Review of methods to measure the effectiveness of state aid to SMEs

Final report to the European Commission

Marco Mosselman M.Sc.

dr. Yvonne Prince

Zoetermeer, November, 2004

EIM
Business & Policy Research

Contents

PART III: Synthesis

Executive Summary

Introduction

Each year, EU Member States spend approximately € 6 billion on state aid to small and medium sized enterprises (SMEs). It is not quite clear to what extent the effectiveness of this state aid is being monitored and evaluated. This study is to identify whether and how the Member States evaluate the effectiveness of state aid to SMEs and to conclude by making recommendations as regards the best method (or methods) for evaluating the effectiveness of state aid to SMEs. The study was performed in 8 steps, in which literature research, interviews, questionnaires, a meeting with the Member States, expert consultation and thorough analysis, provided the information and insights that are combined in this report.

State Aid to SMEs

Due to their relatively high vulnerability to various market failures, the development of small and medium sized enterprises lag behind compared to larger corporations. The most prominent market failures are capital market imperfection and asymmetric information. One way in which Member States respond to these market failures is by providing state aid to SMEs. State aid to SMEs represents a small, but growing fraction of the total amount of state aid. The most important form of aid to SMEs is 'grants', which makes up approximately four fifths of the state aid to SMEs.

State Aid effectiveness studies

Evaluation is the systematic investigation of the effectiveness of economic intervention whereby the performance of support measures is judged against pre-specified standards or criteria. There are many definitions of effectiveness. For the purpose of this study, effectiveness includes not only the extent to which a policy objective has been achieved, but also the (positive and negative) side effects, additionality and displacement. It is important that evaluation studies attach importance to the causality of the relation between the aid measure and the observed effects (e.g. the economic gain).

Types of evaluation studies

There are many classifications of evaluation studies, each classifying many different evaluation approaches. The methodological soundness of these evaluation approaches is beyond dispute, but it is significant that different evaluation methods are suited for different applications. Different evaluation methods should be employed for each stage in the policy cycle and different evaluation methods should also be employed in different policy fields. Economic policy (including state aid to SMEs) needs economic and econometric evaluation methods.

The process of effectiveness evaluation

The process of public policy evaluation is described in literature. In general, four phases can be identified in an evalution study, each phase having its own characteristics: (1) the structuring phase; (2) the data collection phase; (3) the data analysis phase; (4) the judgement phase. There are many different techniques that can be used for designing evaluations, gathering data, analysing data, and making a judgment. The quality of an evaluation is influenced by the evaluation process design. The selection of the appropriate techniques must be based on the research question that underlies the evaluation.

Evaluation practice

The SME state aid evaluation practice varies greatly between the Member States. Of all the Member States that participated in our research, one sixth evaluates all state aid to SMEs, two thirds evaluate some state aid to SMEs and one sixth does not evaluate state aid to SMEs at all. Also the applied

methods differ considerably. They range from quite simple ex-ante multi-criteria analysis, to intricate ex-nunc or ex-post econometric programme evaluations.

Selected methods

There are many techniques that can be used for evaluating the effectiveness of state aid. Based on our research, we arrive at a classification of 8 different 'basic' methods. This 'long list' comprises both qualitative and quantitative methods, and methods with and without control groups. The 8 basic methods reflect an increasing level of sophistication. These methods are:

1 Qualitative description of first order effects.
2 Ex-ante evaluation of the policy theory.
3 Quantitative evaluation of the reach of the state aid.
4 Quantitative evaluation of the first order effects, based on a survey among assisted firms.
5 Quantitative evaluation of the first and second order effects, with control group.
6 Quantitative evaluation of the (first, second and) third order effects, based on an econometric model and a variety of qualitative and quantitative data sources, among which a survey with a control group.
7 Ratio analysis, based on quantitative analyses of secondary data.
8 Goal free programme evaluation.

Assessing evaluation methods

To condense our selection to a limited number of methods that best measure the effectiveness of state aid we apply a multi-criteria analysis. The 8 basic methods were assessed on 13 criteria. These criteria are:

1 Whether or not the method will **quantify** the effect of state aid.
2 Whether or not the method can **prove the causal relations** between the state aid and the effects.
3 Whether or not the method is **objective/neutral.**
4 Whether or not the method is **generally applicable.**
5 Whether or not the method is **transparent.**
6 Whether or not the method is **cost-effective.**
7 Whether or not the method is **time-effective.**
8 Whether or not the **required data** for this method are available or can be gathered easily.
9 Whether or not the method can take account of **deadweight.**
10 Whether or not the method can take account of **displacement.**
11 Whether or not the method can take account of **additionality.**
12 Whether or not the method can control for **company-specific background characteristics.**
13 Whether or not the method can control for **locational and other external factors.**

For the application of the 13 criteria to the 8 methods, we consulted a number of leading scientists and experts in the field of evaluating state aid to SMEs. Based on the results we selected three methods that are best for measuring the effectiveness of state aid to SMEs (methods 5, 6 and 8 as shown above). These methods are best for quantifying the effects of state aid, incorporating such effects as deadweight, additionality and displacement.

Conclusions and recommendations

The most important recommendation of our research is that the national governments should use method 5, 6 or 8 for evaluating the effectiveness of state aid to SMEs. These methods build on each other with method 5 being the least sophisticated and method 8 being the most comprehensive of the three methods. However, if the causality and side effects of the state aid are given less weight and the emphasis of the evaluation is more on improving the process, rather than on measuring the effectiveness, method 4 or 3 could be employed. Of the recommended methods, the one most suited to the

circumstances in, the evaluation experience of, and the constraints faced by the Member State should be selected.

PART I:

BACKGROUND

1 Introduction

Each year, EU Member States spend approximately € 6 billion on state aid to small and medium sized enterprises (SMEs). It is not quite clear to what extent the effectiveness of this state aid is being monitored and evaluated. This study is to identify if and how the Member States evaluate the effectiveness of state aid to SMEs, and to conclude by making recommendations as regards the best method (or methods) for evaluating the effectiveness of state aid to SMEs. The study was performed in 8 steps, in which literature research, interviews, questionnaires, a meeting with the Member States, expert consultation, and thorough analysis provided the information and insights that are combined in this report.

1.1 Motives for the study

Small and Medium Sized Enterprises (SMEs) are widely recognised as the motor for economic growth in the European Union (European Commission, 2003a). The recently adopted Green Paper on Entrepreneurship (European Commission, 2003b)[1] specifically points to the need for creating more new firms, making them grow and become more innovative and competitive. Facilitating SMEs' access to finance, knowledge and information is a key part of the strategy to create a business environment that stimulates the creation and success of SMEs.

Supporting SMEs can take the form of state aid (European Commission, 2003a). Annually, the EU Member States already spend approximately € 6 billion on state aid to small and medium sized enterprises (SMEs)[2]. These funds are distributed through many different support schemes, all with their respective policy objectives and support types. Some of them specifically targeting market failures, others taking a more general support approach.

In this framework, there is a clear need for more knowledge and better understanding of the effectiveness of various types of state aid measures to SMEs used by the Member States. State aid evaluation, however, does not take place in all the Member States. And of the Member States that *do* evaluate, some are more advanced than others. Evaluation methods are still developing and the SME support situation differs from one country to another. In this respect it is important to note that in some of the new Member States, the private SME sector as such did not exist at all until 15 years ago. As a result, these countries have a very different macro-economic starting point and possibly less experience with SME support evaluation than the other Member States.

With these developments in mind, the European Commission wants to have more information on which evaluation methods are suited for evaluating the various forms of state aid to SMEs (European Commission, 2003a).

1.2 Aim of the study

In the frame of this large institutional variety, with possibly many differences in SME state aid evaluation, the European Commission has commissioned a review of the different methods that the Member States employ to measure the effectiveness of state aid to SMEs. This review intends to gather more

[1] European Commission (2003b)

[2] State aid as defined by article 87(1) of the treaty. The amount concerns state aid for SMEs as either primary or secondary objective and excludes general measures. Therefore, the amount does not indicate how much state aid actually went to SMEs. Source: European Commission, State Aid Scoreboard, spring 2004 update.

knowledge and better understanding of the effectiveness of various types of state aid to SMEs used by the Member States. This study is intended to identify which methods are used to measure the effectiveness of state aid, and will make recommendations with regard to the most suitable evaluation methods[3] (European Commission, 2003a). This study will not identify *what* should be measured (the specific indicators) in order to evaluate SME-support, since that depends on the design and objectives of the specific programme.

Research question and objectives

This study focuses on answering the following research question: ***Which methods are best suited for evaluating the effectiveness of state aid to SMEs?***

The project has the following objectives:

1 To make an inventory of the methods used for measuring the effectiveness of state aid to SMEs in the Member States and in the Commission services;
2 To develop and present alternative methods that could be used to measure the effectiveness of state aid to SMEs;
3 To evaluate each of these methods;
4 To conclude by making recommendations as regards the best method (or methods) for evaluating the effectiveness of state aid to SMEs, taking into account data constraints.

In addition to these goals, this report contains an introductory chapter that briefly presents the development of state aid to SMEs over time. This overview of the evolution of aid amounts and aid instruments used by Member States serves as a background for the study.

1.3 Approach of the study

1.3.1 *Structuring the study*

This study is based on the assumption that there are several methods that could be employed to measure the effectiveness of state aid, but that some methods are better suited for this purpose than others. Choosing a method influences the quality of the evaluation's results, as regards the reliability of the outputs. This is due to the fact that the various methods have different ways of coping with such issues as the causality, additionality and comparability of results.

In order to answer the research question and to achieve the study's objectives, we present an overview of applied methods used to assess the effectiveness of state aid programmes, describe alternative methods and assess these methods according to multiple criteria. This study ends with drawing conclusions and making recommendations.

In the framework of this study the following activities were undertaken:

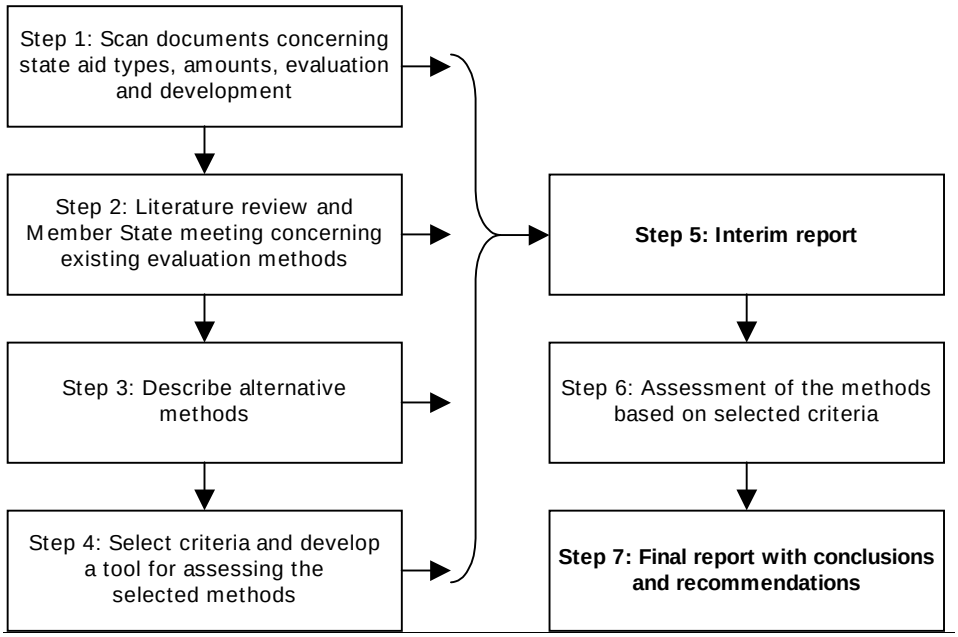
1 Scanning documents concerning state aid types, amounts, evaluation and development;
2 Reviewing literature and conducting a Member State meeting concerning existing evaluation methods;
3 Describing alternative methods;
4 Selecting criteria and developing a tool for assessing the selected methods;

[3] In many disciplines the concept 'methodology' is used synonymously with 'method'. In the remainder of this study, we will try to use the concepts 'method' and 'methodology' more consistently. We use 'method' when referring to *tools of scientific investigation*. We use 'methodology' when referring to *the principles that determine how such tools are deployed and interpreted*. This review is mainly a review of methods concerning one methodology.

5 Compiling an interim report;
6 Assessing the methods;
7 Drawing up a final report with conclusions and recommendations.

The relationships between the activities are displayed in figure 2 and an elaboration of the activities follows the figure.

Figure 1 Approach of the project



1.3.2 *Data collection*

Information about applied and alternative methods was collected from multiple sources. These sources include a literature study, a questionnaire to the Member States, a Member State meeting and the consultation of experts (see table 1).

Table 1 Information needed and data collection tools used

Step	*Information needed*	*Tools for collecting the required information*
1	Types, amounts and developments of state aid to SMEs	– Literature research (EU publications), internet search (State Aid Scoreboard)
2	Classification of state aid types	
3	Description of applied evaluation methods	– Literature research (evaluation reports) – Questionnaire to Member States – Member State Meeting
4	Description of alternative evaluation methods	– Literature research (evaluation literature) – Literature research (evaluation reports of state aid for other objectives, not SMEs) – Consult two evaluation experts
5	Criteria for assessing the selected evaluation methods	– Literature research (evaluation literature) – Consult two evaluation experts
6	Reporting	
7	Assessment of selected methods	– Expert opinions
8	Reporting	

Literature study

The research project started with a study of the existing literature on state aid, followed by a review of the literature on evaluation. An overview of the publications consulted can be found in the bibliography.

The questionnaire

A questionnaire was prepared based on the insights gained from the literature review. The questionnaire, used to identify good practices in all the Member States, was sent by the Commission to the Member States. In total, 21 questionnaires, covering 19 countries, were returned[4].

The Member State meeting

On 25 May 2004, the Commission organised a meeting with the Member States. This meeting was intended to involve the Member States in the study and to facilitate the collection of information about the evaluation methods used by the Member States. The meeting was held at the Commission premises in Brussels and was attended by delegations from 23 countries.

Consulting two evaluation experts

For steps 4 and 5 of the research, two experts in the field of policy evaluation have been consulted. One expert, Professor Peter van Hoesel, is Professor of Public Administration at the Erasmus University of Rotterdam and director of EIM Group. He is an expert in the field of SME policies and evaluation methods.

The second expert, Professor Arthur Ringeling, is also Professor of Public Administration at the Erasmus University. Professor Ringeling has written or edited more than 15 books, more than 100 articles

[4] In Belgium and Italy, the returned questionnaires applied to regions.

and several reports on subjects such as instruments of public policy, policy theories, policy evaluation, administrative organization and culture and public decision-making.

