staff, the greater the degree of inefficiency. However, the need to safeguard the standard of service coupled with social considerations would appear to preclude the possibility of pursuing a reduction in quality standards as a realistic policy. Other conclusions were that factors such as outsourcing, insourcing, higher occupancy rates and smaller department sizes tend to raise a nursing home's efficiency.

DEA was also used to calculate the efficiency scores, and led broadly to the same conclusions.

#### *Conclusion*

The above examples demonstrate the value of statistical techniques. These can be used to specify the precise factors that affect the efficiency of the subject organisation. In other words, they can supply you with a set of explanatory factors backed by statistical evidence, which you can then use to improve the organisation's efficiency.



At the same time, the examples also show that it is not possible either to account for all kinds of inefficiencies or for the organisations in question to influence all possible explanatory factors. This is in itself a crucial consideration: if the analysis shows that most inefficiencies are caused by external factors rather than by shortcomings in the subject organisation itself, this is a good reason for sticking to the same policy.

#### 8.6.3.2 *Non-statistical techniques*

If the information available to you precludes the use of a statistical technique in order to identify factors that can account for differences in efficiency, you can often use graphs instead, consult experts or make comparisons between organisations. Let's take a closer look at these three options.

##### *Graphs*

One way of finding explanations is by drawing a graph or scatter diagram in which every single factor that could possibly account for any efficiency differences is plotted against the relevant efficient indicator, such as the cost per output or the cost per outcome. However, in order to draw conclusions on the basis of such figures, you need to have access to a sufficiently large number of observations, so as to ensure that your conclusion is not simply a question of chance.

If you produce indicators for comparing a large number of organisations, you should make sure that, where possible and relevant, the possible explanatory factors include the *size of the organisation*. This factor can also affect efficiency, which is why you need to take it into account. For example, small, medium-sized and large organisations may all have different cost structures (large organisations tend to be more bureaucratic, resulting in a higher average cost; see also Annexe 8).

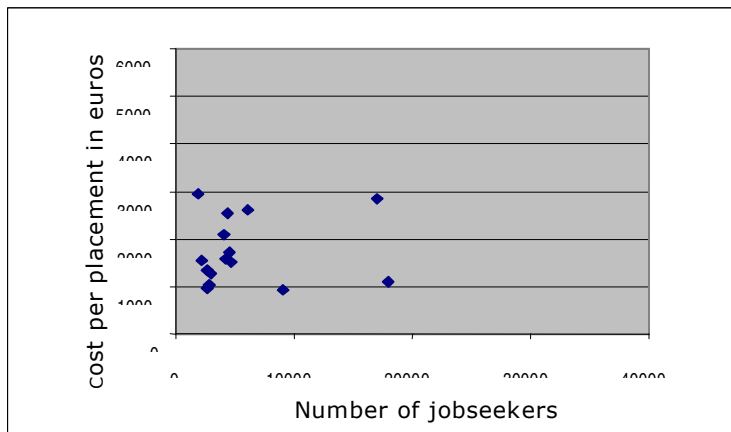
The following example of the usefulness of graphs is taken from the Court's audit of job centres (now known as Centres for Work and Income) to which we have already referred. Generally speaking, those jobseekers who are not easily employable, for whatever reason, do not stand a chance of finding a job until they have attended some form of training course. The expectation, therefore, is that the cost per jobseeker who finds a job tends to rise in inverse proportion to the relative employability of the jobseekers on a job centre's books (all other things remaining equal). This is because the job centres need to do more and hence incur



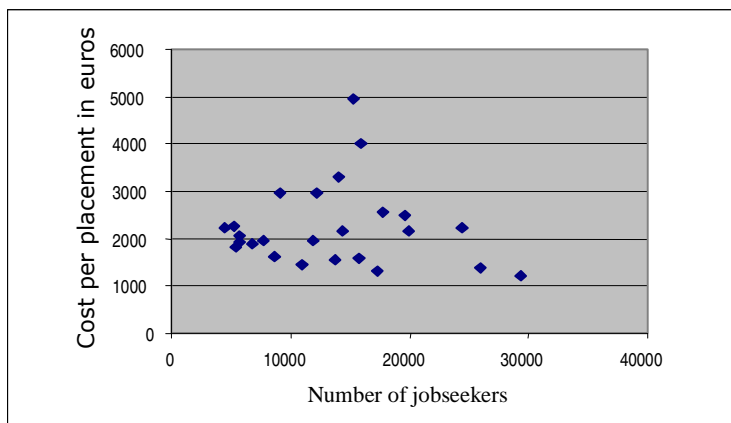
more costs in order to raise the employability of the jobseekers on their books.

The following graphs show the cost per placement in 1998 in relation to those jobseekers for whom the job centres managed to find jobs, broken down according to the relative employability of the jobseekers on the centres' books. The employability categories are based on the composition of the portfolio of jobseekers at each job centre. In order to take account of the effects of size differences, i.e. large, medium-sized and small job centres may conceivably have different cost structures, the costs have been plotted against the size of the job centre, as measured in terms of the number of jobseekers (see Annexe 8 for further information).

A: Employability category 1 (most employable jobseekers)

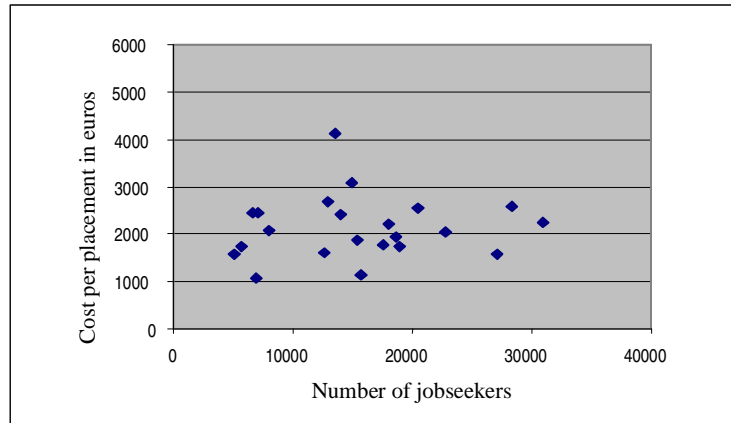


B: Employability category 2 (jobseekers of average employability)





C: Employability category 3 (least employable jobseekers)



The graphs do not point to a clear correlation between the cost per placement and the relative employability of the jobseekers on a centre's books. Depending on the number and relative employability of the jobseekers on a centre's books, the cost per placement is roughly between €1,000 and €3,000.

Obviously, account must also be taken of the influence of the job market. Those job centres that operate in a region where there is a relatively high demand for labour should find it easier than other job centres to find jobs for the jobseekers on their books. However, even if account is taken of labour market conditions, there are still fairly wide gaps in the cost per placement. In other words, there is no clear link between the various explanatory factors and the cost per placement. The findings do not suggest any obvious way of improving the efficiency of job centres.

*Expert opinions*

Another way of finding potential explanations for differences in efficiency is by asking experts to look at the findings (i.e. on the efficiency scores; the cost per outcome or output) and to suggest possible explanations for them. Although their opinions are no more than educated guesses, they could help to identify potential opportunities for improvement. The subject organisation(s) could then try out these suggestions in practice, using a method of trial and error to see whether they actually lead to any improvement.

There are, however, a number of risks associated with such an approach. Given that there is no guarantee that the opinions of experts are genuine explanations for any inefficiencies, and that different factors often affect





each other and it is difficult to know in advance how they are going to interact in practice, any action taken on the basis of expert opinions may well not have any effect at all on the level of efficiency and might even exacerbate the situation.

#### Risks of expert opinions (fictitious example)

Let's say that you consult an expert in connection with an audit you are performing of the efficiency of hospitals. The expert tells you that the cost per patient tends to rise in proportion to the percentage of paramedical staff employed by a hospital. But what if the proportion of paramedical staff affects the composite price? That is, the lower the proportion of paramedical staff, the higher the composite price. Let's say that your expert is not aware of this. The fact is that the higher the composite price, the higher the cost per patient. If the hospital now tries to reduce the proportion of paramedical staff in order to raise its efficiency, there is a risk that the operation will fail as a lower proportion of paramedical staff will result in a higher composite price. This, in turn, will have an adverse effect on efficiency.

#### *Comparing organisations*

If you have compared a number of comparable organisations with each other in order to measure the efficiency of your subject organisation(s), and you find that there are wide differences in the level of efficiency, you will find it useful to take a closer look at the best- and worst-performing of these organisations. This could suggest a number of useful ways in which the worst-performing organisation could adjust a number of its operating processes so as to model them on the processes used by the best-performing organisation.

For instance, the example of the job centres given in section 8.6.2.1 (see figure and table) suggests that the inefficient job centres can raise their efficiency to the level achieved by their reference job centre (i.e. the best-performing comparable job centre). In other words, the inefficient job centres could take a good look round, and learn some useful lessons from the reference job centre located just above them in their own column in the figure. Obviously, this type of approach is best suited to a comparison with actual (rather than virtual) organisations.

There are the same risks with this type of approach as there are with expert opinions.



### *How not to do it*

An example of a DEA-based study in which conclusions were drawn far too readily on the basis of the research findings is the efficiency study of home-care organisations carried out by PricewaterhouseCoopers and Berenschot in 1999 (to which we have already referred).

#### Exceptional Medical Expenses Act

The Dutch government decided in the 1990s to carry out benchmarking studies for all organisations covered by the Exceptional Medical Expenses Act. PricewaterhouseCoopers and Berenschot performed a DEA analysis in 1999 of organisations offering home-care services. The government concluded, on the basis of the audit findings, that this particular tool can be used for analysing and comparing organisational performance, and also for identifying similar characteristics of efficient organisations.

De Groot (2000) shows that the study hardly merits these conclusions. The researchers did not manage to find sufficient explanations to account for the differences in efficiency, with the exception of a relatively trivial link between a high rate of absenteeism due to illness and a low level of efficiency. Moreover, no particular aspects of the operational management pursued by those organisations labelled as 'best-practice organisations' obviously serve to distinguish them from the others. This means that the study cannot be used to justify any recommendations for raising the efficiency of the organisations concerned to the level of the best-performing comparable organisation, which is, after all, the ultimate aim of this type of comparative study.

As is the case with every statistical technique, both DEA and SFA need to be used with care. Moreover, in order for the findings to be useful in practice, it is important to formulate potential explanations and to examine these to see whether they do indeed account for any differences in efficiency revealed by your analysis (see section 8.6.3). In other words, you need to formulate your explanations for efficiency differences very carefully and make sure they are backed by clear evidence. If you are unsure about them, you should put them to the test.