1.3.3 *Data analysis*

All the collected materials were analysed for their merits, using mainly qualitative techniques (such as *frequency counts* and *document analysis*). The analysed data are described and presented in the remainder of this report. This ultimately results in an overview of 8 'basic' methods for evaluating state aid effectiveness.

1.3.4 *Judgement phase*

After approval of the data analysis phase of this study, the selected methods were judged, to find out which of the methods is / are most suited for evaluating the effectiveness of state aid to SMEs. This judgement is based on a multi-criteria analysis. Seven renowned international evaluation experts in the field of SMEs (scientists with practical evaluation experience and Commission experts) delivered their judgement as to how each of the 8 'basic' methods performs in relation to the relevant criteria. The following experts have cooperated:

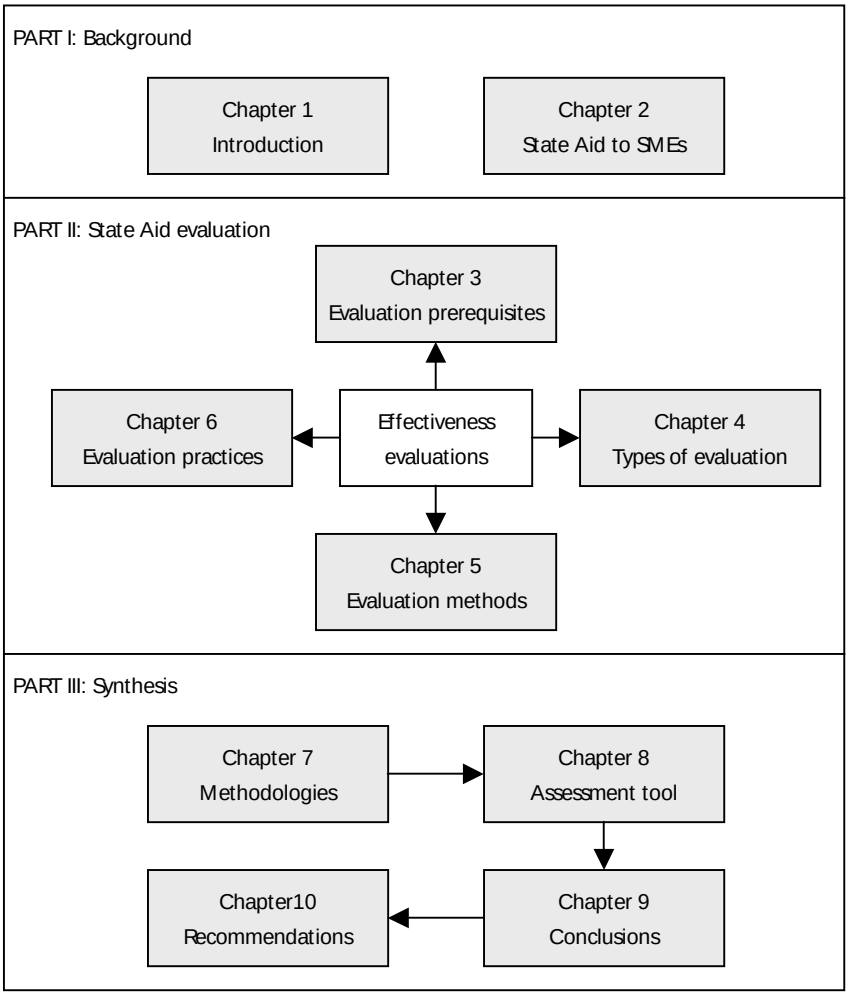
- Prof. David Audretsch — Indiana University, USA
- Prof. Roy Thurik — Erasmus University Rotterdam, the Netherlands
- Prof. Lois Stevenson — University of Ottawa, Canada
- Prof. Bruce Kirchhoff — New Jersey Institute of Technology, USA
- Prof. Johan Lambrecht — European Institute of Higher Education, Belgium
- Miss. Veronica Gaffey — Evaluation expert form the European Commission, Directorate-General for Regional policy, Belgium
- Dr. Yvonne Prince — Director of EIM Business & Policy research, the Netherlands

After the assessment of the 8 'basic' methods, this report was drafted and conclusions and recommendations were drawn.

1.4 Contents of this report

In this report, three topics will be discussed: see figure 3. The background, i.e. the policy environment, state aid situation and objectives for this study (part I); an overview of state aid evaluation, based on theoretical and practical literature and field research (part II); and a synthesis of the evaluation methods with conclusions and recommendations on which methods are best for evaluating the effectiveness of state aid to SMEs (part III).

Figure 2 Structure of the report


PART I: Background
Chapter 1
Introduction
Chapter 2
State Aid to SMEs
PART II: State Aid evaluation
Chapter 3
Evaluation prerequisites
Chapter 6
Evaluation practices
Effectiveness
evaluations
Chapter 4
Types of evaluation
Chapter 5
Evaluation methods
PART III: Synthesis
Chapter 7
Methodologies
Chapter 8
Assessment tool
Chapter10
Recommendations
Chapter 9
Conclusions

2 State Aid to SMEs

Due to their relatively high vulnerability to various market failures, the development of small and medium sized enterprises lag behind compared to larger corporations. The most prominent market failures are capital market imperfection and asymmetric information. One way in which Member States respond to these market failures is by providing state aid to SMEs. State aid to SMEs represents a small, but growing fraction of the total amount of state aid. The most important form of aid to SMEs is 'grants', which makes up approximately four fifths of the state aid to SMEs.

2.1 Background of State Aid to SMEs

The vital importance of SMEs

Small and medium-sized enterprises are of vital importance for the European economy. SMEs stimulate innovation and competition since most of them are flexible, close to their customers and exploit market niches. As a result SMEs often are particularly innovative and contribute to the creation of new employment.

However, due to their size, SMEs often face difficulties in access to finance, information and knowledge. These difficulties may be caused by market failures, of which the most prominent ones are capital market imperfections, including difficult access to finance especially in the creation and start-up phase, and asymmetric information.

State aid to SMEs

One of the instruments used by Member States to address the difficulties faced by SMEs is state aid[5]. In practice, the state aid programmes have very different policy objectives. These policy objectives range from increasing the innovativeness of the SME sector by disseminating innovative technologies among SMEs, through providing financial assistance in order to increase the number of start-ups.

Only state aid that is earmarked exclusively for SMEs (as primary or secondary objective) is considered in the report. Other types of assistance (e.g. state aid earmarked for R&D, but rendered to an SME) are not taken into account.

State aid to SMEs, like all other types of state aid, is subject to control by the European Commission. The aim of state aid control is to ensure that the competition in the internal market is not distorted contrary to the common interest, while at the same time allowing for state interventions that may tackle identified market failures.

2.2 Amounts and development of state aid

The amount of state aid to SMEs

According to the State Aid Scoreboard, in 2002 the 15 EU Member States spent € 48,753 million on state aid[6] (Commission of the European Communities, 2004). This amount represented 0.56% of their

[5] This report focuses on state aid as defined by article 87 (1) of the EC Treaty. General measures not falling with the definition of state aid are not dealt with in this report.

[6] Total State Aid, excluding railways.

Gross Domestic Product (GDP). State aid to SMEs as a primary objective in that same year amounted to € 4,857 million. In addition, some € 920 million was spent on state aid where SME support was the secondary objective[7] (Commission of the European Communities, 2003a). These amounts show only aid that was exclusively targeted at SMEs and that fall under the examination of the European Commission. Other types of state aid (such as state aid that was not specifically targeted at SMEs, or *de minimis* aid) may also have reached SMEs, but are not included in these figures[8].

Development of state aid

The Stockholm and Barcelona European Councils called on Member States to continue to reduce state aid as a percentage of GDP while redirecting it towards more horizontal objectives of common interest, including SMEs. The total amount of state aid (excluding aid to railways) was reduced from € 70,446 million (1.09% of GDP) in 1992, to € 48,753 million (0.56% of GDP) in 2002. In that same period, total aid excluding railways, agriculture, fisheries and transport decreased from a level of € 54,389 (0.85% of GDP) to € 34,005 million (0.39% of GDP).

Not only is the amount of state aid being reduced, it is also being redirected from ad-hoc and sectoral support towards horizontal objectives of common interest. While in 1992, horizontal objectives had accounted for 58% of the total state aid (excluding aid to railways, agriculture, fisheries and transport), by 2002 this fraction had grown to 73%.

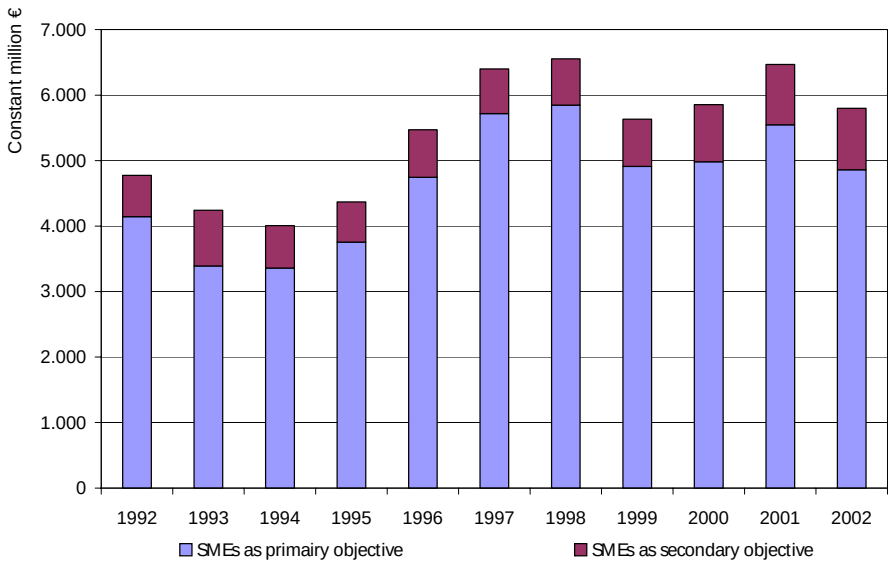
Development of state aid to SMEs

With regard to state aid to SMEs, the trend is somewhat different. Whereas the total amount of state aid is declining, there is no convincing evidence that the annual amount of state aid to SMEs is decreasing. This may be due to the fact that part of the reductions of sector support has been redirected towards horizontal objectives (among which aid to SMEs). As a result, no univocal downward trend of state aid to SMEs can be distinguished. The development of the amount of state aid to SMEs in the EU is shown in figure 4.

Figure 3 Development of state aid to SMEs in the EU, by objective

[7] I.e. 100% of this aid is earmarked for SMEs.

[8] For example, the National Court of Audit calculated that the Dutch Ministry of Economic Affairs run assistance measures for SMEs and entrepreneurship with a 'financial importance' of an estimated € 2.95 billion between 1995 and 1999, while only € 112.4 million was notified to the Commission as state aid. This indicates that state aid forms only a small fraction of the total public support for the private sector. Sources: Algemene Rekenkamer (2001, p. 12) and Commission of the European Communities (2001, p. 134).



Source: EIM, based on data of the European Commission website

State aid instruments

Through time, a trend has become visible in the pallet of instruments that Member States use to disseminate state aid to SMEs. The single most important instrument is *grants*. During the last decade grants have become more important. In 1992, 68.0% of the total amount of state aid had been distributed through grants. In 2002 this increased to 82.5%.

The second most important instrument is *tax exemptions*. Tax exemptions were a popular instrument in the 1990s, but have shown a strong decline since 2000.

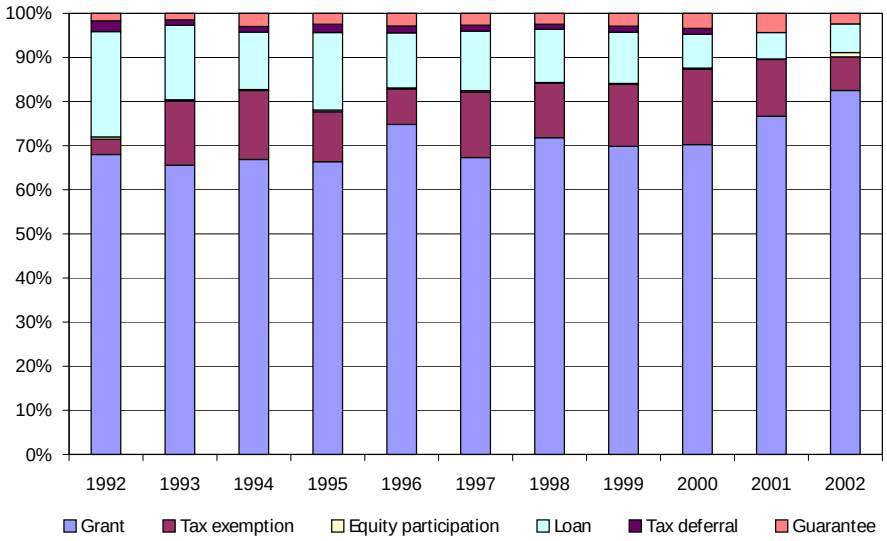
The third most important instrument is *loans*. The relative importance of loans also decreased during the last decade. While loans represented 23.9% of the total amount of primary SME state aid in 1992[9], their relative share was reduced to 6.5% in 2002.

The fourth instrument is *guarantees*. Guarantees represented 1.7% of the total amount of state aid in 1992, but increased to 2.4% in 2002. However, the relative importance of guarantees had built up to 4.4% of the total amount of state aid for SMEs as the primary objective one year earlier (in 2001). The fallback to 2.4% in 2002 might be temporary.

Other forms of state aid to SMEs, such as *equity participations* and *tax deferrals*, played relatively a marginal role.

Figure 4 Development of the relative importance of aid instruments (in % of total annual state aid to SMEs)

[9] As measured in 'cash grant equivalents'.



Source: EIM, based on data of the European Commission website

PART II:

STATE AID EVALUATION

3 State Aid effectiveness studies

Evaluation is the systematic investigation of the effectiveness of economic intervention whereby the performance of support measures is judged against pre-specified standards or criteria. There are many definitions of effectiveness. For the purpose of this study, effectiveness includes not only the extent to which a policy objective has been achieved, but also the (positive and negative) side effects, additionality and displacement. It is important that evaluation studies attach importance to the causality of the relation between the aid measure and the observed effects (e.g. the economic gain).

3.1 Introduction

Since the objective of this study is to identify a method that best measures the effectiveness of state aid to SMEs, the remainder of this report will focus on the evaluation of *effectiveness* of SME support schemes (in contrast to *efficiency* evaluation, which is not part of the current project). This chapter will introduce the concept of effectiveness. Since the report will be made available to a wide public, we will illustrate some of the key concepts with a fictive example.

Example

Throughout this report, we will illustrate the key concepts by using an example. This example comprises a fictitious aid scheme that aims to reduce market failure with regard to innovation in SMEs. Since innovation can improve productivity, it has proved to be an important factor for economic growth. The market failure involves the fact that SMEs may under-invest in innovation processes, since the burden of the investment risk is too great for SMEs. The (intermediate) policy objective and the accompanying state aid scheme of this support measure could be the following:

Policy objective: to achieve economic growth through improving innovation in the SME sector.

Aid scheme: a closed-ended grant scheme, whereby the government refunds 50% of the costs of new R&D projects in small enterprises.

When is an aid scheme effective?

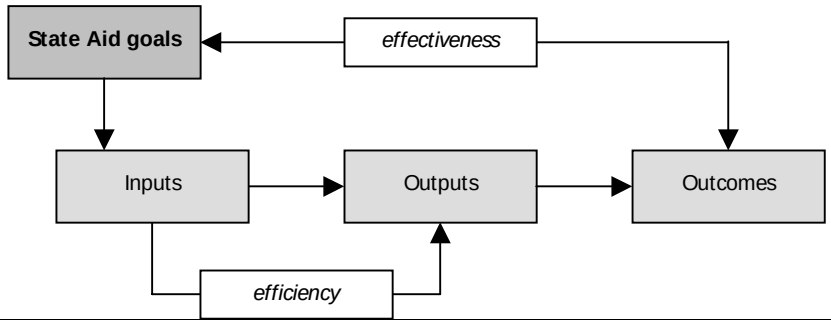
There is not one single, generally accepted definition of effectiveness. Besides, measuring the effectiveness of state aid is difficult for a number of reasons. As we have seen, the only acceptable rationale for state aid to SMEs is that it reduces the market failures that SMEs face. Ideally, evaluation of the effectiveness of state aid should identify whether, and to what extent, the state aid scheme has been able to reduce the targeted market failure. This approach involves quantifying both the market failure and the influence of the aid scheme on the market failure. However, it is quite impossible to quantify a market failure, let alone the impact of a policy measure on the extent of that market failure. Therefore, aid policies use intermediate objectives.

Often, state aid serves several intermediate objectives. According to Professor David Storey, evaluation of effectiveness is impossible unless objectives are clear and measurable. Too often, objectives are either not specified or specified in a way that is vague and incapable of being used as the basis for deciding whether or not the policies are successful. State aid objectives should be quantified and become explicit targets (Storey, 2000). In order to secure that objectives are transformed into quantified targets, evaluators should be involved in the earliest stages of the policy process, that is during the design phase of the policy (OECD, 2004). Measuring the effectiveness of a scheme requires *clear*, *explicit* and *quantified* objectives.

What is effectiveness?

Another question can be raised, namely: what exactly is 'effectiveness'? The simplest definition of effectiveness would be the percentage to which the objectives of the state aid scheme have been reached. As shown in figure 6, effectiveness relates the project's outcomes (= effects) to the state aid goals[10].

Figure 5 Various aspects of a support policy



Source: EIM 2004

Inputs

Inputs are the resources that are provided by the state aid programme. For example a monetary amount, a tax deferral, or human capital. For the purpose of evaluation, inputs are commonly calculated as the 'cash grant equivalent' of the provided resources.

Outputs

Outputs are the direct effects that the state aid provision results in. For example increased production capacity, higher company turnover, or a better-educated workforce.

Efficiency

Efficiency is the term commonly used for relating outputs to inputs. A process is 'efficient' if it requires relatively little inputs to produce a certain amount of output. If a process produces relatively little outputs, compared to the inputs, it is said to be 'inefficient'.

Outcomes

The outcome of a process is 'something that follows as a result or a consequence'[11] from the outputs. In the case of a state aid scheme this is in the sphere of an achieved state of being. Outputs could be: increased competitiveness, or economic growth.

10 The Dutch government uses the SMART-criteria for policy objectives. Policies should be **S**pecific, **M**easurable (quantified), **A**cceptable for the users, **R**ealistic and **T**ime-limited: SMART (Ministry of Economic Affairs, 2002).

11 Webster online dictionary (www.webster.com).

Effectiveness

Effectiveness is the term commonly used to refer to the goal-attainment of a measure, thus relating the outcome of a process to its original goals. A state aid measure is said to be effective if the goals are reached, i.e. if the outcomes match with the goals.

An example: Economic growth depends to a large extent on the creation and growth of small enterprises. However, due to a number of reasons (such as risk-averse investors, monitoring problems, hold-up problems, adverse selection, information asymmetries, etc.), there is a sub-optimal level of financial capital available for SME development. If the ultimate objective of a support measure is to reduce this market failure through providing guarantees (inputs), an intermediate objective could be to provide risk capital to start-ups (intended output), in order to increase SME development (intended outcome).

However, the effects of a certain measure are difficult to quantify, let alone measure. For example: how to measure 'innovation', 'skills' or 'access to finance'? More over, it is often very difficult to demonstrate the causal relationship between an aid scheme and the measured changes (presumed effects). This may result in *deadweight factors* (changes that would have occurred anyway because of other causal or contextual factors, regardless of the aid scheme). Also, *additionality* (which additional investments of the SME sector itself did the support measure induce?), *displacement* (relocation of the unwanted obstacles of firm development to other firms), *substitution* (reaching the desired effect at the expense of non-targeted enterprises) and *cross-substitution* (for example, subsidies for new employment are used for investments in production facilities[12]) may occur. These issues will be discussed below, in more detail.

Evaluation should look beyond effects and *at processes*

According to Professor Ringeling, an evaluation study is successful if it contributes to the public discussion about state aid effectiveness. Therefore, not only the evaluation outcome, but also the evaluation process and additional information is of utmost importance. Evaluation should not only answer the question whether the goals have been obtained, but it should also bring together policy makers and the other actors involved, such as the entrepreneurs, their sector organisations, communities of interest, etc. Evaluation should look beyond the effects of a certain scheme, and should give due attention to the network of actors involved in the policy measure and to the processes that invoke these effects.

Moreover, in the covering letter sent with their contribution to this study, the representatives from the Norwegian Royal Ministry of Trade and Industry made the following remark: "*… the Norwegian authorities are interested in identifying the state aid tools and state aid guidelines that are the most effective in achieving SME policy objectives. We believe that this can best be done by gaining better knowledge and know-how about how the specific state aid measures actually work in fostering enterprise creation, removing barriers to entry for newly created firms providing venture capital and seed capital and increasing SME employment and turnover.*" This remark is important since it focuses not only on the effectiveness of the *output* of state aid (increase in jobs, innovations, start-ups, etc), but also on the effectiveness of the *processes* which state aid initiates (the changes in mechanisms that produced the outputs, the durability of the outputs, etc.).

12 This effect was identified in Kangasharju and Venetoklis (2002).

3.2 What exactly is an evaluation study?

Like any other study, an evaluation is research. Research is carried out in order to answer a clearly defined *research question*. This research question is the single most important factor influencing the design of the evaluation study. Examples of research questions from past state aid evaluations are:

1 Whether the scheme was meeting its objectives and whether there are alternative ways of meeting these objectives (Arup Economics & Planning, 2000)
2 To what extent does the measure lead to more R&D-activities (first order effect) at enterprises? What is the influence of company size, sector, technology, R&D project classification, and type of user on the innovativeness (second order effect) and economic performance (third order effect) for supported and unsupported companies? (Brouwer et al., 2002)
3 Would the subsidised jobs have been created even without subsidies? (Kangasharju and Venetoklis, 2002)

In his recent paper on SME policy evaluation, Professor David Storey of Warwick University argues that evaluation should be an integral part of the policy process (OECD, 2004). Policy makers should use the COTE framework, which represents policy ***C****larity and Coherence*, clearly stated policy ***O****bjectives*, quantified policy ***T****argets*, and ***E****valuation* of the policy in terms of its targets.

According to Prince (1998), an evaluation study should meet at least the following criteria: it is analytical, systematic, objective, reliable, reproducible and user-driven. For the purpose of this study, we define evaluation as follows: ***Evaluation is the systematic investigation of the effectiveness of economic intervention whereby the performance of support measures is judged against pre-specified standards or criteria****.*

Criteria

The standards, or criteria, against which the performance of the support measure is judged, must be identified and quantified before the evaluation process starts (see Bressers and Hoogerwerf (1995), Bennet (2003, p. 68), Rossi et al. (1999)). Ideally, the success criteria have already been formulated in the design phase of the support measure. However, for various reasons, this is often not the case. Intended objectives of support programmes are often multiple, vague, hidden, evolving and sometimes even conflicting (OECD 1999, p. 22).

A report from the Dutch National Court of Audit shows the relevance of clear, measurable and quantified criteria. The report concludes that, over six years after the issuing of the bill '*Werk door ondernemen*' ('*Work through enterprise*') only little is known about the results of the 28 policy measures aimed at the SME sector that were announced by the bill (Algemene Rekenkamer, 2001). The lack of insight in the performance and resulting effects is the consequence of the fact that for many of the policy measures, no clear objective was formulated. Also, many measures have never been evaluated. As a result, it remains unclear whether, and to what extent, the measures have really contributed to the original objectives of the policy: *to stimulate the start-up, growth and expansion of small and medium-sized enterprises and to realise further employment growth*. With the policy measures, a total amount of € 1.25 billion was involved.

In annex no. 4 to its report, the Court of Audit formulated 5 norms that have been applied to the policy objectives (Algemene Rekenkamer, 2001):

1 Explicit : The objective has to be recorded in parliamentary bills.
2 Complete : (a) The objectives have to be formulated in terms of performance (output) and in terms of intended welfare effects (outcome), and
(b) the target group of the policy has to be described, fenced in and motivated.
3 Timely : The objectives have to be formulated *prior to* implementation of the policy measure.
4 Testable : The objectives have to be formulated in a testable manner, where

(a) the intended situation (i.e. if the goals be reached) is described clearly,
(b) concepts are measurable and formulated univocally, and
(c) the time frame within which the goals have to be reached is made explicit.

5 Consistent : (a) *Consistency among the objectives*: objectives must be compatible, objectives of policy components should connect to the overall policy objective, and intermediate objectives must correspond to the final objectives.
(b) *Consistency between objectives and policy information*: the formulation of the objectives must connect to the socio-economic problem analysis, which is based on adequate policy information.

The criteria formulated in the policy objective form the basis for the research question and the design of the evaluation study. These criteria should *not* be formulated in terms of output (performance indicators of the scheme itself, such as number of subsidies or quality of the information services), but should be formulated in terms of outcomes (intended effects, such as number of jobs created, number of new start-ups incubated, or the amount of value added by the supported firms, all as a result of the support measure). However, often the policy objectives are unclear with respect to the intended effects. In such cases, the policy objectives have to be *reconstructed*, i.e. the evaluator has to do research prior to the evaluation in order to rediscover the objectives, goals and intended effects after the policy measure has been implemented (Bulder, 2004).

Example

Success criteria of the R&D funding scheme are the following:
- At least 600 new jobs in R&D (first order effect)
- Increase profitability of the participants with 15% (second order effect)
- Increase turnover from innovative products in the total SME sector by 50% (third order effect)

However, ultimately an evaluation intends not only to judge whether or not the goals of the support scheme have been reached, but also to judge the merit, worth, and value of the entire support scheme. This does not only include intended effects, but also positive and negative side effects (Vedung 1997, p. 59). According to the supporters of the goal-free evaluation approach, pre-occupation with the objectives of the support scheme hampers the identification of such side effects. In this report we stress the importance of incorporating both intended effects (such as additionality) and unintentional effects (such as displacement).

3.3 Causality

An important issue when evaluating policies in general is the causality. Causality implies that there is a causal relationship between two phenomena: the cause and the effect. Without the cause, the effect would not have occurred. In order to demonstrate a causal relationship, three prerequisites have to be fulfilled at the same time:
- *concomitant variation* (the extent to which X and Y occur at the same time, or vary in the expected way)
- *time order of occurrence of variables* (effect Y has to occur parallel to or after cause X, but not prior to X)
- *elimination of other possible causal factors*.

3.4 Deadweight

Deadweight, or counterfactual, has to do with the question ‘what would have happened within the company if the support had not been provided’. How can we be sure that an increase in R&D jobs is

the result of an aid measure and not other contextual factors? Would these jobs also have been created if the scheme had not existed?

Venetoklis (2000, p. 15) demonstrates the importance of the use of deadweight. The use of deadweight as a benchmark for the results provides more balanced judgements than studies that do not use deadweight, or where the judgement of s measure's effectiveness was based on estimates from the recipient firms.

Traditionally, deadweight has been researched using a control group of non-beneficiaries (see for instance Turok, 1991). However, this approach has some drawbacks, one of which is the selection problem. The problem is that there may be structural differences between supported firms and unsupported firms (selection bias). Selection bias may occur as a result of firm self-selection (i.e. the firms that apply for the support are not representative for the total population), or as a result of agency selection (i.e. the agency accepts only the applications that meet the selection criteria) (Wren and Storey, 2002). The structural differences may bias a sound comparison between the research population and the control group[13]. There are statistical procedures to test for selection bias, for instance the Heckman 2-step adjustment procedure[14].

Commissioned by Enterprise Ireland (a public organisation for the economic development of Ireland), Lenihan et al. (2003) conducted an alternative research approach on this topic, using a self-assessment tool. This tool makes use of face-to-face interviews with owner-managers of subsidised SMEs, making a control group superfluous. The authors developed a method to measure the total deadweight in terms of employment, incorporating both 'pure deadweight' (project would have commenced unchanged, even without support) and 'partial deadweight' (project would have commenced, but in a different form) categories. They do this by grouping together the firms that experienced similar developments and comparing the number of jobs created. In this way, they assign the proportion of jobs created to the support scheme. The total deadweight of the state aid (= 'pure' and 'partial' deadweight) was 46.2%.

Based on their research, the authors concluded that the 'pure' deadweight was 19% (i.e. 19% of the projects would have commenced unchanged, even without Enterprise Ireland support). The authors asked the beneficiaries of Enterprise Ireland assistance whether, in the absence of aid, their project would have:

1 Gone ahead as now unchanged, that is, same scale, time and location (i.e. pure deadweight)
2 Gone ahead at a later date – delayed the project (i.e. partial deadweight)
3 Gone ahead but on a reduced scale – changed the nature of the project – removed certain features (i.e. partial deadweight)
4 Combination of 2 and 3 (i.e. partial deadweight)
5 Abandoned the project (including going ahead at a location outside Ireland) (zero deadweight).