## Annexe 1 Literature

Aarts, L.J.M., I. Been and R. Goudriaan (2002).

Productiviteitstaakstellingen in de nationale begrotingen: een internationale vergelijking van de doorwerking in de begroting van de arbeidsproductiviteitsontwikkeling. Den Haag: APE bv.

Baarda, D.B. & M.P.M. de Goede (1996). *Basisboek methoden en technieken*. Houten: Stenfert Kroese.

Baarda, D.B., M.P.M. de Goede & M. Kalmijn (2000). *Basisboek enquêteren en gestructureerd interviewen*. Houten: Stenfert Kroese.

Blank J.L.T. (1998). Benchmarken of de kunst van het vergelijken. In: *Economisch Statistische Berichten*, 29-5-1998, 2.

Blank, J.L.T. (ed.) (2000). *Public provision and performance; contributions from efficiency and productivity measurement*. Amsterdam: Elsevier.

Blank, J.L.T (2000). Doelmatigheid van de publieke sector: enige beschouwingen. In: *Openbare Uitgaven*, 32, no. 6, pp. 260-271.

Blank, J.L.T. (2003). *Data Envelopment Analysis: een goede praktijk*. <http://www.benchmarkenindepubliekesector.nl>: Kenniskamer Benchmarken Publieke Sector, Ministry of the Interior and Kingdom Relations.

Blank, J.L.T. and E. Eggink, assisted by A.I. de Graaff (1996). *Zuinig op zorg; een empirisch onderzoek naar de productiestructuur van verpleeghuizen*. Rijswijk: Social and Cultural Planning Office.



Blank, J.L.T., E. Eggink, A.H.Q.M. Merkies (1998). *Tussen bed en budget*. Een empirisch onderzoek naar de doelmatigheid van algemene en academische ziekenhuizen in Nederland. The Hague: Social and Cultural Planning Office.

Bouckaert, G., W. Van Reeth, T. Auwers & K. Verhoest (1998). *Handboek doelmatigheidsanalyse prestaties begroten*. Brussels: Ministry of the Flemish-Speaking Community.

Bouckaert, G. & T. Auwers (1999). *Prestaties meten in de overheid*. Brugge: die Keure.

Bressers, J.T.A. (1982). *Beleidseffektiviteit en waterkwaliteitsbeleid; een bestuurskundig onderzoek* ('Policy effectiveness and water quality policy: a study of public administration'), dissertation submitted at Technical University of Twente.

Brinks, R. & T. Witteveen (ed.) (1999). *Toegang tot de rijksbegroting*. The Hague: Sdu.

Coelli T.J. (1996-1). *A guide to Frontier Version 4.1: A Computer Program for Frontier Production Function Estimation*. CEPA Working Paper 96/7. Armidale: Department of Econometrics, University of New England.

Coelli T.J. (1996-2). *A guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program*. CEPA Working Paper 96/8. Armidale: Department of Econometrics, University of New England.

Coelli T.J., D.S. Prasada Rao and G.E. Battese (1998). *An introduction to efficiency and productivity analysis*. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

Cooper W.W., L.M. Seiford and K. Tone (2000). *Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-solver software*. Dordrecht: Kluwer Academic Publishers.

Netherlands Court of Audit (1990). *Subsidies*. House of Representatives, 1989-1990, 21 080, nos. 1-2.



Netherlands Court of Audit (1991a). *Toetsbaarheid* kinderopvangbeleid. House of Representatives, 1991-1992, 22 233, nos. 1-2.

Netherlands Court of Audit (1991b). *Voorlichtingscampagnes van het Rijk*. House of Representatives, 1990-1991, 22 152, nos. 1-2, p. 5.

Netherlands Court of Audit (1991c). *Verslag van de Algemene Rekenkamer over 1990*. House of Representatives, 1990-1991, 22 032, nos. 1-2.

Netherlands Court of Audit (1995a). *Fondsbeheer in de sociale zekerheid*. House of Representatives, 1994-1995, 24 065, nos. 1-2.

Netherlands Court of Audit (1995b). *Ontslagitkeringen bij universiteiten, academische ziekenhuizen en onderzoeksinstituten*. House of Representatives, 1994-1995, 24 224, nos. 1-2.

Netherlands Court of Audit (1996a). *Gesubsidieerde arbeid*. House of Representatives, 1996-1997, 25 070, nos. 1-2.

Netherlands Court of Audit (1996b). *Vergunningen*. House of Representatives, 1995-1996, 24 656, nos. 1-2.

Netherlands Court of Audit (1996c). *Verhaalsrecht in de ziekenfondswet*. House of Representatives, 1996-1997, 25 025, nos. 1-2.

Netherlands Court of Audit (1997a). *Afvalpreventie*. House of Representatives, 1996-1997, 25 375, nos. 1-2.

Netherlands Court of Audit (1997b). *Fonds voorheffing pensioenverzekering*. House of Representatives, 1996-1997, 25 401, nos. 1-2.

Netherlands Court of Audit (1997c). *Informatievoorziening en kengetallen*. House of Representatives, 1996-1997, 25 471, nos. 1-2.



Netherlands Court of Audit (1997d). *Ontwikkeling en herverdeling politiesterkte 1994-1995*. House of Representatives, 1996-1997, 25 385, nos. 1-2.

Netherlands Court of Audit (1998a). *Fiscale faciliteit voor het leerlingwezen*. House of Representatives, 1997-1998, 26 060.

Netherlands Court of Audit (1998b). *Geluidshinder Defensie*. House of Representatives, 1998-1999, 26 230, nos. 1-2.

Netherlands Court of Audit (1998c). *Groeicijfers Schiphol*. House of Representatives, 1998-1999, 26 265, nos. 1-2.

Netherlands Court of Audit (1998d). *Kwaliteit uitvoering Europese akkerbouwregeling*. House of Representatives, 1998-1999, 26 240, nos. 1-2.

Netherlands Court of Audit (1999a). *Collegeprogramma 1999-2004: positionering en prioriteiten*. The Hague: Court of Audit.

Netherlands Court of Audit (1999b). *Integratiebeleid etnische minderheden*. House of Representatives, 1998-1999, 26 426, nos. 1-2.

Netherlands Court of Audit (1999c). *Procesmanagement bij het Ministerie van Onderwijs, Cultuur en Wetenschappen* (letter to the Minister of Education, Culture and Science). The Hague: Court of Audit.

Netherlands Court of Audit (2000a). *Organisatie van de beleidsevaluatie*. House of Representatives, 1999-2000, 27 065, nos. 1-2.

Netherlands Court of Audit (2000b). *Museale collecties van het Rijk*. House of Representatives, 2000-2001, 27 470, nos. 1-2.

Netherlands Court of Audit (2001a). *Efficiëntie van arbeidsbureaus*. The Hague: Court of Audit.

Netherlands Court of Audit (2001b). *Begeleiding en herplaatsing van voortijdige schoolverlaters*. The Hague: Court of Audit.



Netherlands Court of Audit (2001c). *Specifieke gegevens regio Zuid/Oost Brabant*, annexe to the above-mentioned audit report entitled *Begeleiding en herplaatsing van voortijdige schoolverlaters* (copies of which were sent exclusively to local recipients). The Hague: Court of Audit.

Netherlands Court of Audit (2002a). *Bestrijding uitstoot broeikasgassen*. House of Representatives, 2001-2002, 28 272, nos. 1-2.

Netherlands Court of Audit (2002b). *Reactie van de Algemene Rekenkamer op het Referentiekader mededeling over de Bedrijfsvoering*. The Hague: Court of Audit.

Netherlands Court of Audit (2002c). *VBTB in begrotingen 2002*. The Hague: Court of Audit.

Netherlands Court of Audit (2002d). *Van beleidsbegroting tot beleidsverantwoording*. Lijst van vragen en antwoorden. House of Representatives, 2000-2001, 26 573, no. 75.

Netherlands Court of Audit (2002e). *Preventie en bestrijding jeugdcriminaliteit*. House of Representatives, 2001-2002, 29 282, no. 2.

Netherlands Court of Audit (2002f). *Toelating bestrijdingsmiddelen voor de landbouw*. House of Representatives, 2002-2003, 28 615, nos. 1-2.

Netherlands Court of Audit (2003a). *Handreiking meten van doelmatigheid*. The Hague: Court of Audit.

Netherlands Court of Audit (2003b). *Wonen, zorg en welzijn van ouderen*. House of Representatives, 2002-2003, 28951, nos. 1-2.

Netherlands Court of Audit (2004). *Handleiding Normen en begrippen*. The Hague: Netherlands Court of Audit (Part of HANDAR, Netherlands Court of Audit manual).

Eggink, E. and J.L.T. Blank (2001). *Verstandig verzorgd*. Een empirisch onderzoek naar de efficiëntie van de intramurale zorg voor



verstandelijk gehandicapten. The Hague: Social and Cultural Planning Office.

Finansministeriet (2001). *Budgetredegørelse 2000/2001 (Budget Review 2000/2001)*. Albertslund: Schultz Information.

Fried, H.O., C.A. Knox Lovell and S.S. Smidt (1993). *The measurement of productive efficiency: Techniques and applications*. New York/Oxford: Oxford University Press.

GAO (1990). *Case study evaluations*. <http://www.gao.gov>.

GAO (1991). *Designing evaluations*. <http://www.gao.gov>.

Goudriaan et al. (1989). *Doelmatig dienstverlening. Een onderzoek naar de produktiestructuur van vier voorzieningen in de kwartaire sector*. Rijswijk: Social and Cultural Planning Office.

Goudriaan, R., B. Jongbloed and D.C. van Ingen (1998). *Kostendeterminanten en doelmatigheid van het Nederlandse hoger onderwijs*. Zoetermeer: Ministry of Education, Culture and Science.

Groot, H. de (2000). *Tussen markt en politieke arena; over managementcontrol in non-profitorganisaties*. Amsterdam: Free University.

Groot, H. de & R. Goudriaan (1991). *De produktiviteit van de overheid. Over prestaties, personeel en uitgaven in de publieke sector*. Schoonhoven: Academic Service.

HM Treasury et al. (2001). *Choosing the right FABRIC. A Framework for Performance Information*. London (UK): HM Treasury, Cabinet Office, National Audit Office, Audit Commission, Office for National Statistics.

Hoogerwerf, A. (1980). *Overheidsbeleid*. Alphen aan den Rijn: Samson.