The main drawback of this approach is that it is subjective to, and susceptible for, strategic or socially desirable answers.

13 For example: if a support measure is provided to fast growing firms only, and the control group also contains slow and medium growing firms, then the analysis of the impact of the support on firm growth is influenced by a selection bias.

14 The Heckman 2-step adjustment procedure formulates a single equation to explain the selection procedure and then formulates a second equation that explains performance change on the basis of the factors included in the selection equation (also see: Storey, 2002). For a more detailed explanation see annex III.

Example

In our example, recipients may receive a subsidy worth up to 50% of the total project costs. However, it could be that the project would have commenced anyway, even without the subsidy. In that case, the support measure substitutes private investments by 100%. An unintended side effect of the scheme could be that the money saved is then invested in a promotion campaign, production facilities or new housing for the company and should be subtracted from the deadweight effect. Such a side effect was measured by Bager-Sjörgen (2003).

3.5 Additionality

Additionality is defined as 'whether a support measure induces private investments that would otherwise not have been made' (Brouwer et al. 2002). However, the scope of the definition varies. For instance, Brofoss et all (2004) limit the scope of additionality to an enterprise's capabilities in a certain area that can be deployed in the market. Others also include knowledge spillovers to the business sector (broad dissemination of public knowledge in the private sector). Capabilities and spillovers are especially relevant in the sphere of R&D support and funding for (private) SME consultancy services (Brofoss et al. 2004). For the purpose of evaluation third order effects of state aid to SMEs, the appropriate scope might be to also include additional welfare, which would not have occurred without government intervention (Licht, 2003).

In practice, additionality has been measured particularly in the sphere of R&D subsidies. Evaluation studies that include state aid additionality have addressed three appearances of this phenomenon:

1 The question of input additionality; did firms spend more of their own resources on the intended activities because they were subsidised? (A formal model for calculating input additionality is presented by Lach and Sauer (2001))
2 The question of output additionality; did the activities' outputs increase due to the support scheme? (for example the number of innovations, patents, jobs, firm start-ups, etc.) (A formal model for calculating output additionality in relation to displacement is presented by Meeusen and Janssens (2000))
3 The question of behavioural additionality; permanent changes in firm behaviour, inducing a more efficient transformation of inputs into outputs.

Measuring additionality is difficult, but not impossible. Brouwer et al. (2002) used a questionnaire combined with econometric analyses to measure input additionality. They analysed whether or not the investments in R&D increased more than the value of the R&D subsidy that the recipient firm obtained. Brofoss et al. (2004) measured output additionality through the use of a questionnaire, but approached the matter in a more qualitative way. They included questions such as 'how important was the subsidy for getting the project started at all?'.
Clarysse et al. (2004) have developed and tested a questionnaire tool for measuring input, output and behavioural additionality of R&D grants. Their conclusion is that for analysing additionality of SME aid, a standardised telephone[15] survey questionnaire can be used (with a few customised parts for special groups within the target population).
An alternative way of measuring input, output and behavioural additionality is through the use of control groups.

15 In such a complex domain as R&D subsidies, the interviewers can process information in a more homogeneous way than the interviewees can. Therefore the survey should be conducted in face-to-face interviews or as a telephone survey. The authors strongly discourage future researchers from doing postal surveys with their questionnaire.

Example

Within the scope of our example of R&D subsidies, input additionality could be defined as the additional investments that the recipient companies make, that they would not have made if the support measure has not existed. For example: a certain recipient would not have commenced with a certain R&D project at all without the subsidy. Let's presume that the total R&D project costs € 50.000, of which 50% is subsidised. Then the input additionality is € 25.000, or 100%.
Output additionality is more difficult to measure. What are the outputs? Jobs, innovations, new products? This depends on the objective of the support scheme. In our example, output additionality would be achieved if the support resulted in an extra increase in the number of innovations that would not have been marketed if no feasibility subsidy had been awarded.
Behavioural additionality could be achieved if the support scheme induced the recipients to adjust their R&D processes or improved the recipients' attitudes towards innovation, and that these adjustments also increased the efficiency of other processes. For example if they performed feasibility studies in the future too (without the use of a subsidy) or if they set up structural cooperation with other companies.

3.6 Displacement and (cross) substitution

Displacement occurs when public support crowds out or replaces private investments. Displacement is a negative effect of state aid, which (partly) annuls the state aid effects. Therefore, effectiveness measurements must stake account of displacement.
Displacement can occur *within* a company (e.g. a company receives a subsidy for making an environmentally friendly investment, but the company would have made the same investment without the subsidy anyway). But displacement can also occur *among* companies (e.g.: subsidising the R&D activities of one firm may result in the abortion of R&D activities by one of its competitors).

The topic of displacement is difficult to investigate, but was also researched by Lenihan, Hart and Roper (2003). The authors calculated the displacement using the fraction of the assisted firms' sales sold on the domestic market and the fraction of sales that competes with other national suppliers. They multiplied these two figures and arrived at a total displacement of 8.1%. Although disputable from a method point of view, this approach is simple and transparent.

Example

In our example, subsidies to company A could displace private investments by company B. Suppose that two small ice-cream companies produce for the same market. Company B intends to develop an innovative new way of producing ice cream using only private money for the investment. To earn back the investment, B's sales have to grow by at least 25%.
However, company A receives a subsidy to improve its production facilities. The development will make the production process more efficient, so that the company can produce 25% more ice cream. Assuming there is not enough room in the market for both companies to grow, if company A receives the subsidy, company B will not be able to make the investment. In this case, the subsidy displaces private investments made by another company.

3.7 Other side effects

Besides negative side effects, state aid may also have positive side effects. Two of the most important side effects that deserve to be mentioned in this report are knowledge spillovers and increasing de-

mand. Knowledge spillovers occur in the sphere of R&D support and training support. Examples are the dissemination of knowledge through cooperation or through job changes of employees. In this way, other companies may also benefit from knowledge that has been created with public financing. Increased demand appears when the state aid measure leads to growth of SMEs, that subcontract more work to suppliers. In that case there is a multiplier effect resulting in improved economic benefits from the aid.

3.8 Conclusion

A narrow definition of effectiveness is: the extent to which the goals of a certain policy measure have been achieved. A broad definition would involve an assessment of all the effects – both intended and unintended – of the policy measure. In practice, definitions are adopted that lie between these extremes.

Recent literature often curtails the definition of effectiveness by including deadweight, additionality and displacement (and substitution), but often fencing out other side effects. With regard to our study, effectiveness will be defined in the broadest sense, namely the integral assessment of all the effects – both intentional and unintentional – of the support measure, including deadweight, additionality, displacement (and substitution).

4 Types of evaluation studies

There are many classifications of evaluation studies, each of which classifies many different evaluation approaches. The methodological soundness of these evaluation approaches is beyond dispute, but it is important to note that different evaluation methods are suited for different applications. Different evaluation methods should be employed for each stage in the policy cycle, and also different evaluation methods should be employed in different policy fields. Economic policy (within which state aid to SMEs falls) needs economic and econometric evaluation methods.

4.1 Introduction

State aid intends to realise a certain policy objective. When designing an aid scheme, policy makers choose an aid type (grant, tax deduction, soft loan, etc.). This aid type is governed by certain criteria (e.g. only firms with less than 50 employees may apply), restrictions (e.g. maximum amount of aid is € 5,000 per firm per annum) and procedural rules (e.g. an application form that should be completed and submitted in triplicate, together with authentic signed statements from the firm's bank, lawyer and accountant). Together these make up the state aid scheme.

The large monetary amounts, the vast diversity of policy objectives and the variety of instruments associated with state aid bring about uncertainty with regard to the question which aid scheme design is most effective in reaching the desired effects. Researching the effectiveness of an existing aid scheme serves two purposes: (1) accountability to the principal (financer) of the scheme, and (2) the identification of inefficiencies so that the scheme can be improved (see also: Venetoklis, 2002, p. 5).

Put differently, there is a need for systematic evaluation of the effectiveness of the various state aid policies and their respective instruments. In various Member States[16], state aid programme evaluations have provided valuable insights for improving existing programmes and have identified learning items for new support programmes. In this sense, the cost of evaluation studies can be re-earned through gains in efficiency and effectiveness.

An evaluation of the effectiveness of state aid intends to monitor whether the aid scheme is designed in such a way that the policy objective is achieved effectively[17]. Thus, an evaluation can result in adjusting the aid type, criteria, restrictions or procedures captured in the aid scheme design, if it concludes that there is room to improve the effectiveness.

4.2 Classifications of evaluation approaches

In evaluation literature there are many classifications of effectiveness evaluation studies. For example:

– Classification according to the *phases in the policy process*: fixing the agenda, problem diagnosis, policy design, decision-making, implementation, and the carrying-out of the scheme (note that with professional organisations, the policy process is a cycle that starts again, and where evaluation is a continuous activity);

16 The results of the survey show that this concerns at least Austria, Belgium, Czech Republic, Finland, Ireland, the Netherlands, Norway, Portugal, Slovak Republic and the United Kingdom

17 Note that evaluations of the effectiveness of an aid scheme do in principle not question the policy *objectives*, which are often outcomes of political negotiations. However, Storey argues that evaluations can lead to changes in the policy's *targets* (OECD, 2004).

- Classification according to *the moment at which the evaluation takes place*: Ex-ante, ex-nunc and ex-post evaluations;
- Classification according to the *level of sophistication of the method and the type of results*: from descriptive through highly analytic;
- Classification according to *the frequency of data collection*: *constant* (monitoring), periodic, or ad hoc;
- Classification according to *the evaluation theories and principles applied*: goal-attainment, goal-free, stakeholder approaches, etc.;
- Classification according to *the type of analysis techniques employed*: ranging from purely qualitative through entirely quantitative (note that this can be a continuum);
- Classification according to *the data sources*: direct impact measurements at the target group (i.e. SMEs), versus perception measurements at stakeholders;
- Classification according to *the measured effects*: evaluation investigating zero-order effects (has the state aid activity really been realized?), first-order effects (has the stimulated/financed activity really led to an implementation), second-order effects (has the implementation led to more innovative products, more employment, more exports, etc.), and finally third-order effects (what has been the contribution of one Euro state aid to the competitiveness of a Member State?)[18];
- Etc.

Some of the issues raised in the different classifications will be discussed below.

4.2.1 *Phases in the policy process*

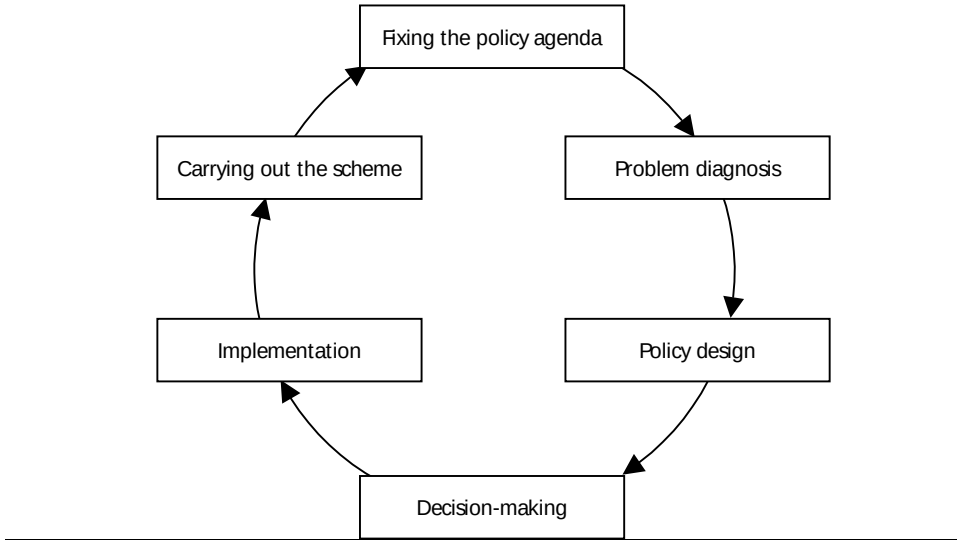
Modern literature treats the process of policy formation and implementation as a cycle: problems are identified and studied, solutions designed and implemented, effects are evaluated and the policy is adjusted. The policy cycle consists of six steps:

1	Fixing the agenda:	social problems (i.e. market failures) are identified
2	Problem diagnosis:	the nature and extent of the problem is assessed
3	Policy design:	one or more policies for solving the problem are designed (non-intervention should be one of these policies)
4	Decision-making:	the best approach is chosen
5	Implementation:	a department or an agency is assigned or set up, that will execute the policy
6	Carrying out the scheme:	the policy instrument (i.e. the state aid measure) is carried out

Based on evaluation of the policy process, the cycle starts again: what issues have not been addressed (yet), and how can we improve the problem solving. The policy cycle can be depicted as shown in figure 7 below:

Figure 6 The policy cycle

18 For a more detailed elaboration see: MEET (2000), Prince (2000) and Muizer and Prince (2000)



Source: EIM, 2004

The above diagram clearly indicates that there are many steps between the identification of a socio-economic problem and the implementation of a support measure. According to Professor Peter van Hoesel these intermediate steps should also be part of an evaluation of the effectiveness of a scheme, that is if the objective of the evaluation is to learn what can be improved. An assessment of the effects of a scheme may identify that the scheme fails to reach certain objectives, but it does not yet make clear *why* a scheme fails at a certain point. Therefore, an evaluation should also look at what choices were made, why they were made and the processes with which they were implemented.

Professor Peter van Hoesel provided an overview of research methods that facilitate the policy process (see table 8 in annex II). The methods mentioned are those most frequently applied (and published) in the various stages of the policy process[19].

4.2.2 *Ex-ante versus ex-post evaluations*

Evaluations can be carried out before (ex-ante), during (ex-nunc) or after the completion (ex-post) of an economic intervention programme. Ex-ante evaluations can be part of the preparation of an intervention and can measure the estimated or expected impact of an intervention beforehand. Ex-ante evaluations often are unformalised, iterative processes that contribute to the fine-tuning of programme specifications. Ex-nunc, or interim evaluations examine the ongoing activities of a programme in order to give relevant feedback on its effectiveness. Referring to an ex-nunc evaluation, a programme administrator could make adjustments to the programme during its duration. Ex-post evaluations assess the effectiveness and efficiency of a programme in hindsight. Because economic impacts usually need some time to materialise, ex-post evaluations usually take place some time after the completion of the programme. Ex-post evaluations should be used to learn from "what went right and what went wrong" during a specific programme, if they are to be more effective and/or more efficient intervention programmes in the future.