House of Representatives (2001). *Vaststelling van de wet inzake het beheer van de financiën van het Rijk (Comptabiliteitswet 2001)*. Explanatory Memorandum. 2000-2001, 27 849, no. 3, pp. 44-45.





- Institute for Research into Public Spending. *Openbare Uitgaven*, Theme: Efficiency in the public sector, no. 1, 1998.
- Israël, F.J., A.G.M. Matheeuwsen, M.V.A.M. Roelofs and F.V. Roijackers (2001). Werken arbeidsbureaus efficiënt? In: *Openbare uitgaven*, 33, no. 3, pp. 126-132.
- Jones, R, & M. Pendlebury (2000). *Public sector accounting*. Harlow (UK): Pearson Education.
- Knaap, P. van der (2000). Evaluatieonderzoek en prestatiegegevens bij het rijk: nieuwe kaders en perspectieven. *Beleidsanalyse*, 3 and 4, pp. 35-49.
- Kraan-Jetten, A. (1991). *Effectiviteit van overheidsbeleid*. Systematisering van inzichten, getoetst aan het beleid ter bevordering van het eigen-woningbezit. Beusichem: Centraal Service Bureau.
- Leeuw, F.L. (1992). *Productie en effectiviteit van overheidsbeleid*. Institutionele analyse en effectmeting. The Hague: Vuga.
- Lovell, C.A.K. (1998). Doelmatigheidsonderzoek in de publieke sector. In: *Openbare uitgaven*, 30, no. 1, pp. 6-13.
- Lulofs and Schuddeboom (1991). Het vaststellen van de mate van doelrealisatie. In: J.T.A. Bressers en A. Hoogerwerf (ed.), *Beleidsvaluatie*. Alphen aan de Rijn: Samsom H.D. Tjeenk Willink.
- Ministry of Finance (1999). *Van beleidsbegroting tot beleidsverantwoording*. The Hague: Ministry of Finance.
- Ministry of Finance (2001). *Regeling prestatiegegevens en evaluatieonderzoek rijksoverheid*. The Hague: Ministry of Finance.
- Netherlands Bureau for Economic Policy Analysis (2004). *Lerend beleid: het versterken van beleid door experimenteren en evalueren*. CPB document 48. The Hague: Netherlands Bureau for Economic Policy Analysis.



New South Wales Treasury (2001). *Guide to economic performance measurement for general government sector agencies*. New South Wales: New South Wales Treasury.

PriceWaterhouseCoopers/Berenschot (1999). *Benchmarkonderzoek Thuiszorg biedt aanknopingspunten voor instellingen en overheid*. PriceWaterhouseCoopers/Berenschot, Utrecht.

Rao, D.S. Prasada (1999). *Workshop on Productivity and efficiency analysis, 15-18 November 1999*. Groningen. University of Groningen, Systems, Organization and Management Research Institute.

Scheel, H. (2000). *Effizienzmasse der Data Envelopment Analysis*. Wiesbaden: Gabler.

Social and Cultural Planning Office (2001). *Maten voor gemeenten*. Een verkennende studie naar prestatie-indicatoren in het kader van het 'Plan van aanpak transparantie in de financiële verhouding'. Working document no. 74. The Hague: Social and Cultural Planning Office.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Mahwah (NJ): Lawrence Erlbaum Associates.

Swanborn, P.G. (1987). *Methoden van sociaal-wetenschappelijk onderzoek*. Meppel: Boom.

Swanborn, P.G. (1996). *Case-study's, wat, wanneer en hoe?*, Meppel: Boom.

Swanborn, P.G. (1999). *Evalueren*. Meppel: Boom.

Swedish National Audit Office (1999). *Handbook in performance auditing. Theory and practice*. RRV 1999: 17. Stockholm: RRV.

Tacq, J. (1991). *Van probleem naar analyse*. De Lier: Academisch Boekencentrum.

Taphoorn, R.T.G, K. Paalman, and H.J. van der Meulen (1998). *Twee jaar agentschap: doelmatigheid bij de Dienst Justitiële Inrichtingen*. In: *Openbare Uitgaven*, no. 1, pp. 32-40.



Turksema, R.W. (2000). *The supply of day care*. Utrecht: ICS.

Vermeulen, W.J.V. (1992). *De vervuiler betaald: onderzoek naar de werking van subsidies op vier deelreinen van het milieubeleid*. Utrecht: Van Arkel.

Weimer, D.L. & A.R. Vining (1992). *Policy analysis. Concepts and practice*. Englewood Cliffs (NJ): Prentice Hall.



## Annexe 2 Data collection: (longitudinal) surveys

A survey is a systematic means of questioning people or organisations about a large number of issues. The questions may be asked in writing, over the phone or in the form of a personal interview. Although surveys are held at a single point in time, the questions may relate to a number of different periods. Surveys form a good basis for making pronouncements of both a descriptive and an explanatory (causal) nature.

In the case of longitudinal surveys, data are collected at different points in time. There are two possibilities in this respect: collecting the same type of data from the same group of respondents (which is known as a panel study) and collecting the same type of data from different groups of respondents (known as a trend study).

Surveys can also be used for making pronouncements on the effectiveness of government policy, provided that the three causality conditions referred to in section 4.3.1.1 are met.

### *Population and sample (units)*

Because it is generally impractical (and unnecessary) to survey an entire population, surveys tend to be based on *samples* taken from a given population. If you ensure that your sample is selected at random, you can make pronouncements about averages, linkages and effects that apply to the population as a whole.

The performance audits undertaken by the Netherlands Court of Audit are based on different populations. First of all, we are talking about central government and legal persons with statutory tasks. Other suitable populations (particularly in the case of facilitation audits) are provincial and local authorities (as was the case with the audit on the preparation of disaster plans). You might even consider including citizens in your audit population (as in the case of the audit report on juvenile crime).



However, populations may also consist of policy tools such as licences, investment incentive schemes and grants.

If your sample is to be statistically reliable, you must have a good idea of the size of the population. This is fairly easy to measure in the case of local and provincial authorities, but becomes more difficult in relation to grants, as there is no database in the Netherlands containing full information on all government grants.

#### Further information

For further information on how to sample and deal with non-response, you are advised to consult either the course material from the *Methods and techniques* course (accessible on the DDO domain on the Intranet) and/or one of the statistical advisers.

#### Questionnaires (written interviews)

Questionnaires are the best way of collecting data for surveys. The advantage of questionnaires is that they allow you to reach a large number of people whom you can then ask a relatively large number of questions. Compiling a good questionnaire is something of an art, though. You are advised to devote plenty of time to this.

#### Further information

The following are examples of publications you could consult for further (accessible) information on longitudinal and other surveys, measurement scales, reliability, validity and/or questionnaires: Baarda & De Goede (1995), *Basisboek Methoden en Technieken*; Baarda, De Goede & Kalmijn (2000), *Basisboek enquêteren en gestructureerd interviewen*; Swanborn (1987), *Methoden van sociaal-wetenschappelijk onderzoek*.



## Annexe 3 VBTB budgets

New-style ministerial budgets should have the following format:<sup>98</sup>

New-style ministerial budget
<ul style="list-style-type: none"> <li>• Bill               <ul style="list-style-type: none"> <li>- Budget statement</li> </ul> </li> <li>• Summary of contents</li> <li>• <i>Policy section</i> <ul style="list-style-type: none"> <li>- <i>Policy agenda</i></li> <li>- <i>Policy articles</i></li> </ul> </li> <li>• Section on operational management</li> <li>• Departmental agencies</li> <li>• Annexe with detailed information</li> </ul>

The policy *agenda* describes the ministry's policy priorities. This means in any event spelling out what exactly the ministry intends to do during the coming year in order to put its policy plans into effect. The policy agenda also contains a review of the financial consequences of the proposed policies. Finally, the main spending increases are also shown, as are any proposed cuts.

The bulk of the policy section is taken up by the policy *articles*, which are intended to provide a means of establishing a direct link between (1) the objectives of the ministry's policy, (2) the tools the ministry intends to use and the outputs it is planning to deliver in order to achieve these objectives, and (3) the available financial means. The policy articles should be drafted in such a way that the policy fields are clearly identifiable and sufficiently homogeneous. Moreover, the policy articles should always contain performance data on the outcomes the ministry is seeking to achieve, even if the minister in question is not (directly) responsible for all of them, either because this responsibility has been

<sup>98</sup> Netherlands Court of Audit, 2002c.



transferred in full or in part to an external body or because it has been devolved to a provincial or local organisation.

So as to make it easier to compare policy articles with each other, the new-style policy articles have a fixed format. This is as follows:

Policy article new style
<ul style="list-style-type: none"><li>• General policy objective</li><li>• Operationalised policy objectives</li><li>• Budgetary consequences of policy</li><li>• Budget flexibility</li><li>• Assumptions about outcomes, efficiency and estimate</li></ul>



## Annexe 4 Criteria for assessing the formulation of policy objectives

Main criterion	Sub-criteria	Notes
Specific: a. outputs and outcomes (processes, where relevant) b. target group	<p><b>a. Outputs and outcomes</b></p> <ul style="list-style-type: none"> <li>Policy objectives should preferably be formulated in terms of outputs (i.e. what specific outputs is the auditee planning to deliver?) and the social outcomes the policy is intended to achieve (i.e. what social situation or change does the auditee wish to bring about?);</li> <li>If this is not possible, process-related objectives may be formulated.</li> </ul> <p><b>b. If the policy is not targeted at the entire population:</b></p> <ul style="list-style-type: none"> <li>the target group must be defined and delineated as clearly as possible;</li> <li>the auditee must explain why it has selected this particular target group.</li> </ul>	<p>An output objective relates to a policy output which the auditee is able to deliver, without it being dependent on others or on social trends over which the policy in question does not have any influence. In the case of a social impact objective, the policy is generally one of a number of determinants.</p> <p>If it is not possible to define an output or a social impact objective, a process-related objective may be formulated as an alternative. Product quality, which is often difficult to measure, is a case in point here. In this case, the objective may be defined in terms of the progress achieved by measures to improve quality (i.e. the process).</p>
Measurable	<ul style="list-style-type: none"> <li>The objectives must be defined in measurable terms, i.e. a clear description should be given of the situation that should exist once the objectives have been achieved, and terms must be clearly defined (i.e. they</li> </ul>	<p>Om te weten waarop de minister aanspreekbaar is, moet worden gelet op definities van begrippen en de praktische meetbaarheid van genoemde grootheden. Een nulmeting is niet altijd noodzakelijk. Identifying the precise scope of a minister's accountability requires you to look very closely at the way in which</p>