19 In annex III, the 'X's indicate which of the methods and data collection techniques can be applied for ex-ante, ex-nunc, or ex-post effectiveness evaluation.

Maes and Sels (1999) analyse the type of effects that can be measured by ex-nunc, direct ex-post and long-term ex-post evaluations of company training programmes. Ex-nunc evaluation can measure the participants' (employees') reactions to, and knowledge gained from, the training. One reaction is, for example, participants' satisfaction about the training. However, although satisfied course members are important for stimulating taking part in future training, "there is no point in measuring the level of entertainment". Learning concerns employees' attitudes, knowledge and skills. Testing the attitudes, knowledge and skills (e.g. by the use of exams) can be useful for measuring whether the training is effective in transmitting the course material to the members. The same types of evaluations can be applied with regard to certain non-financial aids to SMEs (testing whether the transfer of knowledge to SMEs is effective).

Direct ex-post evaluation is an evaluation immediately after the support is provided. This type of evaluation can measure changes in course members' behaviour that are a result of the fact that the members followed the course. This is useful, since this type of evaluation measures whether the skills learnt (output) have been transferred to the job context (outcome).
Long-term ex-post evaluation is the only way to measure the true results of the scheme. These results include increase in productivity, lower error margin, fewer complaints, cost reduction, or improvement of product quality (Maes and Sels, 1999). Eventually, these are the intended effects of the training (behavioural additionality), and these effects should contribute to the overall objective of increasing company profitability.

According to Vedung (1997, p. 7) evaluation is retrospective *by definition*. Indeed, although a judgement about the *expected* effects of a policy measure can be made beforehand, a systematic and reliable assessment of the effectiveness of a certain public policy intervention in the market system is, by definition, retrospective. Since some results of state aid need some considerable time to take effect, the evaluation may take place some years after the support has been granted. Like Vedung, we are of the opinion that retrospective does not necessarily concern entire support programmes only. Evaluation can also take place after the completion of a significant number of *support cases*. Thus, with support schemes that have a lead-time of several years, both ex-nunc and ex-post evaluations can measure the impact on supported beneficiaries, and provide inputs for improvement of the scheme in the future.

4.2.3 *Storey's six steps to heaven*

In his article 'Six steps to heaven', Professor David Storey (2002) describes six steps of appraisal of particular programmes of small business support, ranked in terms of sophistication (step 1 being the least sophisticated and step 6 the most sophisticated approach for appraising state aid). This classification is based on various forms of monitoring and evaluation practice in Great Britain.

Table 2 Storey's six steps to heaven

Monitoring		
Step 1	Take up of schemes	Involves a description of scheme and assisted firms' characteristics.
Step 2	Recipients' opinions	Assisted firms are asked for their opinions with regard to the value of the scheme and the application procedures.
Step 3	Recipients' views of the differences made by the assistance	Firms are asked about the effects, additionality and displacement invoked by the support measure.
Evaluation		
Step 4	Comparison of the performance of 'assisted' with 'typical' firms	Measures the development of employment and sales, and survival rate of assisted firms, compared to the SME sector's average.
Step 5	Comparison with 'match' firms	Instead of the average SME, assisted firms' performance is compared with the performance of a 'match' with the same characteristics as the assisted firms.
Step 6	Taking account of selection bias	Same as step 5, but now the 'match' firms are selected, taking account of selection bias, utilising the statistical Heckman technique.

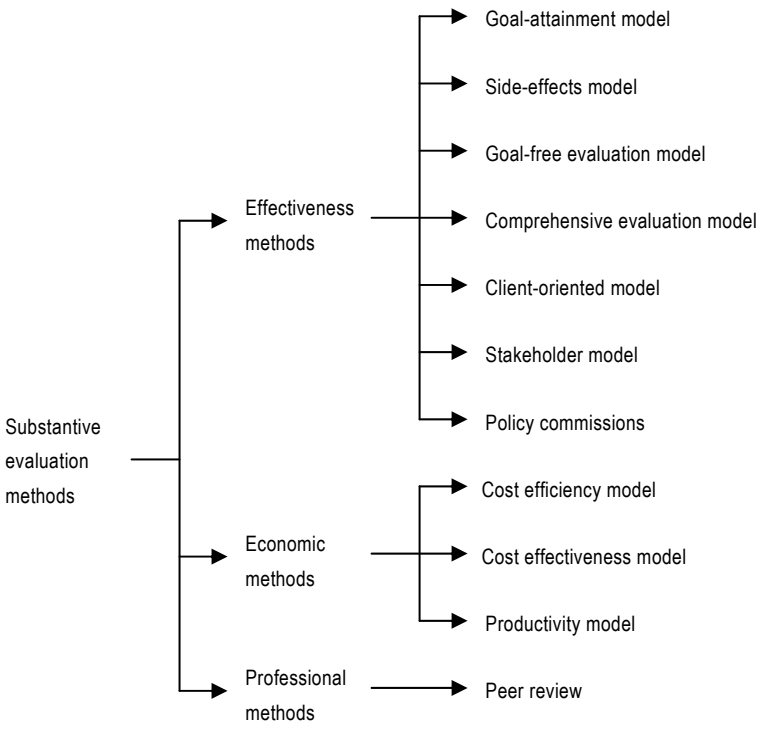
Source: Storey (2002)

From the viewpoint of research reliability, it is important for the most sophisticated analysis possible to be undertaken, since the more sophisticated steps are able to specify the deadweight more accurately than the less sophisticated steps. Inadequate specification of the deadweight invariably leads to overestimation of the state aid programme's effects. However, in his recent paper Storey argued that not all policies need to be evaluated at the same level of sophistication – unless policy makers want to compare the results of one programme with another (OECD, 2004). According to Storey, step 5 is perfectly valid if there is no systematic selection of assisted firms. If selection bias does occur (which is the case with most schemes) step 6 is an evaluator's 'heaven'.

4.2.4 *Substantive evaluation models*

Evert Vedung (2000) gives the following review of approaches for evaluating government policies. These approaches present the general basic principles for designing an evaluation, not the technical details concerning data collection and analysis.

Figure 7 Overview of types of substantive evaluation models for scrutinising government policies



Source: EIM, based on Evert Vedung (2000)

Economic methods

– **Goal-attainment** This is the basic evaluation approach, where the evaluator judges whether the goals of the programme have been reached, and the effects are a result of the support measure.

– **Side effects** This approach does not only take the goals of the support measure into account, but also looks at the side effects of the measure. What other effects does the support measure invoke, be they positive or negative?

– **Goal-free evaluation** This approach assesses the effects of an (economic) intervention, regardless of the objectives of the measure. The supporters of this approach are of the opinion that pre-occupation with the objectives of the measure narrows the view of the evaluator.

– **Comprehensive evaluation** This approach incorporates the implementation and sometimes even the planning process of the support measure in the evaluation. In other words: parts of the intervention other than the outputs and outcomes are evaluated, for example the processes of implementation and feedback.

– **Client-oriented** This approach takes clients' (beneficiaries') goals, expectations, concerns or needs as the criterion of merit. This approach is based on the question whether a measure satisfies the clients' concerns and expectations (in contrast with the question whether the measures' goals have been met). This market-drive perspective acknowledges the fact that support

recipients' objectives and drives do not necessarily coincide with the programme management's goals.

- **Stakeholder** This approach acknowledges the effects of the intervention in the recipients' clients, competitors, suppliers, etc. The stakeholder approach organises an evaluation around the organisations (people) that have an interest in or are affected by the intervention. This is a fundamentally different approach from the goals attainment or even the side effects models.
- **Policy commissions** This approach is the Swedish alternative to the stakeholder approach, where the stakeholders are not consulted but perform the evaluation. Stakeholders are invited by the government to take place in an ad hoc policy commission that should advise the government on the effectiveness of the scheme. The government does specify the issues that should be part of the evaluation, but does *not* interfere with its completion. The tradition of the policy commissions is that they are future oriented. Commissions' analyses are focused much more on alternatives for future action than on impacts of past policies. However, in practice the work of these policy commissions is much more a political enterprise than thorough research work.

Economic methods

Many effectiveness models neglect the costs of a measure. Economic models do incorporate the costs. They could calculate a ratio between the overall costs (including negative side effects) and the overall benefits (including positive side effects). Economic methods have the advantage that they produce ratio's that can be communicate to the public easily. However, these methods do not allow any nuances and the ratios can start 'leading their own lives'.

- **Cost efficiency** This approach measures both programme inputs and outcomes in monetary terms. These two are divided by each other, so that if the outcome has a higher value than the inputs, the ratio is larger than 1.
- **Cost effectiveness** This approach measures inputs in estimated monetary terms and outcomes in terms of actual impact. These two are also divided in such a way that the cost per unit of outcome is quantified (also see figure 6).
- **Productivity** A simple example of an economic productivity evaluation of a job creation programme would be the net number of jobs created, divided by its total costs in euros. Net means corrected for displacement and additionality. This ratio determines the production of 1 euro of the taxpayers' money. More advanced ratio analysis could include also other effects, such as welfare increase.

– **Peer review** This approach measures the effects of a measure against the criteria of the sponsored profession. Peers are expected to judge their colleagues against the rules of their own profession. In a peer review lawyers evaluate lawyers and surgeons evaluate surgeons. This model seems inappropriate for our purpose, since the evaluation of economic intervention programmes does not involve the evaluation of the performance of individual professionals, managers or entrepreneurs, but the evaluation of the effectiveness of specific programmes. Only in some cases (where several authorities provide the same type of aid) could peer reviews be applied.

4.3 Conclusion

Because of the previously mentioned issues and due to the fact that each state aid scheme has its own specific characteristics, there is not *one best way* to measure the effectiveness of state aid (OECD 1999, p. 21); nor is there one single method that can be applied in all circumstances. The European Commission claims that there is no golden rule with regard to evaluation techniques: 'there is no single evaluation method which is universally applicable. Instead the choice of techniques should be determined by the particular evaluation problems at hand' (European Commission, 1997a, p. 42).

Moreover, scientific (effectiveness) evaluation methods are still being developed, so it would be unwise to seek the 'best' approach that could be applied to all schemes (Rossi, Freeman and Lipsey 1999). Methods for evaluating effectiveness should be tailored to (1) the aim of the evaluation, (2) the specific circumstances, characteristics, objectives and structure of the aid scheme (*e.g.*, the existence of *ex-ante* measurements, amount of individual aid grants, data availability, etc.) and (3) the characteristics and structure of the targeted enterprises.

The various classifications of methods give an indication of the types of methods that could be used to assess the integral effects of state aid measures. Ideally, we would combine the advantages of the presented approaches. The ideal evaluation method combines a measurement of goal attainment with a goal-free identification of the scheme's side effects (i.e. displacement, additionality and other side effects), stakeholders' concerns, and the efficiency with which each euro taxpayers' money is spent. From the article by David Storey (2002) it is clear that evaluation of the effectiveness of SME support cannot do without a carefully selected control group.

Do's and Don'ts in evaluation research

Do's

- Start thinking about evaluation *before* implementation of a support measure
- Design support measures with quantified objectives
- Outsource the evaluation to an independent evaluator
- Evaluate according to the programme's objectives
- Combine quantitative and qualitative techniques
- Identify all the stakeholders and their interests and objectives of the intervention
- Choose appropriate sample sizes: not too

Don'ts

- Think of evaluation as a necessary evil
- Postpone evaluation until the support measure has ended
- Limit the evaluation to simple goal attainment
- Expect unbiased results from self-evaluations or surveys among aid recipients
- Think that evaluation is easy
- Use only one source of information
- Apply a standard approach to evaluate unique intervention programmes
- Forget to include deadweight (counterfactual), displacement and additionality.

small but not too large either – Use a control group – Take account of selection bias – Integrate evaluation in the entire policy process – Learn from evaluation results	– Neglect the positive and negative side effects of interventions – Think that evaluation is too difficult and too expensive

Source: EIM, 2004

5 The process of effectiveness evaluation

The process of public policy evaluation is described in literature. In general, four phases can be identified in an evalution study, each phase having its own characteristics: (1) the structuring phase; (2) the data collection phase; (3) the data analysis phase; (4) the judgement phase. There are many different techniques that can be used for designing evaluations, gathering data, analysing data, and making a judgment. The quality of an evaluation is influenced by the evaluation process design. The selection of the appropriate techniques must be based on the research question that underlies the evaluation.

5.1 Introduction

It is obvious that the outcome of an evaluation should reflect the actual situation. However, the *outcome* of effectiveness studies may be influenced by the quality of the evaluation. Venetoklis (2000) demonstrated that the research methods, the type of data used, and the independence of the evaluator influence the outcome of the evaluation study. Purely descriptive evaluation methods yield far more positive results than econometric/statistical evaluation methods, which yielded much more balanced results[20]. Therefore, we suggest using a combination of both primary and secondary data and, using an array of analysis tools and techniques, of both qualitative *and* quantitative methods (see also: Ziegelaar, 2004).
Secondly, Venetoklis found that studies using primary data (survey results) yielded far more positive conclusions than studies using secondary data (existing statistics and Tax Office data). Possibly socially desirable responses are responsible for part of this phenomenon.
Thirdly, Venetoklis demonstrated that evaluation studies commissioned by the aid distributing agencies seemed to produce more favourable results than independently initiated evaluations[21].

To balance the above issues, thorough evaluation demands a well thought out and systematic approach in order to mirror the true effects of an aid measure. Ideally, the (quantifiable) criteria are established prior to the design of the aid programme. As we have seen, the evaluation criteria may be based on the underlying policy objective, but this is not necessarily the case. Programmes can also be judged using other criteria, such as cost effectiveness, scope or reach of the programme, its performance relative to other programmes, or the programme's contribution to the national welfare.

The literature describes the process of public policy evaluation. In general, the following phases within an evaluation study can be identified (European Commission, 2003):

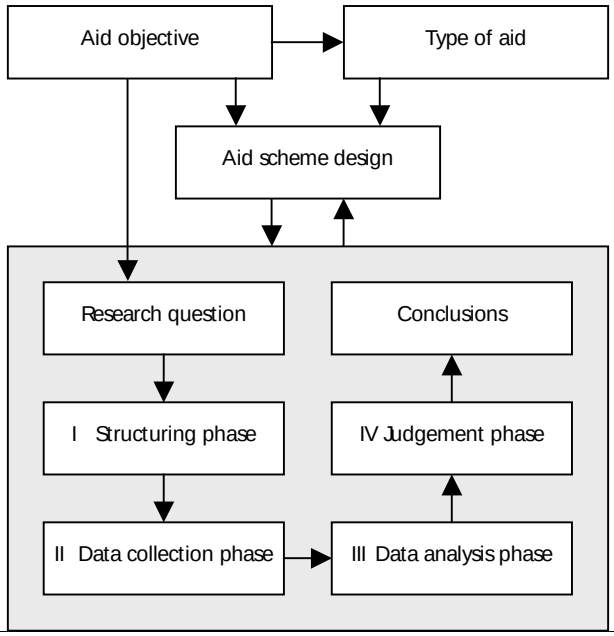
1 Structuring phase
2 Data collection phase
3 Data analysis phase
4 Judgement phase

[20] Venetoklis assessed 22 state aid evaluation studies. Of these 22 studies, 9 were descriptive (qualitative) and 13 were statistical / econometric studies. All 9 descriptive studies reached positive conclusions (i.e. that the support scheme was successful), while 6 of the 13 statistical / econometric studies reached a negative verdict (i.e. that the support measure was ineffective).

[21] On the other hand, Professor David Storey argues that evaluations, like any other audit, should always be performed by independent external evaluators. However, "if a key role of evaluation is to contribute to making continuous improvements in the policy, then the evaluators need the active co-operation and involvement of both policy-makers and deliverers" (OECD, 2004).

However, prior to these phases comes a possibly more important activity: the definition of the research question. Since an evaluation is research and each research starts off with a research question, good evaluations also are based on a research question. The aim of the evaluation research is to answer that question, bearing in mind the application of the evaluation: learning and improving policies and implementations. With this in mind, the evaluation process can be depicted as follows (see figure 9):

Figure 8 The different phases of an evaluation study



Source: EIM, 2004

5.2 Research question

The first and single most important step of an evaluation study is the formulation of a research question (Rossi et al., 1999). The research question is the foundation on which the other parts of the study are built. Bekkers and Ringeling (2003), propose a framework for basic research questions in the scope of policy evaluations (see figure 10):

- Does it work?
- Is it allowed?
- Is it correct?
- Is it suitable?

Figure 9 Framework for policy evaluation research

	Action theory	Valuation theory
Logic of consequence	Instrumental rationality: Does it work?	Legal rationality: Is it allowed?
Logic of appropriateness	Politic rationality: Is it suitable?	Normative rationality: Is it correct?

Source: EIM, based on the interview with Professor Ringeling, and Bekkers and Ringeling (2003, p. 22)

State aid effectiveness is concerned with the consequences of an action. Therefore, the evaluation questions should fall in the upper left quadrant of the above matrix. The questions should be assessed from the viewpoint of instrumental rationality: *does it work*? The 'it' refers to the state aid measure or programme.

According to Bekkers and Ringeling (2003, p. 22), the basic research question 'does it work' is represents a problem-oriented approach of policy analysis. Problem oriented policy analysis is aimed at an analysis as regards the contents of the causes of social problems. 'A policy works or is effective when the chosen policy measures lead to attaining the goals of the policy' (Bekkers and Ringeling, 2003, p. 23). The question if the policy was effective is uncoupled from the question whether the policy (aim) is just, fit and/or allowed.

Research questions with regard to the evaluation of effectiveness

For the purpose of performing the evaluation research, the basic research question 'does it work' should be translated into derived research questions. Since this report is concerned with evaluations of the effectiveness of state aid measures, these derived research questions will have to take into account the total range of (both positive and negative) effects of the support measure. However, in literature there is no agreement about what effectiveness actually is. As we have seen in a previous chapter, effectiveness could be identified in terms of:

- The outputs (i.e. performance indicators),
- The outcomes (i.e. the socio-economic effects of the support),
- The effectivity (i.e. how effective outputs are transformed into socio-economic effects),
- The goal attainment,
- The cost-effectiveness, or
- The total macro-economic impact of the support.

Special attention should be paid to:

- Causality
- Additionality
- Displacement

– Deadweight

Since the (external) evaluator often does not know the criteria against which the policy makers will judge the success of a support measure, the principal of the study often formulates the research questions. These research questions must be based on the success criteria of the scheme (derived from the targeted market failure), and also on other factors described in this report. It is also important for the evaluator to identify possible changes in the original and intermediate objectives. Sometimes these objectives have been changed during the course of the support programme. Insight into these changes can prevent unnecessary problems during the evaluation study.

Example

An example of research questions from a Dutch evaluation study of the Law Stimulating Research & Development (WBSO) are (Brouwer et al. 2002):

– Does WBSO stimulate the R&D expenditures of companies?
– What are the displacement and additionality effects?
– Are these relationships *causal*?
– What is the role of WBSO in the decision making of companies about R&D?
– What other effects does WBSO have on the R&D process?
– What are the private returns of WBSO investment?
– What is the contribution of WBSO to innovation objectives?
– To what extent is the target population reached?
– How efficiently is the scheme executed?
– What possible amendments are necessary to improve the scheme?

5.3 Structuring phase

After the research question of the evaluation study has been formulated, the evaluation study must be designed. During the structuring phase, the research process is designed in such a fashion that the research question can be answered in a cost-effective way and with high quality. This implies formulating the general principles and hypotheses of the study (a 'picture' of the relevant aspects of the 'real world' that the researchers expect to find), selecting the aspects of the 'real world' that need to be analysed, making a choice about what data is needed to answer the research questions, identifying and selecting data sources, and choosing research techniques with which these data should be analysed. In the structuring phase, choices are made that determine the remainder of the effectiveness study[22]. When the effectiveness study is outsourced, the external consultants often perform the structuring phase during the drafting of the proposal. The principal commissioning the study then chooses the consultant whose approach is most cost effective and/or will provide the highest quality.

When designing an evaluation study, policy makers and evaluators should ask themselves such questions as:

– What should be evaluated: a single support measure, or an entire support programme?
– What should be measured: the outputs (i.e. performance indicators), the outcomes (i.e. the socio-economic effects of the support), the effectivity (i.e. how effective outputs are transformed into socio-economic effects), the goal attainment, the efficiency, the cost-effectiveness, or the total macro-economic impact of the support.

22 For example: in this stage the consultant needs to think about whether or not he will make use of a control group, whether companies that receive multiple types of aid should be excluded from the study because this complicates the analyses, whether a survey will be used or not, etc.

– How should this be measured: literature research, interviews, case studies, experts / Delphi, secondary statistical data, questionnaire surveys, file research, application data, group discussions, etc.?
– How should we analyse this data: quantitatively, qualitatively, or otherwise?
– Should we use a control group?
– How should we make a judgement: cost-benefit analysis, multi-criteria, or otherwise?

There are several techniques for structuring an evaluation. These techniques include:
– The intervention logic (Jansen Schoonhoven and Vos, 1998; European Commission, 1995, p. 25)
– The colour vote (European Commission, 1999a, volume 3, page 39 – 58)
– The impact matrix (European Commission, 1999a, volume 3, page 39 – 58),
– Inductive logic (Levin-Rozalis, 2000),
– Abduction (Levin-Rozalis, 2000), and
– Realistic evaluation (Pawson and Tilley, 1997).

In this report we will describe only the intervention logic, since this is the most appropriate technique. Readers are referred to the references if they wish to know more about the other techniques.

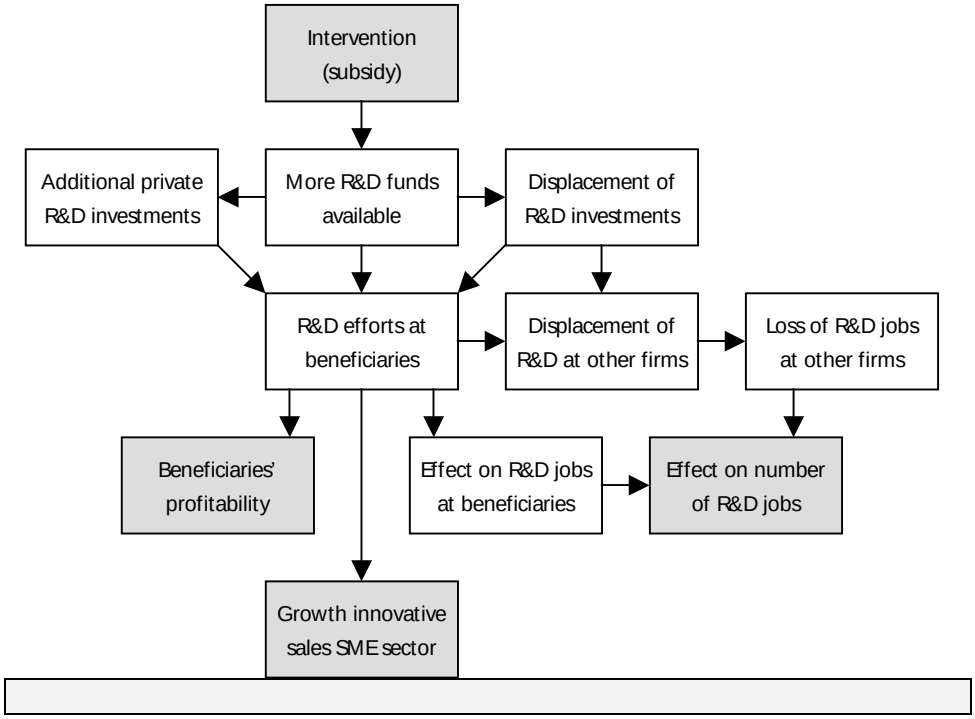
The intervention logic method

The intervention logic method analyses the expected effects of intervention in a logic system of causal relationships. A chain of causal relationships could be: available capital leading to higher spending on R&D, which in turn leads to increased innovation, that improves the competitiveness and profitability of a firm, so stimulating the national economy. These presumed causal relationships are often depicted in cause-and-effect diagrams (also called impact diagrams), or as result chains, both showing more or less clearly the (expected) causal relations between phenomena (see also Jansen Schoonhoven and Vos (1998), and European Commission (1995, p. 25)).

The intervention logic analyses the expected effects of an (economic) intervention in the chain of causal relationships. For example, an intervention could be a subsidy for R&D projects. This subsidy is expected to have an impact on the behaviour of firms, and thus induces the chain of causes and effects. The logic result of the intervention would be a change in the R&D structure, and possibly in the innovation, competitiveness and profitability of the subsidised firms. The testable hypothesis would be that subsidies improve the competitiveness and profitability of the firm.

Example

The diagram below depicts a fictitious impact diagram for an R&D subsidy scheme:



Control groups

Many tools for data analysis work with control groups. There are various ways of establishing control groups. Which alternative should be chosen depends on various factors, such as the timing of the design of the evaluation (prior to, during or after the support scheme has been carried out).

Control groups are possible only if a programme has no full coverage, i.e. not all firms that fulfil the relevant criteria are supported. With full coverage programmes, all the firms that apply to the criteria have received support. Here, unsupported firms (that could in principal have been considered for the control group) differ from the supported group considerably on a number of relevant criteria.

As a result, with full-coverage programmes, unsupported companies are inherently different from supported companies. The systematic bias is called the sample selection bias (Wren and Storey, 2002). Sample selection bias means that analyses are influenced (biased) by certain structural differences between the control group and the firms that were selected for assistance. These differences occur because of self-selection or agency selection of assisted firms. Due to this systematic bias, it is difficult to construct a proper control group, but there are several techniques that can cope with the sample selection bias problem. In any case, with full coverage programmes the deadweight (or counterfactual) should be based on the state of being before intervention.

– **Input-output models** Several outcome and input variables are interrelated. Input-output analysis is used to identify the linkages between the variables and thus can be useful in developing forecasts. This technique should be applied when performing ex-ante evaluations.

– **Econometric studies using regression models**

According to Venetoklis (2000), ex-post evaluations should utilise both descriptive and econometric methods. Econometric regression models measure the causality between the policy intervention (*cause*) and the observed impact (*effect*) (Venetoklis 2002). However, econometric studies of the impact of economic interventions suffer from a lack of reliable data. Therefore, econometric models have to use dummies or proxy variables, this has an adverse effect on the validity of causal relationships (Ederveen et al., 2002).
This technique should be applied whenever possible. These models are very laborious and require large amounts of high quality quantitative data,

– **Experimental design**

This is a method where the target group is divided in an 'experimental group' (= supported enterprises) and a 'control group' prior to the execution of the scheme. The experimental group and the control group are compared prior to and after completion of the programme. In methodological literature, the experimental design is regarded as the best causal-analytical design. However, experiments also have disadvantages, such as selection bias, availability of complete and reliable data, the fact that only an limited number of variables can be included in the evaluation research, the fact that experiments can not identify the reasons for possible policy failure ("the 'black box' between input and output remains closed"), and its labour intensiveness. An experiment often costs a group of experienced evaluators 1 to 3 years (see: Verschuren, 1998 and Ziegelaar 2004).
This technique should be applied when evaluations are designed prior to execution of the intervention and only a part of the target group is going to be supported.

– **Randomised design**

This is an experimental design technique where firms are assigned to the experimental group or th control group randomly. According to Rossi et al. (1999, p. 279), "The randomised controlled experiment is the strongest design for assessing net impacts of interventions" of support programmes with partial coverage. In practice it is hardly used because it is very expensive, elaborative and it would create inequality because some firms that are entitled to assistance, are denied this assistance because they must be placed in the control group.

– **Quasi-experimental designs**

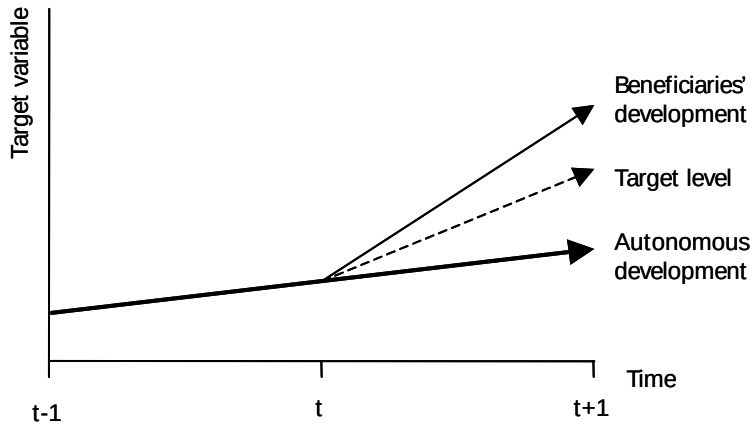
These are used in environments where companies have not been randomly assigned to the 'experiment' or control group. There are various techniques for creating a control group in a quasi-experimental design, the most important are:

- <u>Statistical control designs</u>

are impact designs where a control group is statistically equated by a multivariate statistical procedure on characteristics that are associated with the programme outcomes. This is a method for constructing an ex-post control group.
This technique should be applied when a control group cannot be established prior to the execution of the scheme.