Main criterion	Sub-criteria	Notes
	<p>should not be open to interpretation) and measurable;</p> <ul style="list-style-type: none"> <li>Where objectives are defined in terms of a given reference point, information must be available on the state of the reference point, i.e. a baseline survey must have been performed.</li> </ul>	<p>terms have been defined and at the practical measurability of the variables used. Baseline surveys are not always required. For example, if an objective is formulated as follows: 'In 2010, the number of people suffering from noise pollution may not be greater than it was in 1985', a baseline survey needs to be performed of the situation in 1985. However, a baseline survey does not need to be performed if the objective is formulated as follows: 'In 2010, the number of people suffering from noise pollution may not be greater than 12,000.'</p>
Achievable	<ul style="list-style-type: none"> <li>The policy objectives must in any event be formulated in consultation with the relevant actors (e.g. the House of Representatives and executive agencies).</li> <li>Any decisions taken to adjust the policy objectives whilst the policy is in progress must be taken in consultation with the House of Representatives and executive agencies.</li> </ul>	<p>The policy objectives must be formulated in consultation with the House of Representatives and the relevant executive agencies. This particular criterion is all about the link between policy and implementation, i.e. policymakers must seek to establish first of all whether the policy they wish to pursue is achievable in practice.</p>
Realistic	<ul style="list-style-type: none"> <li>The policy objectives must be chosen in such a way that they can be achieved if the scenarios for the external variables are plausible.</li> </ul>	<p>There are two aspects to the realism of policy objectives:</p> <ul style="list-style-type: none"> <li>they must be defined in realistic terms;</li> <li>when formulating policy objectives, policymakers must take account of any constraints that may prevent them from achieving them (e.g. external factors and shared responsibilities for policy or policy outcomes).</li> </ul>
Time-related	<ul style="list-style-type: none"> <li>The auditee should state the</li> </ul>	<p>If no clear time restrictions (such as a</p>



Main criterion	Sub-criteria	Notes
	<p>date by which or the period within which the objective must be achieved.</p> <ul style="list-style-type: none"> <li>In the case of a long-term objective, the auditee should specify a time scale including clear milestones (such as intermediate objectives or the pace at which the objective is to be achieved).</li> </ul>	<p>deadline and/or milestones) have been set in advance, it is difficult to decide whether the objectives have been achieved in accordance with the predefined conditions.</p>
<p>Consistent</p>	<p>There are two types of consistency:</p> <p>1. Consistency between objectives:</p> <ul style="list-style-type: none"> <li>the policy objectives must be compatible with each other;</li> <li>the objectives formulated for individual aspects of the policy must be consistent with the general policy objectives (hierarchy);</li> <li>any intermediate objectives must be consistent with the ultimate objective;</li> <li>in the case of decentralised policy, the aims of central government should be consistent with those of the bodies responsible for implementing the policy (such as local authorities or executive agencies);</li> <li>the objectives must be formulated in a manner that is consistent in time. Any changes in the objectives must be explicitly stated and explained.</li> </ul> <p>2. Consistency between objectives and basic data:</p> <ul style="list-style-type: none"> <li>the objectives must be</li> </ul>	<p>Type 1 consistency:</p> <p>Policies often come with a whole range of objectives. These can be arranged in the form of an 'objective tree':</p> <ul style="list-style-type: none"> <li>in a hierarchy: objectives for individual policy measures → objectives for individual aspects of policy → general policy objectives;</li> <li>in time: intermediate objectives → ultimate objective.</li> </ul> <p>It is important that these objectives should be consistent with each other. The achievement of one objective should not imply that another objective cannot be met or makes it difficult to achieve another objective.</p> <p>Type 2 consistency:</p> <p>If a policy objective states that the quantity of waste should not grow from 40 million tonnes in 1986 to more than 50 million tonnes in 2000, the basic data on which it is based is that 40 million tonnes of waste was produced in 1986. If it subsequently emerges that waste production in 1986 was in fact 50 million tonnes rather than 40 million tonnes, the</p>



Main criterion	Sub-criteria	Notes
	<p>consistent with the data on which they are based. The objectives must be adjusted to take account of any changes in the basic data.</p>	<p>objective is no longer consistent with the data on which it is based.</p>

Below follow four examples of audits including assessments of the way in which policy objectives had been formulated.

#### *Example 1*

The SMART criteria were used in the audit of the policy on the prevention of greenhouse gas emissions (Netherlands Court of Audit, 2002a). The auditors sought to assess whether the policy objectives formulated by the Ministry of Housing, Spatial Planning and the Environment, the Ministry of Economic Affairs, the Ministry of Transport, Public Works and Water Management, the Ministry of Agriculture, Nature and Food Quality, and the Ministry of Finance in relation to greenhouse gas emissions were SMART, i.e. specific, measurable, achievable, realistic and time-related.

#### *Example 2*

The audit of the government's policy on the integration of ethnic minorities (Netherlands Court of Audit, 1999b) included an assessment of the following aspects of the policy objectives:

- *Completeness*: the auditors sought to ascertain how the objectives of policy measures had been formulated, i.e. as outputs, outcomes or both. The criterion here was that the objectives should be formulated in terms of both outputs and outcomes.
- *Measurability*: the objective of a measure was described as being measurable if (1) a measurable description had been given of the situation existing once the objectives had been achieved (i.e. the indicators in question should be clear and the desired situation should be formulated in quantitative terms), and (2) there was a clear time horizon, so that it was clear within which period or by which date the objectives should be achieved.



- *Consistency*: the auditors sought to ascertain whether the objectives specified for the policy measures in question were consistent with the general objectives of the government's policy on ethnic minorities.

### *Example 3*

The audit of the government's policy on waste prevention (Netherlands Court of Audit, 1997) included an assessment of the following aspects of the policy objectives:

- *Specificity*: the Netherlands Court of Audit disagreed with the Minister as to what exactly could and could not be classified as an objective. The Minister felt he was not accountable for policy objectives defined in information documents produced by the National Institute of Public Health and the Environment. The Netherlands Court of Audit accepted the Minister's standpoint because these documents were not parliamentary papers. The Netherlands Court of Audit found objectives in parliamentary papers which the Minister said should not be regarded as such, but rather as policy measures and the expected results of such measures. The Netherlands Court of Audit stuck to its view that the aims were objectives for which the Minister was accountable.
- *Measurability*: the objectives in question related to the prevention of waste. The added difficulty in this case involved measuring how much waste was not produced. The criterion used by the auditors was that the objectives should provide clarity on the following points: the year in which the objectives were to be achieved, the reference year and the situation in the reference year, the forecast for the autonomous growth in the volume of waste production, the pace at which the objective should be achieved, the waste substances covered by the objective, and the way in which the auditee was planning to measure the achievement of the objective.
- *Consistency*: the consistency of the objectives was operationalised in four different ways:
  1. *Consistency in time*. Although the objective was changed on a number of occasions, not all the changes were announced and explained in parliamentary papers.
  2. *Consistency with the basic data*. The objective was based on outdated information.



3. Consistency *between the general policy objective and the objectives for individual aspects of policy*. The objectives formulated for a number of high-priority waste substances had to concur with the general objective for the total quantity of waste.
4. Consistency *between central government objectives on the one hand and those applying to local authorities and firms on the other*. The auditors assessed whether the objectives formulated for central government policy had been translated into corporate environmental policy plans and environmental licensing procedures adopted by local authorities. The report referred to this aspect as the 'local impact'.

#### *Example 4*

The audit of the measurability of the government's childcare policy is a good example of the way in which the Netherlands Court of Audit deals with policy objectives that cannot be measured (Netherlands Court of Audit, 1991a). Because the aims of and the assessment criteria applying to the childcare incentive scheme were not formulated clearly enough, the Netherlands Court of Audit asked the minister a number of questions. The Netherlands Court of Audit used the answers given by the minister in combination with the underlying documentation to construct the final objective for the policy in question, together with a number of secondary objectives.

Although the minister was not prepared to accept all the objectives as formulated by the Netherlands Court of Audit, he did clarify the aims of the policy. In this instance, the auditors did not seek to establish whether the objectives had indeed been achieved.



## Annexe 5a Criteria for assessing policy information and its usefulness

Main Criteria	Sub-criteria	Detailed information on sub-criteria	Notes
<b>Relevance</b>		<p>Starting question / initial selection: If you come across information that is not needed for the purpose of reviewing and adjusting the policy in question (i.e. it is irrelevant), you will need to find out whether the information in question has been collected for this purpose.</p> <ul style="list-style-type: none"> <li>• If not, you should not give any opinion on it. The information may be relevant for another purpose. You should disregard it.</li> <li>• If it has, the opinion you should give is 'irrelevant'. You do not need to assess the quality of this information.</li> </ul>	<p>If data are to be collected efficiently and if efficient use is to be made of the funds set aside for this purpose, unnecessary data should not be collected. Which particular aspects are relevant depends on the nature of the policy field under review and the question that needs to be answered. In many cases, the question will depend on the stage that the policy in question has reached.</p>
	Complete	<ul style="list-style-type: none"> <li>• <i>All</i> information that should be available, in the light of the policy aims, in order to adjust the policy or account for it, should indeed be available.</li> <li>• No relevant information should be withheld.</li> </ul>	<p>Information is needed in order to understand a particular issue. Data need to be available on <i>all</i> relevant aspects.</p> <p>In virtually all cases, future scenarios and models are based on assumptions and simplifications of reality. It is important that these should be made clear. If certain relevant aspects and factors are not taken into consideration, this should be explicitly stated and explained.</p>
	Up-to-date	<ul style="list-style-type: none"> <li>• All the information present should be sufficiently up-to-date (i.e. not obsolete).</li> </ul>	<ul style="list-style-type: none"> <li>• The up-to-date requirement may not lend itself to quantification.</li> <li>• Data on outputs and outcomes should be reviewed at least once every five years, unless a different frequency was agreed when the policy was announced.</li> </ul>
	Punctual	<ul style="list-style-type: none"> <li>• The necessary information should be available on time (i.e. not too late).</li> </ul>	<p>At issue here is the order in which data on the one hand and decisions and actions on the other hand are made available. The sequence of steps in the policymaking chain should be logical. For example:</p> <ul style="list-style-type: none"> <li>• a problem analysis and an <i>ex ante</i></li> </ul>