	- Heckman 2-step adjustment procedure	formulates a single equation to explain the selection procedure and then formulates a second equation that explains performance change based on the factors included in the selection equation (also see: annex III). This technique should be applied when self-selection is expected to have taken place.
	- Matching	is a method where each participant is coupled to a mirror firm that shares the same characteristics. Matching can be based on various statistical or qualitative criteria, such as number of employees, annual turnover, export region, sector, location, etc. This technique should be applied when there are mirror firms available.
–	**Reflexive controls**	Reflexive controls measure the outcome variables taken on participating targets before intervention and use these as control observations. The net effect is calculated as follows: [outcomes for participants after intervention] – [outcomes for participants before intervention] ± [effects of other processes at work during the intervention] + [design effects and stochastic error][23].
–	**Time-series analysis**	Time-series analyses compare the outcomes (e.g. the level of R&D expenditures) before the start of the support with the outcomes after the completion of the support.

Figure 10 Time series analysis



Source: EIM, 2004

Results of the structuring phase

The structuring phase results in a model, which is used to describe the support measure and to provide evidence of the measure's effects (European Commission, 1997a, p. 42). Using this model the evaluator can determine what data need to be collected for the evaluation study. Such data could be:

– Qualitative or quantitative data

23 Source: Rossi, 1999, pp. 347-348

- Primary or secondary data
- Information about various subjects, such as participants, non-participants (for example in the case of a control group), the aid measure itself, stakeholders, the economy as a whole, etc.

Also, the structuring phase should result in an approach how the deadweight (counterfactual), additionality and displacement will be identified.

We can conclude by stating that the structuring phase is very important, because in this phase decisions are taken that influence the quality and cost-effectiveness of the evaluation study. It is important that issues and assumptions are made explicit in the structuring phase, so that the policy makers and/or the evaluators take the right decisions with regard to the aspects that they want to analyse, the data that they need to collect and the methods with which they will analyse the collected information.

5.4 Data collection phase

Once the structure of the evaluation has been sorted out and it is clear which data need to be collected, a choice of methods to collect the data has to be made. One could decide to use existing data or collect new data. There are many different methods for data collection, and each of them has advantages and disadvantages. Some methods are suited to obtain a specific type of data. For example, statistical data on macro-economic development can be more easily obtained from secondary data sources than from surveys or interviews. On the other hand, surveys, interviews and other primary data sources allow for asking exactly the right questions and including the desired control variables whereas with secondary data sources the evaluator is surrendered to the existing sample size, definitions and control variables, etc.

But surveys are not well suited for collecting information about *why* things are as they are. Therefore, techniques for collecting qualitative data should be employed, such as interviews, or case studies. All and all, the type of information needed (that has been made explicit during the evaluation design) determines which tools and techniques ought to be used for collection data during the evaluation. A study to identify the potential side effects of a support measure is likely to use interviews and a survey, whereas a study that investigates the overall economic effects should analyse existing economic statistics. The following matrix relates the several data collection techniques to each other.

	Primary data	*Secondary data*
Qualitative data	– Interview – Case study – Group discussions	– Literature research – Expert panel
Quantitative data	– Application form data – Information during lead-time – Questionnaire surveys	– Existing statistics – Firm financial data from public registers

The table below gives some characteristics of the most important data collection methods[24].

[24] For information on these and other methods, see: Babbie (2004), Neuman (2000) and Hart et al., (1998)

Method	*Advantages*	*Disadvantages*
Qualitative		
Literature research	– Learn from previous studies – Cost efficient – learn about the to be expected causal relations	– Definitions not always correct – No control over selections and methods used by the authors – Only relevant for the design of the scheme
Interview	– Interactive, questions can be adapted to the situation, experience and knowledge of the respondent – More time available than with surveys – More in-depth, qualitative information can be found (why-questions) – Investigate processes, explaining behaviour and arguments for certain behaviour	– Expensive – Longer lead-time – Difficult to judge whether 'facts' can be generalised – Subjective opinions cannot be generalised – Difficult to quantify the effect of the scheme
Case study	– A good, in-depth understanding of *how* the measure works can be obtained. Combination of several data collection methods is possible	– Very expensive – Depth of analysis comes at the costs of loosing broad coverage. Only a limited number of participants are described. Therefore, the outcome of the case studies cannot be generalised – External validity
Expert panel	– Expert panels can be consulted in areas where objective reliable data is absent – Relatively inexpensive and quick – Large group of experts can produce statistically significant results – Convergence of opinions is possible – Creating involvement and public support for evaluation results	– Although expert opinions can be collected and summarised systematically, they remain subjective – Possible disagreement between experts – Quality of the data depends on the quality and reliability of the experts
Group discussion	– Interaction between the participants	– Subjective opinions, often not statistically significant – External validity
Quantitative		
Collecting existing statistics	– Cheap – Quick – Specialised agencies (e.g. statistics office or tax office) often have larger samples than can be obtained when collecting primary data	– Variables were constructed for another purpose – Questions were formulated otherwise than optimal for the evaluation – Impossible to correct data collection during the process
Application form data, firm financial data and information during the lead time of the project	– Cheap and quick – All users of the measure are included	– Evaluation design needs to be made before the scheme goes 'live' – Information from support receivers only, no information about a control group (step 1 of Storey, 2002) – Socially desirable answers
Questionnaire survey	– Collecting large datasets – Systematic – Quick	– Questionnaires can be quite costly – Selection problem: supported firms may give 'socially desirable' answers – Potential response bias

As we see above, there are many techniques for collection the required data. Existing evaluation studies all use different techniques for collection data. Which technique should be used, under what circumstances and for what purpose? According to Venetoklis (2000), estimations of the impact of business subsidy programmes should not be based on primary data (from interviews or questionnaires filled in by recipient firms), but rather on secondary data (from financial statements of firms). In our opinion, however, data from both primary and secondary sources should be collected and used to determine the effects of state aid. Also, qualitative and quantitative types of information should be used. These sources complement each other, and qualitative information can be used to explain the phenomena that are identified using quantitative data.

For example: if statistics from the National Statistics Office indicate that supported firms have increased their labour force significantly more than other firms, the evaluator can use interview techniques to identify *why* these supported firms have expanded the number of staff on their payroll, and what type of personnel it concerns; thus showing the causality. It could be that the assistance made it possible to increase production, or it could be that the aid required so much paperwork that the company had to hire someone to handle the administrative burden of the state aid received. Thus, qualitative information from primary sources complements the quantitative secondary data.

5.5 Data analysis phase

Data comprises facts and figures. Data themselves are however not yet an answer to the research question of the evaluation study. It takes an expert to transform data into information, on the basis of which the research question can be answered.

This expert may use a variety of tools to analyse evaluation data. Some of the tools are very obvious. Statistical data will be analysed with the help of statistical techniques. But a large variety of statistical techniques exist, ranging from straightforward and easy to interpret, through very complex and difficult to interpret. And statistical techniques in particular pose certain requirements as to the amount (number of observations, or 'cases'), scale (nominal/ordinal/interval/ratio) and quality (e.g. missing values) of the data. Below, the most frequently used data analysis methods are listed[25]:

Qualitative

- **Document analysis techniques, such as:**
  - Coding & abstraction: Coding & abstraction is a technique for systematising qualitative data, such as literature data, interview data and case study data. Categories of concepts used to label data are identified, (coding); linked categories of data are grouped and conceptualised at a higher level of abstraction to produce conclusions. This technique should be applied when there are large quantities of unsystematic qualitative data that need to be processed.
  - Data matrices: Key themes or dimensions are identified and data is sorted in respect to these themes, hence making it easier to draw out patterns across data. This technique should be applied when systematic qualitative data are available, but not surveyable.

25 For information on these and other methods, see Babbie (2004), Neuman (2000) and Hart et al., (1998)

- Time-series qualitative data analysis	Data are ordered chronologically to provide an account of activities and events to identify causal relationships. This technique should be applied when the evaluator wants to draw conclusions from qualitative data in case studies. This technique is not robust enough to make larger generalisations.
- Frequency counts	The identification of key themes and assertions and counting the number of times that they occur in the data. This technique should be applied when qualitative data are systematic and can be classified into clear and distinct categories.
– **SWOT-analysis**	Qualitative analysis method, whereby the Strengths, Weaknesses, Opportunities and Threats of a scheme are identified. This technique should be applied e.g. in the sphere of interviews with stakeholders to identify all aspects of a support measure.
– **Delphi**	Procedure to develop judgemental forecasts in which experts can react in several rounds to each other's position without the potential biases inherent to group meetings (dominant positions etc). Cycles can proceed until consensus emerges. This technique should be applied when experts with in-depth knowledge are asked to participate in a group discussion. Delphi is a powerful tool for group discussions with experts.
Quantitative	
– **Statistical techniques, such as:**	
- Univariate analysis	Analysis of a single variable for multiple groups. This technique should be applied when the evaluator wants to show a difference between two or more groups of entities, on one characteristic (e.g. compare the mean R&D expenditures of the treatment group with the mean R&D expenditures of the control group).
- Bivariate and multivariate analysis	Analysis of the relationship between two or more variables in which the variables are to be analysed simultaneously (search for differences or association). This technique should be applied when the evaluator wants to demonstrate the effect of a change in one variable on the value of another variable (e.g. the association between the heights of the aid-amount and the level of R&D of the recipient company).

- Regression analysis	Regression analysis is a statistical technique where the evaluator studies the relationship between the policy variable (e.g. the amount of support), parameters (e.g. economic climate) and the target variable (e.g. the number of SMEs created). The relation between these variables is calculated in the form of the regression coefficient. In practice the policy variable is not the only variable that completely explains changes in the target variable, but other factors also influence the behaviour of the target firm. Therefore, it is very important to think about the regression model accurately. Including the correct intermediate, latent and autonomous variables in the regression model is important, since the selection of variables determines the quality (validity and explanatory power) of the regression coefficient (see: Verschuren, 1998 for further details). This technique should be applied when the evaluator wants to explain the level of the target variable (e.g. level of innovation) as the result of two or more causes (e.g. amount of aid, firm size and industry).
- Factor analysis	A statistical approach to analyse interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors). This technique should be applied when the evaluator cannot grasp the essence of a phenomenon in one single variable or survey question. The values of several variables are combined in one single factor.
- Variance analysis	Technique that determines whether samples are from the same population with equal means, based on one or more dependent measures. This technique should be applied when the evaluator wants to know whether the target group and the control group vary significantly on the target variable (e.g. level of innovation).

5.6 Judgement phase

In the judgement phase the results of the analysis are judged against the pre-specified criteria. Often, judgement involves comparisons: before vs. after, inside vs. outside the region, supported vs. unsupported enterprises, etc. Based on this comparison politicians or scheme administrators may decide to continue, adapt or terminate the aid scheme.

Often, the researcher performing the evaluation study finds additional results during the evaluation project. For example, ideas may have been generated about how the scheme could be improved or why the scheme should be continued although it does not meet its objectives. These additional results should also be included in the evaluation report because they may provide valuable insights in the *overall* performance of the aid measure.

Judgement techniques

Jansen (1984) gives a comprehensive overview of judgement techniques that can be used to compare alternative ways to implement environmental policy (see annex I). Some of these methods can also be applied for the *systematic investigation of the effectiveness of economic intervention programmes*

whereby the performance of the programme is judged against pre-specified standards or criteria. In the view of this definition, only the following judgement techniques are relevant:

– **Cost/benefit analysis**	This analysis translates all costs, negative and positive effects of the measure into monetary values. Costs and negative effects are calculated according to their market price. Intangible negative aspects are valued at 'shadow prices' through the making of assumptions. The monetary value of the positive effects is then divided by the monetary value of the negative effects. This technique should be applied when the principal wants to know whether or not the costs exceeded the benefits of the scheme.
– **Cost-effectiveness analysis**	This approach judges whether the alternative policy measures (or for our purpose, the achieved results) are worth the costs. Evaluation criteria are operationalised quantitative target values. This approach does not take account of the weighting of the goals. This technique should be applied in ex-ante evaluations, when one of a group of policy measures should be chosen.
– **Planning Balance Sheet method**	This is an extension of the cost/benefit method. However, distinction is made between the costs and benefits for various stakeholder groups. This technique should be applied when stakeholders' interests vary and/or are opposed.
– **Goal Achievement matrix**	This approach aims to measure the extent to which the support scheme reaches the quantified targets. This technique should be applied during ex-post evaluations, whenever possible. However, attention should also be paid to other effects than the target.
– **Graphic representation**	Evaluation studies with relatively limited amounts of information can often be displayed graphically. Graphic representation makes the analysis results clearer for the reader, although it does not take the weights of the various targets/variables into account. This technique should be applied when the readers require a low-barrier presentation of the results.

Other judgement techniques, not listed by Jansen (1984) are:

- **Economic impact analysis** — Is the impact of the measure on the national welfare (the sum of consumer and producer surplus) positive, i.e. is the situation with the support measure more efficient than the situation without the support measure? If performed correctly, the situation without the support measure should also account for lower taxes, where the market allocates the extra tax money itself (see Martin, 2001, pp. 98-104 and pp. 159-165 and Sheperd 1997, p 416).
  There are two approaches: The Economic Efficiency Approach concerns maximising the wealth of the total society, whereas the Distributional Justice Approach concerns the way that wealth is distributed among society's individuals (see Venetoklis 2002, p. 36).
  This technique should be applied when complete support programmes are evaluated for their long-term benefit to the macro-economic development of society as a whole.
- **Benchmarking** — The benchmark approach estimates how well the programme performs, compared to other (reference, or 'benchmark') programmes (see European Commission 1999a, vol. 3).
  This technique should be applied when the relative performance is judged against the performance of other support measures (e.g. another support instrument that targets the same market failure, or the same support measure in another region).
- **Multi criteria analysis** — The performance of the support measure is judged by more than one criterion. These criteria can be weighted (see European Commission 1999a, vol. 3). The weight factors can be determined using a questionnaire survey among stakeholders, followed by confirmatory factor analysis (for technical details, see Henry 2002)
  This technique should be applied when there are several success criteria for the support measure (either ex-ante or ex-post).

5.7 Pitfalls

Apart from the aforementioned phases of the evaluation process, various authors mention a number of pitfalls with regard to policy evaluations.