Main Criteria	Sub-criteria	Detailed information on sub-criteria	Notes
			<p>evaluation should be conducted before policy tools are used.</p> <ul style="list-style-type: none"> <li>objectives (preferably including operational evaluation criteria) should be formulated when the policy is launched.</li> </ul> <p>You should take account of the cyclical nature of policy. This means that objectives may be adjusted or refined after the policy has been launched. There is nothing wrong with this, provided that the auditee says that this has been done and explains why it was decided to do so.</p> <ul style="list-style-type: none"> <li>An <i>ex post facto</i> evaluation should be conducted before any decision is taken on the continuation or adjustment of the policy.</li> </ul>
Reliability / accuracy	Reliable	<p>The measuring instrument should be used accurately. A series of measurements should generate the same results.</p> <ul style="list-style-type: none"> <li>Policy information should be sufficiently reliable.</li> <li>The reader should be able to judge the reliability of the data.</li> </ul> <p>If you're not sure about the reliability of the data:</p> <ul style="list-style-type: none"> <li>the degree of unreliability should be clearly stated (if possible, quantified, for example in the form of a reliability interval);</li> <li>the margin of uncertainty should be acceptable;</li> <li>if there is a high degree of uncertainty, you should compare data from different sources.</li> </ul>	
	Valid	<p>The measuring instrument should be technically adequate.</p> <ul style="list-style-type: none"> <li>Concepts should be measured validly, i.e. operationalised and measured in such a way that you measure what you want to measure..</li> </ul>	<p>There follow a few examples to illustrate the difference between reliable and valid:</p> <p>If you measure the temperature with the aid of a thermometer, but don't read it properly, the measurements are valid but unreliable. The discrepancies caused by unreliable measurements are random errors (i.e. a series of measurements produces irregular results; sometimes they are too high and at other times they are too low).</p> <ul style="list-style-type: none"> <li>Invalid measurements are caused by the use of an inadequate measuring instrument. If you measure the temperature with a barometer, or with a thermometer with an inaccurate scale, the measurements are not valid even though they may well be reliable. A lack of validity leads to systematic errors,</li> </ul>



Main Criteria	Sub-criteria	Detailed information on sub-criteria	Notes
			<p>i.e. the measurements are consistently too high or too low. Invalid measurements may also result if you try and measure concepts that do not lend themselves to measurement. The point here, therefore, is to decide whether the indicators you have selected (and the assumption being that these <u>are</u> measurable) are capable of approximating the concept you wish to measure as closely as possible. For example, the number of unemployed people can be measured by taking the number of people registered at job centres. However, this figure is valid only if all unemployed people do indeed register with job centres.</p> <p>Another cause of systematic distortion is the tendency for people to give socially desirable answers when they are interviewed or asked to complete a questionnaire. In practice, it is sometimes hard to differentiate between reliability and validity. The important thing is to make clear why the data in question are inaccurate, and where the problems lie. You can usually assess the accuracy of data by establishing whether a number of conditions have been met:</p> <ul style="list-style-type: none"> <li>• The data should have been compiled impartially. Evaluations should preferably be performed by staff who are not members of the department or institution responsible for preparing or implementing the policy.</li> <li>• The method of data collection should satisfy the relevant technical requirements. You should base your assessment of the quality of data sources on the requirements set out in the Dutch Central Government Performance Data and Evaluative Studies Regulations.</li> <li>• The accuracy of the data must be guaranteed (for example, by performing a peer review or by checking the data input).</li> <li>• When assessing data derived from other sources, you can include the quality of the relevant sources in your assessment (in the case of models, for example, this means assessing the quality of the input data used).</li> </ul>
<p>Conformity with statutory regulations and authoritative guidelines for the contents</p>	<p>The quality requirements for evaluations set out in the Dutch Central Government Performance Data and Evaluative Studies Regulations</p>	<p>Quality requirements for standard performance data set out in the Dutch Central Government Performance Data and Evaluative Studies Regulations</p> <p>Examples of these quality requirements:</p> <ul style="list-style-type: none"> <li>• the object of the system should be validly defined;</li> <li>• the system should be well designed;</li> <li>• the definitions and terms used</li> </ul>	





Main Criteria	Sub-criteria	Detailed information on sub-criteria	Notes
	<p>The following are examples of quality requirements contained in the Dutch Central Government Performance Data and Evaluative Studies Regulations:</p> <ul style="list-style-type: none"> <li>• a description should be available of the purpose of the evaluation;</li> <li>• the instructions and questions applying to the evaluation should follow logically from its stated purpose;</li> <li>• the findings of the evaluation should serve a practical purpose; the conclusions and/or recommendations should be in line with the stated purpose of the evaluation.</li> </ul>	<p>should be clear;</p> <ul style="list-style-type: none"> <li>• it should be possible to make comparisons over a number of years.</li> </ul>	
Comprehensibility	Accessible / clear	It is important that the data should be accessible to those who use it (e.g. policymakers, members of the House of Representatives, and so forth).	
	Unambiguous	The data should not be open to more than one interpretation.	
Comparability	Consistent	<p>There are three types of consistency:</p> <ol style="list-style-type: none"> <li>1. Consistency in time: <ul style="list-style-type: none"> <li>• data from different periods should be comparable with each other;</li> <li>• any changes in previously supplied data should be</li> </ul> </li> </ol>	<p>Type 1 consistency:</p> <p>In the case of consistency in time, the point is whether it is possible to compare data for different years or periods of time (e.g. are the subject you wish to measure and the measuring method you are planning to use consistent in time?).</p>



Main Criteria	Sub-criteria	Detailed information on sub-criteria	Notes
		<p>clearly stated.</p> <p>2. Consistency between different parts of the policymaking chain:</p> <ul style="list-style-type: none"> <li>• data should correspond with the way in which policy has been defined;</li> <li>• the way in which outputs are measured should correspond with the policy tools used and the objective formulated for the output in question; the way in which outcomes are measured should correspond with the social problem in question and the objective formulated for the outcome in question.</li> </ul> <p>3. Consistency between different data sources:</p> <ul style="list-style-type: none"> <li>• different data on the same subject should be mutually consistent;</li> <li>• data from different sources on the same subject should be compared with each other and preferably presented alongside each other;</li> <li>• if certain data or aspects of data are inconsistent with each other, this should be explicitly stated and, if possible, explained.</li> </ul>	<p>Type 2 consistency:</p> <p>In the case of consistency between different parts of the policymaking chain, the point is that data on social problems, aims, inputs, outputs and outcomes should be consistent with each other. For example, data on outputs and outcomes should correspond with the way in which the policy aim in question has been defined. Key indicators should consistently be used in the same way (e.g. the definition of the target group, years, definitions and operationalisations used). If not, reasons should be given.</p> <p>NB: Inconsistency in data on successive parts of the policymaking chain may also be caused by problems in the logical structure of the policy in question. If there is no logical relationship between the various elements of the policy and between the various stages of the policymaking process, the relevant data are not likely to be mutually consistent.</p> <p>Type 3 consistency:</p> <p>If different data sources are inconsistent with each other, this may be because of a difference in measuring techniques, samples, response, etc. The problem may also be caused by conflicting interests. The risk with inconsistent data, if no explanation is given for the inconsistencies, is that the data itself may be controversial and/or that the parties concerned will decide to adopt those findings that correspond most closely with their own interests.</p>
Conformity with statutory regulations and authoritative guidelines for the form	The quality requirements for evaluations set out in the Dutch Central Government Performance Data and Evaluative Studies Regulations	<p>Examples of these quality requirements:</p> <ul style="list-style-type: none"> <li>• The evaluation report should contain a summary.</li> </ul>	



## **Annexe 5b Details of the 'completeness' sub-criterion in relation to policy information (part of the 'relevance' criterion)**

The type of data required depends on the stage that the policy in question has reached. We have broken the process down into the following three stages (see the three columns in the table below):

1. the preparation of new policy. This may involve both data from a previous policy cycle and other data (i.e. prior to t-0);
2. decision-making, in which the foundations are laid for the policy in question (i.e. t-0);
3. implementation and enforcement. Here, the evaluation is performed against the background of a desire to continue, adjust or terminate the policy in question (i.e. t-1).

The availability of data on a limited number of subjects has a bearing on every policy field and should therefore form part of every audit. The availability of data classified as 'optional' does not form part of the basic remit. You will have to make a judgement, for each policy field and each individual audit, as to whether the information is relevant from a policymaker's standpoint. If so, you should be able to include it as part of the audit material on which your judgement is based. At the same time, the audit team should try and ascertain, in relation to every objective specified (see the third column in the table below), what sort of information is required given the nature of the objective.



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
Social problem	<ul style="list-style-type: none"> <li>The nature and scale of the social problem (in qualitative or quantitative terms);</li> <li>The main causes of the social problem and the extent to which policy is capable of influencing these.</li> </ul>	<p>Optional:</p> <ul style="list-style-type: none"> <li>the reasons why the auditee is planning to deal with this problem;</li> <li>the actors involved in the problem, and their interests.</li> </ul> <p>Optional: reasons given for the policy</p>	<ul style="list-style-type: none"> <li>The nature and scale of the social problem, both when the policy was launched (i.e. t-0) and today (i.e. t-1). Incidentally, if the objective is defined in terms of the social outcome, this will also be a means of measuring the outcome.</li> </ul> <p>Optional:</p> <ul style="list-style-type: none"> <li>The situation between the two dates.</li> </ul>
Objectives	<ul style="list-style-type: none"> <li>The objective that the auditee is seeking to achieve.</li> </ul> <p>Optional:</p> <ul style="list-style-type: none"> <li>Factors taken into consideration when choosing the objective.</li> </ul>	<p>2. The objective that the auditee is seeking to achieve.</p> <p>Optional:</p> <p>3. Reasons given for the formulation of the objective.</p> <p>4. The objective that the auditee is seeking to achieve.</p> <p>5. In the case of a future objective (t-2): deducing from this</p>	



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
		<p>what should have been achieved on the date on which the data were collected (i.e. t-1), for example by means of interpolation .</p>	
<p>Input: policy tools</p>	<ul style="list-style-type: none"> <li>- The policy tools that the auditee is planning to use.</li> </ul>	<ul style="list-style-type: none"> <li>- The policy tools that the auditee is planning to use.</li> </ul>	<ul style="list-style-type: none"> <li>- The policy tools that the auditee actually used.</li> </ul>
<p>Reasons for choice of policy tools</p>	<ul style="list-style-type: none"> <li>- alternative Alternative policy tools</li> <li>- The reasons given for the choice of policy tools (i.e. Why is the tool in question better than the rest? Have all relevant arguments been given?)</li> <li>- The logical relationship between policy tools and policy aims (i.e. plausible reasons why the tools help the</li> </ul>	<p>Optional:</p> <ul style="list-style-type: none"> <li>- Underlying reasons</li> </ul>	<p>Not applicable</p>



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
	<p>auditee achieve its aims, partly in the light of the controllable factors identified by the problem analysis).</p> <p>Optional:</p> <ul style="list-style-type: none"> <li>- Realism of objective, anticipated outcomes and potential side effects.</li> </ul>		
Input: cost and funding	<ul style="list-style-type: none"> <li>- The cost of implementing the proposed policy (i.e. budgeted cost).</li> <li>- The way in which the auditee is proposing to fund the cost.</li> </ul>	<ul style="list-style-type: none"> <li>- The cost of implementing the policy.</li> </ul>	<ul style="list-style-type: none"> <li>- The cost of implementing the policy (i.e. the actual cost, where applicable as compared with the budgeted cost).</li> </ul>
The parties involved and division of tasks and responsibilities	<ul style="list-style-type: none"> <li>- Who is responsible for what?</li> </ul> <p>Optional:</p> <ul style="list-style-type: none"> <li>- Have all the relevant actors been consulted?</li> <li>- If more than one government body is involved in the policy field in question: does</li> </ul>	<ul style="list-style-type: none"> <li>- Who is responsible for what?</li> </ul>	<ul style="list-style-type: none"> <li>- Who did what, against the background of everyone's individual responsibilities?</li> </ul>



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
	<p>the proposed central government policy interfere with those policies pursued by the other government bodies?</p>		
<p>Outputs and outcomes (where relevant, the process):</p>	<p>Optional:</p> <ul style="list-style-type: none"> <li>- Arrangements made for the evaluation of outputs and outcomes (who, when?).</li> </ul>	<p>Optional:</p> <ul style="list-style-type: none"> <li>- Arrangements made for evaluations.</li> </ul>	<ul style="list-style-type: none"> <li>- You must be able to assess whether the outputs and outcomes (and, where relevant, the process) are adequate, by comparing them with the objectives. The data must relate to all the objectives that have been defined. The minimum data requirement depends on the way in which the objective has been defined: if the objective has been defined in terms of outputs, policy outputs will need to be measured accordingly. If the objective has been formulated in terms of social outcomes, the outcomes will need to be measured accordingly. In the case of an objective that has been defined in terms of a process, data on the process will be required. Ideally, the objective should have been formulated in terms of both outputs and social outcomes, in which case data will need to be</li> </ul>



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
			<p>available on:</p> <ol style="list-style-type: none"> <li>1. the policy outputs generated by the auditee;</li> <li>2. the social outcomes;</li> <li>3. the outputs and outcomes must be linked with the policy aims (i.e. you will need to make a pronouncement on the degree to which the auditee has achieved its aims).</li> </ol> <p>Optional:</p> <ul style="list-style-type: none"> <li>- Data should also be available on any side effects (i.e. unintended outcomes, which may be either desirable or undesirable).</li> <li>- Providing a plausible explanation to what extent social outcomes have been achieved thanks to the policy outputs (i.e. establishing a link between a. and b.). There is also a tougher requirement in this respect, which is to provide empirical evidence for this causal link (i.e.</li> </ul>





Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
			<p>measuring the effectiveness of policy).</p> <ul style="list-style-type: none"> <li>- Providing a plausible explanation to what extent the side effects may be ascribed to the policy outputs.</li> </ul>
<p>The degree of efficiency</p>	<p>Optional:</p> <ul style="list-style-type: none"> <li>- The efficiency criteria</li> </ul>		<p>Optional:</p> <ul style="list-style-type: none"> <li>- The nature and scale of the outputs or products delivered, if necessary taking account of their quality.</li> <li>- The cost incurred (in terms of time and money).</li> <li>- The cost per output or product.</li> <li>- The actual cost needs to be compared with a target figure.</li> </ul> <p>The above target figure is generally relative, i.e. it involves making a comparison over time or with other bodies (efficiency).</p> <ul style="list-style-type: none"> <li>- An explanation should be provided if the actual cost does not match the target figure.</li> </ul>
<p>Implementation and enforcement problems / success and failure factors</p>	<p>Optional:</p> <ul style="list-style-type: none"> <li>- What sort of problems are likely to occur?</li> <li>- What steps has the auditee taken to</li> </ul>		<p>Only if the aims have not been achieved or are not likely to be achieved:</p> <ul style="list-style-type: none"> <li>- Information on the reasons why it has not proved possible to achieve a certain aim. These causes</li> </ul>



Stage: Information on:	Preparation of new policy (i.e. prior to t-0)	Decision-making (t-0)	Implementation and enforcement (t-1)
	<p>prepare for these problems?</p> <ul style="list-style-type: none"> <li>- What steps has the auditee taken to ensure that the launch of the new policy proceeds smoothly?</li> </ul>		<p>may be linked with:</p> <ol style="list-style-type: none"> <li>1. practical aspects surrounding the enforcement of the policy in question, such as the degree to which and the way in which certain policy tools have been used;</li> <li>2. the division of tasks and responsibilities;</li> <li>3. the extent to which all relevant actors have contributed as planned (by giving their assistance or support, for example);</li> <li>4. unforeseen social changes;</li> <li>5. the realism of the objective.</li> </ol>
<p>Any changes made during the course of policy planning and implementation</p>	<ul style="list-style-type: none"> <li>• Changes in data during the process of policy planning: these should always be explicitly stated.</li> </ul>		<ul style="list-style-type: none"> <li>• Changes in data since the policy was launched: these should always be explicitly stated.</li> </ul>



## Annexe 5c Examples of the use of criteria taken from previous audits

Of the examples cited in this annexe, only the audit reports on the growth figures for Amsterdam Schiphol Airport and on the tax relief for apprenticeship schemes relate to the policy planning stage. All other reports relate to the implementation stage.

### FIRST MAIN CRITERION: RELEVANCE

#### Example 1:

In 1997, the Netherlands Court of Audit examined the quality of the indicators used in ministry annual reports (Netherlands Court of Audit-, 1997c). Readers should have a clear idea of the meaning of any indicators quoted in the explanatory notes. A link needs to be established between policy and policy objectives on the one hand and the indicators on the other, and there should be a clear link between each indicator and the figure to which it relates. In many annual reports, this was not the case and the Netherlands Court of Audit concluded that insufficient information was provided on the relationship with policies and budgets (pp. 16-17). The data provided in this situation could be described as being irrelevant.

Incidentally, the *Dutch Central Government Performance Data and Evaluative Studies Regulations* state explicitly that systems of standard performance data, like regular evaluations, should be fit for their stated purpose (Ministry of Finance, 2001, p. 37).

#### Example 2:

The proposal for an audit of accounting procedures in the healthcare sector (Annual Healthcare Review 3, January 1999) contains a set of criteria. The aim of the audit is to ascertain whether both the technical data on healthcare provision and the financial data are relevant. Given



the nature of the target group, i.e. members of the House of Representatives, the data needs to be highly aggregated and should form part of a time series covering a number of years. Data is deemed to be relevant if it relates to:

- vital aspects of the current state of healthcare provision;
- the main trends;
- the main aspects of policy.

### **Sub-criterion: completeness**

Many audits of the Netherlands Court of Audit seek to assess whether the available policy information is complete. Below follow descriptions (by no means exhaustive) of a number of cases. These have been arranged in accordance with the aspect of policy information which the auditors were seeking to examine.

#### *General*

##### Example 1:

The proposal for an audit of accounting procedures in the healthcare sector (Annual Healthcare Review 3, January 1999) contains a set of criteria. The technical data on healthcare provision has to meet the following requirements in respect of completeness. Information on policy should in any event contain a description of the external policy aims, the results (in terms of outputs and outcomes) and the relationship between the aims and the results. A link should also be established with the inputs used. Where possible, technical data on healthcare provision should be provided in the form of indicators.

The Annual Healthcare Review is not the only source of information, incidentally. Other sources are *De staat van de gezondheidszorg* ('The state of the healthcare sector') and *Volksgezondheid Toekomstverkenningen* ('The outlook for public health'). Any overlaps should be avoided.

#### Social problem

No suitable examples available.

#### Policy tools



Example 1:

The audit of noise pollution caused by military activities (Netherlands Court of Audit, 1998b) included an assessment of the measures planned and taken.

Example 2:

The audit of waste prevention (Netherlands Court of Audit, 1997a) included an assessment of the development, use and impact of policy tools.

*The reasons given for the choice of policy tools*

Example 1:

The audit of tax relief for apprenticeship schemes (Netherlands Court of Audit, 1998a) included a clear description of the criteria for policy planning (see Annexe 1 to the audit report). Many of these criteria have a bearing on the policy information provided at the time when new policy is prepared. The most crucial requirement is stated as being the following: an adequate plausibility test of the assumed link between the policy tool which the auditee is planning to use and the aim which the auditee is seeking to achieve. This requirement was not met in the case of tax relief for apprenticeship schemes.

Example 2:

The audit of central government public-information campaigns (Netherlands Court of Audit, 1991b) stated that the reasons given for using public-information campaigns as a policy tool should satisfy the following requirement: public information that is intended to influence attitudes and/or behaviour is acceptable only if (a) the policy in question has been ratified by parliament, (b) the policy does not concern a controversial issue, and (c) the campaign is based on technically sufficient and accurate information that enables those concerned to reach an independent judgement.

Example 3:

The audit of government grant schemes (Netherlands Court of Audit, 1990) sought to establish whether *ex-ante* evaluations had been performed prior to the introduction of grant schemes worth more than



one million guilders. Where no such evaluation had been performed, a written explanation had to have been provided. An *ex-ante* evaluation of a grant measure should examine, *inter alia*, the assumed relationship between the grant as a policy tool and the policy objective in question. In other words, the auditee must provide a plausible explanation of the role played by the tool in achieving the policy objective. There is a stricter requirement, which is that a comparison should be made of the various possible tools and possible combinations of tools on the basis of a number of criteria, including in any event the extent to which the tool (or combination) in question is likely to help the auditee achieve its policy objective.

#### The cost of policy

##### Example 1:

The audit of government licensing schemes (Netherlands Court of Audit, 1996) sought to establish whether the government had a clear idea of the cost of such schemes. The following expenses and receipts were included in the audit: the cost of issuing licences, the value of the administrative charges received and other licence-related receipts (i.e. money received from renting out state-owned property). The audit did not include the cost of enforcement.

##### Example 2:

The June 1994 report describes two audits which sought to analyse implementation costs (the Netherlands Court of Audit, 1994). In the first of these, involving environmental grants, the Court examined the cost of implementing five grant schemes for private-sector firms (pp. 42-56). The absence of a system of accounting for time and cost meant that it was not possible to calculate the cost of implementation. In the second audit, on the nature conservation policy plan, the Netherlands Court of Audit found that accounting procedures at the ministry in question had not been designed to provide clear information on the funds spent in putting the plan into effect.

#### The parties involved and division of tasks and responsibilities

##### Example 1:

The audit of the management of social security funds (Netherlands Court of Audit, 1995a) examined the division of tasks and responsibilities in



relation to the management of three social security funds. The funds in question were deemed to be properly managed if it was clear from the relevant regulations that:

- the regularity of the amounts involved should be audited by a person or body other than the body responsible for their collection or distribution, i.e. an external auditor, a fund manager, a supervisory authority or an inspector working for the civil service;
- the inspector should possess the necessary authority to carry out inspections, such as the right to inspect accounts and documents;
- the audit responsibilities of the fund manager and the inspector did not overlap;
- the fund manager was able to determine the scope and findings of the inspection.

#### Example 2:

One of the issues examined in the audit of the Pension Insurance Advance Payment Fund (Netherlands Court of Audit, 1997b) was whether the way in which tasks and responsibilities were divided over a number of bodies, in pursuance of the Act and the Decree on the Pension Insurance Advance Payment Fund was both comprehensive and cohesive (pp. 12-14).

#### Implementation and enforcement problems / success and failure factors

#### Example 1:

The reason for the audit of severance payments made by universities, university teaching hospitals and research institutes (Netherlands Court of Audit, 1995b) was the growth in both the number and the monetary value of such payments, despite the existence of a policy of reducing them on both these counts. The Netherlands Court of Audit sought to establish the causes of this disparity.

#### Efficiency

#### Example 1:

In its audit of the enforcement of the EU's scheme for arable farming, the Netherlands Court of Audit interpreted the term 'efficiency' as meaning that the scheme should be implemented in such a way that a preset standard of service was delivered at a minimum level of cost (Netherlands Court of Audit, 1998d). The executive agency had already performed an enquiry of its own into the efficiency of five regional offices, and had



found that the amount of time taken to process a single application varied widely from one office to another. The Netherlands Court of Audit sought to establish whether the data used were complete, accurate, timely and regular. It also examined the explanations given for the discrepancies in handling times, and assessed whether account had been taken of all possible factors (pp. 17-19).

Example 2:

The audit of the right of recovery under the Health Insurance Act (Netherlands Court of Audit, 1996c) was aimed at identifying whether health insurance funds recovered the costs incurred in treating physical injuries caused to their members from the perpetrators or, as the case may be, from the latter's insurance companies. The fact was that there were wide differences, both between health insurance funds and over time, in the average sums recovered per insured person. The Netherlands Court of Audit tried to explain these differences and sought to establish whether the health insurance funds were in a position to exert any influence on the causes. The Netherlands Court of Audit used the results of this study to formulate minimum requirements for the recovery of costs.

Any changes made during the course of policy planning and implementation

Example 1:

The Netherlands Court of Audit made the following recommendation in its report on the audit of the growth figures for Amsterdam Schiphol Airport: 'Any changes in input data that have a material effect on the results of calculations should be explicitly stated before a decision is taken.' (p. 61.)

Example 2:

One of the findings of the audit of waste prevention (Netherlands Court of Audit, 1997a) was that there had been a shift in emphasis in the government's waste prevention policy. The House of Representatives was not informed of this shift until after the Netherlands Court of Audit had completed its audit.

Example 3:





When it audited the efficiency of job centres (Netherlands Court of Audit, 2001), the Court collected data on the cost and performance of job centres and made full preparations for using this data to calculate their relative efficiency and to identify explanations for any discrepancies. The Court used a range of econometric techniques for this purpose. The results were so bizarre, however, that the Court was forced to conclude that the underlying data were too poor to be used as a basis for making pronouncements about the efficiency of the job centres in question.

*Outputs and outcomes (achievement of policy objectives)*

Example 1:

When the Court audited the growth figures for Amsterdam Schiphol Airport (Court of Audit, 1998c), it examined the data available on the achievement of policy objectives. The Court used environmental impact assessment reports to decide whether it was going to be possible to meet the environmental targets for 2015 set in the government's Key Planning Decision. The Court concluded that the government was unlikely to achieve its aims if there was no change in the current trend. In other words, there is no need to wait until the target year itself in order to form a judgement on the auditee's success or otherwise in achieving its policy objectives. You can ascertain at an earlier juncture whether the targets are realistic or whether they are likely to be achieved at some future point in time.

Example 2:

The Court's audit of government grant schemes (Court of Audit, 1990) showed that a ministry is more likely to know whether grants are having their desired effect if an *ex-ante* evaluation has been performed, if evaluation criteria have been drawn up in advance and if an *ex post facto* evaluation is conducted. At the same time, the mere fact that these three 'conditions' have been met is no guarantee that the ministry will automatically know whether grants are having the desired effect.

Example 3:

The Court's report on its audit of central government public-information campaigns (Court of Audit, 1991b) divided information on results and the achievement of objectives into three distinct categories:



- Reach: the degree to which the target group has been brought into contact with one or more communication tools, and has received, understood and accepted the message.
- Impact: the changes in the target group's knowledge, attitudes and/or behaviour resulting from the information campaign.
- Effectiveness: the contribution of the information campaign to the achievement of the policy objectives.

The report also quoted percentages for reach, impact and effectiveness from a large number of previous campaigns.

The Netherlands Court of Audit said that it preferred *ex post facto* evaluations to data on circulation and distribution, such as viewing figures and statistics on the number of people participating in certain activities. Where the campaign in question is relatively low-cost (e.g. costing less than, say, NLG 500,000 or NLG 1 million), a cost-benefit analysis may prompt the auditee to do without an *ex post facto* evaluation.

#### **Sub-criterion: up-to-dateness**

Example 1:

When the Netherlands Court of Audit audited the growth figures for Amsterdam Schiphol Airport (Netherlands Court of Audit, 1998c), it found that the calculation of external security appeared to be based on outdated data on aircraft takeoff weights.

#### **Sub-criterion: punctuality**

Example 1:

When the Netherlands Court of Audit examined the growth of Amsterdam Schiphol Airport (Netherlands Court of Audit, 1998c), it found that certain data had been published too late to be used in the debate and decision-making process.

### **SECOND MAIN CRITERION: RELIABILITY AND ACCURACY**

Example 1:

When the Netherlands Court of Audit audited the growth figures for Amsterdam Schiphol Airport (Netherlands Court of Audit, 1998c), it assessed the reliability of the policy information. It concluded that the



data underlying the Key Planning Decision was clouded in uncertainty because one assumption had been built on top of another. This meant that the Key Planning Decision was a high-risk decision. The Court also assessed the accuracy of models by asking the following questions:

- Are the assumptions underlying the forecasting models correct?
- Does the model take account of all relevant factors?
- Are the forecasts adjusted to take account of recent changes?
- Have the correct input data been used?

#### Example 2:

In its 1997 regularity audit, the Netherlands Court of Audit assessed the quality of the reporting information included in the Budget Memorandum. The criterion applied in this case was that any doubts should be reported and, if possible, quantified (p. 43).

#### Example 3:

When it audited the development and redistribution of policing levels in 1994-1995 (Netherlands Court of Audit, 1997d), one of the investigated aspects was the data on policing levels and the reporting information on the way in which the relevant financial resources were spent. The audit followed up on three previous audits on the same subject. The criterion applied by the Netherlands Court of Audit was that the accuracy and completeness of the data needed to be guaranteed. This requires not simply the presence of sufficient guarantees for the way in which data is recorded and checked, but also that the terms and definitions used should be clear and easy to operationalise.

For instance, no clear definition was provided of the phrase 'officers on the beat', which did not equate in practice with the term 'executive staff' (an example of insufficient validity). In addition, no definition was given of the term 'actual policing levels', which meant, for example, that one regional police force included volunteers, trainee police officers and/or parking wardens in the headcount, whereas others did not (an example of insufficient reliability). The Netherlands Court of Audit also found that the data had not been checked by an independent third party (a guarantee of reliability), and that inaccurate information had been supplied to the House of Representatives on the ratio of staff to material expenses.

It is clear from the response from the Minister of the Interior and Kingdom Relations that he believes that reporting information should be subject to stricter reliability requirements than policy information.



#### Example 4:

The proposal for an audit of accounting procedures in the healthcare sector (Annual Healthcare Review 3, January 1999) contains a set of criteria. The Netherlands Court of Audit will be using the following criteria to verify the accuracy of the financial data:

- The reporting figures must be accompanied by an adequate auditor's report. Where public funds are involved, the report should take the form of a 'regularity statement'. If the funds involved are private-sector funds, the report should take the form of a 'true and fair view' statement.
- When assessing the actual-cost figures, the Netherlands Court of Audit will apply a margin of tolerance representing 1% of the total money flow as reported for the relevant sub-sector.

### **THIRD MAIN CRITERION: COMPARABILITY**

#### **Sub-criterion: consistency**

#### Example 1:

The audit of the government's policy on the integration of ethnic minorities (Netherlands Court of Audit, 1999b) revealed various types of inconsistency in the data:

- The minister was found to be using a different definition of the term 'waiting list' from that used by the auditors (a question of consistency between different parts of the policymaking chain).
- The availability of Dutch courses was sometimes expressed in terms of places and sometimes in terms of teaching periods (inconsistency between different sources of information). The absence of a conversion formula meant that it was not possible to add up these data.
- By comparing the findings of different evaluations and monitors, the auditors found that certain sources were consistently at odds with the rest. This was regarded as indicative of the situation as a whole. An explanation was subsequently found for the discrepancies in the form of a non-representative sample and response (inconsistency between different sources of information).

#### Example 2:

When the Netherlands Court of Audit audited the growth figures for Amsterdam Schiphol Airport (Court of Audit, 1998c), it found that the Key



Planning Decision was internally inconsistent because the traffic forecasts were based on a decline in air fares, whereas the substitution calculations were based on an assumption that fares would rise in future (i.e. a lack of consistency within the same data source; see pp. 30 and 37). The Netherlands Court of Audit made the following recommendation in this connection: 'Particularly in a situation in which there is a high degree of uncertainty, factors that can affect the outcome of forecasts should be treated on a consistent basis' (p. 60). The Netherlands Court of Audit went on: 'The transparency of the decision-making process is enhanced if forecasts are compared with forecasts produced by other parties, particularly when there is a high degree of uncertainty and long-term forecasts are made' (consistency between different sources of information, p. 61).



## Annexe 6 Description of the modus operandi method

The modus operandi method (which literally means 'way of working' in Latin) was devised by Michael Scriven, who also calls it the 'detective paradigm'. In fact, this is just a rather grand title for a method of elimination. Just as the police rule out various suspects one by one when they try to track down the perpetrator of a crime, so this method seeks to rule out the influence of a number of rival factors by examining them one by one.

The method consists of four steps, the final two of which are crucial, as they involve finding evidence for the role played by individual factors. This is done by establishing whether each factor had any effect by working in its characteristic way. The four steps seek to answer the following questions:

- What potential explanations ( $X_i$ ) are there for phenomenon Y?
- Which factors, that are capable of providing the explanations ( $X_i$ ), were indeed present?
- Which characteristic features reveal the causal paths leading from  $X_i$  (in so far as  $X_i$  is present) to Y?
- Of which  $X_i$  were such characteristic features indeed present in reality, and how much of these were present?

The procedure is repeated for every rival factor until a picture is obtained of the degree to which Y is the result either of the use of the policy tools under review or (either in full or in part) of rival factors.

The modus operandi method was used, *inter alia*, in a study performed by Bressers of the effectiveness of the government's water quality policy (Bressers, 1982).



## Annexe 7 Checklist: can you calculate an efficiency score yourself?

You are advised first to read Chapter 8 of this manual before completing the following checklist.

*Is the objective clear enough?*

1. Is the objective clear enough to enable the auditors to decide which outcomes and outputs the auditee is seeking to achieve with the aid of the policy in question, and if not, is the objective relatively easy to formulate?

Can the outcomes and outputs be measured?

2. Is it possible to measure the outcomes of the policy (if necessary, with the aid of an indicator or a proxy variable)?
3. Is it possible to measure the outputs of the policy (if necessary, with the aid of an indicator or a proxy variable)?

First decision:

When you take the first decision, you should think back to the step-by-step approach outlined in Chapter 8. So start at the point where the greatest number of opportunities lie. If the answer to question 0 is 'no', you will need to choose a different objective or refine the definition of the objective in question. If the objective is clear enough, and the answer to question 1 or 2 is 'yes', you can proceed.

If you can answer 'yes' to question 1, but not to question 2, you should take the phrase 'outcomes and outputs' as used below as referring solely to outcomes. If you can answer 'yes' to question 2, but not to question 1, you should take the phrase 'outcomes and outputs' as used below as referring solely to outputs.



Are the data<sup>99</sup> of sufficient quality?<sup>100</sup>

1. Are there data of sufficient quality on the outcomes and outputs, or are such data relatively easy to collect? Are the data available for at least two different periods or organisations?
2. Are there data of sufficient quality on inputs, or are such data relatively easy to collect? Are the data available for at least two different periods or organisations?
3. Are there data of sufficient quality on the quality of the outcomes and outputs, or are such data relatively easy to collect? Are the data available for at least two different periods or organisations?
4. Are there data of sufficient quality on explanatory factors, or are such data relatively easy to collect? Are the data available for at least two different periods or organisations?

Second decision:

In order to calculate an efficiency score, you should at least have been able to answer 'yes' to each of questions 3 to 5. If the answer to question 6 is also 'yes', you can start exploring the factors that may explain the differences in efficiency.

See section 8.6.1 for further information on those situations in which you should exercise caution in drawing conclusions and making recommendations.

---

<sup>99</sup> Or an approximation of the data, in the form of indicators or proxy variables.

<sup>100</sup> The term 'quality' means relevant, reliable and comparable. Relevant means, for example, that all outcomes and outputs *have a bearing on the policy* and that all inputs *have a bearing on the delivery of the outcomes and outputs*, etc. See Chapter 8 for further details.





## Annexe 8 General conditions applying to DEA and SFA analyses

The results of SFA and DEA analyses need to satisfy various conditions before they can be used. These conditions apply across the board to all situations. You can use the relationship identified with the aid of an SFA analysis to assess the degree of compliance with these conditions. This is less easy to do with the results of an SFA analysis.

In other words, if you discover, when performing an SFA analysis, that the conditions have not been met, you know in theory that the same applies to a DEA analysis. This means two things:

1. if you have not yet performed a DEA analysis, you might as well not perform one;
2. if you have performed both DEA and SFA analyses, you cannot use the findings of either of them. This is because there is an implicit assumption, when you use these techniques, that the conditions have been satisfied. If the conditions have not been satisfied, it is possible that the audited organisations could wrongly be labelled as being either efficient or inefficient.

What conditions are we talking about?

- The marginal cost is predominantly positive (based on the principle that there are no free lunches). The marginal cost is the additional cost associated with the delivery of a *single* extra unit of output. Additional inputs are needed if an organisation wishes to deliver one more unit of output, and this therefore means that additional inputs will have to be used. If the analysis shows that the opposite applies to a large number of organisations, i.e. that many organisations have a negative marginal cost, clearly then, something is very wrong (although it's often difficult to find out why) and the results of the analysis are not fit for use.



- If the analysis reveals the existence of variable returns to scale in the sector under review, i.e. that a change in all inputs leads to a disproportionate increase or decrease in the quantity of outputs delivered, this will generally mean that:
  - the smallest organisations generate rising marginal returns, i.e. a proportionate increase in all inputs leads to a disproportionate increase in the quantity of outputs delivered;
  - the medium-sized organisations generate constant marginal returns, i.e. a proportionate increase in all inputs leads to a proportionate increase in the quantity of outputs delivered;
  - the largest organisations generate falling marginal returns, i.e. a proportionate increase in all inputs leads to a disproportionate decrease in the quantity of outputs delivered.
- The situation may not be the reverse, i.e. first falling, then constant and then rising marginal returns.
  
- In conjunction with the previous point: if the analysis reveals the existence of variable returns to scale in the sector under review, in theory, the average cost for small and large organisations will generally be relatively high, whilst the average cost for medium-sized organisations will generally be relatively low, producing a U-shaped average cost curve. Most small organisations do not employ enough staff to enable them to specialise and have relatively high overheads. Large organisations tend to be more bureaucratic, because the management cannot control all aspects of the production process, leading to relatively high overheads. The situation may not be the reverse.
  
- There are three possible explanations for a situation in which these general conditions (applying to SFA and DEA analyses) are not satisfied:
  - the data is of poor quality;
  - not all relevant factors have been taken into account in assessing the relationship between the inputs used and the outputs generated (i.e. the detailed information is either incomplete or wrong);
  - the behaviour of organisations operating in the sector under review is chaotic and/or totally irrational.

The following example (i.e. the audit of job centres) illustrates how a prior analysis can be used to test a range of elements before performing the audit itself.



### *Job centres*

During the audit by the Netherlands Court of Audit of the efficiency of job centres (Netherlands Court of Audit 2001a), SFA was used to assess the relationship between the costs incurred by the job centres and the outputs they delivered (in terms of the number of jobseekers signing contracts of employment). The analysis showed that, for a large number of job centres, the level of cost tended to decrease as the number of outputs increased, i.e. the marginal cost is negative. The analysis also revealed that the average cost is the highest for medium-sized job centres, which is contradictory to the economic theory underlying SFA and DEA analyses. The latter finding is borne out by the efficiency indicators for job centres shown in the graphs in section 8.6.3.2, where the cost per jobseeker placed is plotted against the relative employability of the jobseekers on the centres' books. Although one would expect less employable jobseekers to cost more than more easily employable jobseekers, the graphs do not point to a clear correlation between the cost per placement and the relative employability of jobseekers. Independent of the number and relative employability of the jobseekers on a centre's books, the cost per placement is roughly between €1,000 and €3,000. Even if explicit account is taken of labour market conditions, there are still fairly wide gaps in the cost per placement.

This means that the conditions for the use of SFA and DEA are not satisfied.



## Annexe 9 Glossary

*Audit of the achievement of objectives:* an audit aimed at identifying whether the outputs and/or outcomes the policy was intended to achieve have actually been achieved.

*Comparability:* the data collected should enable comparisons to be made over time and/or between different organisations.

*Cost:* the value of the inputs used for operational management purposes.

*Effectiveness audit:* an audit aimed at identifying whether government policy is producing the desired results.

*Expenditure:* payments made on the basis of obligations.

*External causes:* (potential) explanatory factors that are beyond the control of the ministry and/or the organisation involved in implementing the policy in question.

*Facilitation audit:* an audit aimed at identifying whether the conditions the minister has put in place for an effective implementation of central government policy at regional, provincial or local level have worked successfully.

*Inputs:* staff, equipment, outsourced services and expenditure on financial and non-financial policy tools (such as grants, loans and public-information campaigns).

*Internal causes:* (potential) explanatory factors that are rooted in the ministry and/or the organisation involved in implementing the policy in question.

*Outcome efficiency audit:* an audit aimed at identifying the efficiency or cost-effectiveness of policy outcomes, by finding out:



- whether the outcomes in question could have been delivered with fewer inputs; or
- whether the same inputs could have delivered more outcomes.

*Outcomes*: the publicly perceptible results of policies.

*Output efficiency audit*: an audit aimed at identifying the efficiency of operational management, by finding out:

- whether the outputs in question could have been delivered with fewer inputs; or
- whether the same inputs could have delivered more outputs.

*Outputs*: the results of the work processes used by an organisation in seeking to achieve its policy objectives.

*Performance data*: there are three types of performance data:

- policy objectives, which should be operationalised as far as possible in terms of desired and actual *outcomes*, linked to certain target groups and to a specified time horizon;
- information on the cost price and quality of *outputs* delivered or to be delivered by the government (i.e. products and services);
- programme expenditure, supported as much as possible by *volume and price data* (such as the expected number of recipients of a given grant and the value of the grants paid).

*Policy information*: information on policy objectives, the intended and/or actual outcomes, the outputs either delivered in the past or to be delivered in the future, and the associated cost.

*Quality*: differences in the characteristics of the same outputs or outcomes.

*Relevance*: information which adequately reflect either the policy pursued by or the activities undertaken by the ministry or organisation(s) in question is said to be 'relevant'.

*Reliability*: there may not be any material inaccuracies or omissions in the data collected.

*Side effects*: the unintended social outcomes of policy, which may be either desirable or undesirable.

*SMART-C*: the criteria for the formulation of policy aims, i.e. Specific, Measurable, Agreed upon, Realistic, Time-Related and Consistent.



*Target-group audit:* an audit aimed at identifying whether the target group selected for the policy (e.g. citizens, public-sector organisations or private-sector firms) has been reached, and if so, how.

*Throughput:* all activities performed by a ministry or a third party in converting inputs into outputs.