The first pitfall concerns diversity management. Since the diversity of the community is integrated in the public sector, the public sector itself is characterised by conflicting norms. According to Brugge and Run (2004) and Professor Arthur Ringeling, the judgement phase of an evaluation should mirror the diversity in expectations of the various stakeholders that are involved with the policy process. Although a policy measure may include formal objectives and criteria, the various actors in the network involved with the policy measure all have their own norms, criteria and expectations (see Hufen and Ringeling 1990). A support measure may be successful for the EU-policy makers, since it creates additional jobs, but at the same time it may be unsuccessful for the national politicians and workers (the new jobs are underpaid), for the employers (the measure involves a heavy administrative burden that outweighs the benefits), for the unions (because they lose power), and for the agency executing the measure (because it is implemented in a way that deteriorates the agency's efficiency). Policy evaluation should

take account of this diversity in expectations. Policy makers and policy evaluators should organise this diversity in their processes. The discussion whether a policy measure succeeded or failed, should be held together with all the stakeholders of the policy measure. Majone (1996) called this process 'multiple evaluation': several judgements of a single policy are made at the same time. This will also help the policy makers to look at the incentive effects of a support measure from the perspective of the intended beneficiaries.

The second pitfall regards stakeholders' resistance to evaluation. Historically, evaluation has been performed as a means to render account of public expenditures. But also in the frame of the 'new' objective of state aid evaluation[26], evaluation is developing towards a 'tool for learning' (Uusikylä and Virtanen, 2000). This learning can result in the redirection of certain financial resources, or even termination of unsuccessful support schemes. This implies that current stakeholders may be worse of in the future. Therefore, stakeholders may be reluctant to cooperate with state aid evaluation procedures (Brugge and Run, 2004).

A third pitfall is the availability of suitable data. Many authors point out the importance of reliable data that evaluations could use to quantify the effects of a certain policy measure. A lack of reliable data can make evaluations more expensive (if new data needs to be collected), or even impossible (if information about the situation prior to implementation of the policy is not available). Therefore it is very important to start thinking about evaluation prior to the implementation of a policy measure, so that the benchmark for the evaluation study can be documented. A well-mapped starting-point is essential for determining whether the chosen route leads in the right direction!

A forth pitfall is the position of the evaluator. As indicated in the first paragraph of this chapter, the position of the evaluator may influence the outcome of the evaluation. Storey (OECD, 2004, p. 5) stresses the fact that 'ideally those undertaking the evaluations need to be independent of those responsible for the programmes, as in any audit role'. He (OECD, 2004, p. 31-32) also argues that there are three possible positions of evaluators, all of which have advantages and disadvantages. The departments responsible for delivering the programme have unique inside knowledge and understanding that will not be available to any outsiders. They are also more likely to receive the support of the people providing the programme. On the other hand there is the risk that their independence of political influence is at stake. Alternatively, evaluation specialists within the government could perform the state aid effectiveness evaluation. The advantage is that they have specialised evaluation skills and are independent of the department providing the state aid programme. However, they have less programme knowledge and will rely more on the co-operation of the programme providers.
A third option is to outsource the evaluation to consultants or academics. These outsiders are clearly under less political influence than the two previous groups and are likely to be specialists on their subject. The main disadvantage of the employment of such outsiders is that the discussion of the evaluation results risks being less engaging when it is led by outsiders, who may be viewed as less well informed, than when it follows from an evaluation conducted by those responsible for the programme. Therefore, the question 'who should perform the evaluation?' is an important question that should be answered in an early stage of the evaluation process, balancing the above-mentioned issues. In all three constellations, however, an explicit and formalised evaluation approach procedure as discussed in the previous sections of this chapter should be followed.

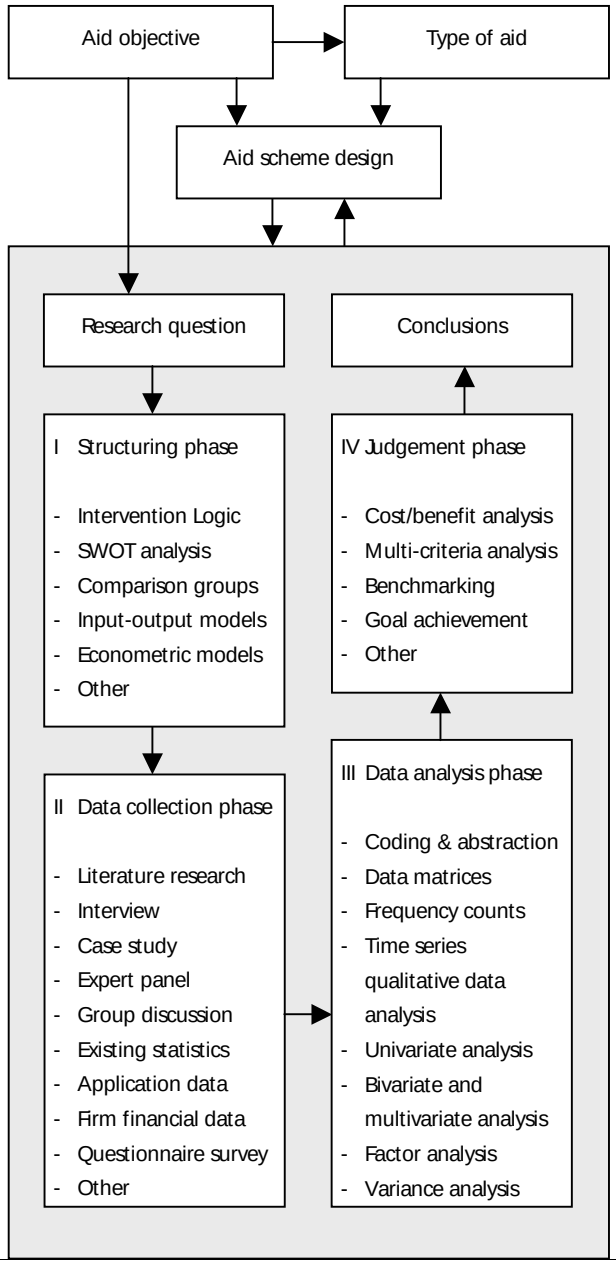
5.8 Conclusion

Evaluation uses scientific research procedures (tools and techniques) that are adapted to the specific political, economic and organisational environment of state aid. Evaluation projects are tailor made. No

[26] i.e. learning for future projects, improving decision making, priorities and distribution of public resources

two evaluations are the same, and it is impossible to identify 'one best way' for evaluating state aid. The design of evaluation projects depends on a number of factors, such as the type of economic intervention, the objective of the support policy, the goals and objectives of the evaluation study, the availability of adequate and reliable data and, of course, the evaluation budget. The judgement of the effectiveness of a support measure should include not only the goal attainment, but also the side effects and other actors' interests.

Figure 11 Tools and techniques



Source: EIM, based on European Commission, 2003

6 Evaluation practice

The SME state aid evaluation practice varies considerably between the Member States. Of the Member States that participated in our research, one sixth evaluates all state aid to SMEs, two third evaluate some state aid to SMEs and one sixth does not evaluate state aid to SMEs at all. The applied methods also differ greatly. They range from quite simple ex-ante multi criteria analysis, to intricate ex-nunc or ex-post econometric programme evaluations.

6.1 Introduction

In practice state aid to SMEs constitutes more than a half percent of the total GDP of the EU Member States. During the Member State Meeting, held on 25 May 2004, the Member States agreed that state aid evaluation is necessary to improve state aid policies and state aid schemes. Evaluation of state aid to SMEs can help to improve state aid measures as policy instruments, and to make use of taxpayers' money more effectively.

This chapter aims to provide an overview of the existing evaluation practices in the Member States. To arrive at this overview, we employed a dual approach. On the one hand, the Member States were invited to complete a questionnaire concerning their SME-support evaluation practices and to attend a meeting in Brussels on 25 May 2004. On the other hand we conducted a literature research. For this literature research, we scanned periodicals, the EIM economic library, and the internet for SME-support evaluations.

Issues raised at the Member State meeting

At the Member State meeting, held on 25 May 2004, delegations from the member states were invited to give their view on the topic of state aid evaluation. Besides comments on their approaches to evaluation of SME state aid, the following issues were raised:

1 Socially desirable answering by the respondents of surveys has to be avoided (aid receivers will never say that the provided support was not necessary)
2 The benefits of evaluation studies should outweigh the costs and limitations. This calls for intricate evaluation methods in complex circumstances where large amounts of support are involved, and simple yet powerful evaluation studies in cases where possible improvements and/or savings are limited (also see: OECD 1999).
3 Variety in the macro economic situation, state aid objectives, circumstances and available data make it difficult to identify a single best practice
4 Ideally, evaluations should also include a chapter on the 'collateral damage': the bureaucracy, the administrative burden for the enterprises, and the market distortion.
5 Market distortion is captured in the phenomenon of displacement, and has to be considered in the evaluation.
6 Evaluation should take place during the lead time of the aid scheme (intermediate or ex-nunc evaluation), in order to be able to implement learning points and adjust the scheme

Many of these issues (e.g. different macro-economic bases, market disturbance, etc.) relate to the deadweight, additionality and displacement, described in chapter 3. Administrative burden and bureaucracy are part of efficiency studies, and thus do not fall within the scope of this research (which is effectiveness).

6.2 Evaluation practice in the Member States

Based on the completed questionnaires, received from the Member States, the overview below shows the methods of state aid evaluation that are employed by the Member States. This overview does not intend to give a complete picture of the evaluation practice in the Member States, but it does intend to identify a broad range of good practices of state aid evaluation methods. Indeed, the overview shows that in practice various types of data collection and data analysis techniques are used.

Table 3 Summary of the evaluation of State Aid to SMEs in the Member States, EEA and Candidate Countries.

Country	*Q*	*A*	*E*	*Short description*
Austria	y	y	s	Some schemes are evaluated, partly internally, partly outsourced. Evaluations are mainly quantitative (using questionnaire surveys), with attention for qualitative aspects (e.g. 'improvement of market position' of the supported firms). Outputs include sets of ratios and multipliers that measure the effects in terms of money value granted. An example of this would be the ratio additional sales prompted per € 1 assistance, or the subsidy amount per job created. Austria applies both ex-nunc monitoring and ex-post evaluations.
Belgium	y	y	s	Aid and evaluation are the responsibility of the regions. Some regions do evaluate. For example, the Walloon region recently commissioned an evaluation of consultancy aid. This extensive study describes the (entry) problems SMEs experience, the support provided, and the effects of the support. The study employs various tools and techniques for data collection (among users and non-users of the support), such as literature research, file research, telephone survey and face-to-face interviews. Also, they use various methods for data analysis, such as graphic representations, statistical analyses and control groups. There is a separate chapter on synthesis and recommendations.
Bulgaria	n	n	?	
Cyprus	n	y	?	
Czech Republic	y	n	s	The ministry of Trade and Industry evaluates some schemes. A major evaluation of the Czech-Moravian Guarantee and Development Bank (whose objective it is to stimulate the SME sector) is currently underway.
Denmark	y	y	s	Denmark does evaluate state aid, but not specific SME schemes. Some of the evaluations are outsourced. Reports are in Danish.
Estonia	n	y	?	
Finland	y	y	s	Ex-ante, ex-nunc and ex-post evaluations are mandatory by law, but in practice they do not always take place. As a result, there is an impressive record of Finnish evaluation studies. However, a government working group judged the evaluation efforts *unsystematic*. Therefore, the working group has made 10 recommendations so that evaluation results wil be integrated in the policy process in a systematic manner.
France	n	y	?	
Germany	y	y	s	Germany evaluates some of its state aid. Part of these evaluation studies are performed internally, other studies are outsourced. Evaluation has helped Germany to improve state aid objectives considerably during the last 5 years. In general, German evaluations of State Aid to SMEs use quantitative approaches, where the results of assisted firms are compared with the results of the non-assisted firms. German evaluators also apply time series analysis.

Country	*Q*	*A*	*E*	*Short description*
Greece	y	y	n	In Greece, state aid measures are not evaluated yet. A method for the appraisal of aid granted is being developed.
Hungary	y	y	s	Some Hungarian state aid schemes are evaluated, partly by external consultants. Evaluations are often based on statistical analyses, but an interesting good practice in Hungary is the analysis of 80 randomly chosen micro-credit documents (including the micro-credit contracts and the recipients' business plans).
Ireland	y	y	a	
Italy	y	y	s	Italy evaluates some of its aid schemes. Examples include a quantitative analysis of the net impacts of SME subsidies on employment. This evaluation uses literature data, data from the tax and social security offices and a questionnaire. Based on this, indicators of the support scheme's effects are calculated. The indicators are applied to measure the extent to which the support objectives have been reached.
Latvia	y	y	n	Latvia does not evaluate its state aid to SMEs
Lithuania	y	y	?	
Luxembourg	n	n	?	
Malta	n	n	?	
The Netherlands	y	y	a	It is stated in the law that: - for each new or adjusted policy objective or instrument it should be considered whether ex-ante evaluation is useful. - each existing support measure should be evaluated, at least once per five years. For ex-post evaluations, the Dutch Ministry of Economic Affairs uses one standard format. The main purpose of evaluation is to check whether the policy-instrument contributes to the main goal for which it was created.
Norway	y	y	s	In Norway state aid is targeted at horizontal objectives, most of which concern SMEs. Besides evaluation results, political and administrative priorities have driven changes in the Norwegian state aid policy.
Poland	y	y	n	Poland does not evaluate state aid to SMEs.
Portugal	y	y	s	A large proportion of Portuguese state aid to SMEs is included in the Community Support Framework. An ex-ante evaluation of this framework was carried out and in 2003 the support measures (including state aid) under this framework were subject to mid term evaluations using a combination of qualitative (case studies) and quantitative (financial data, physical indicators, beneficiary questionnaires) tools, based on the MEANS quality criteria. Only state aid schemes that are not included in the Community Support Framework are not evaluated.
Romania	n	n	?	

Country	*Q*	*A*	*E*	*Short description*
Slovakia	y	y	a	The State Aid Act obliges the Slovak government to evaluate all state aid. This is done using one standard evaluation technique for all state aid. This technique consists of statistical (input-output) analyses of economic characteristics of aid recipients and a (total population) control group, (turnover, value added, profits, taxes, exports and number of employees). The evaluation focuses on four topics: technology transfer, innovation, exports and competitiveness. The outputs of the analyses are data, such as: number of jobs created through the scheme, number of companies established due to the measure, etc. The analyses are performed both on micro and macro levels. Aided SMEs are followed for 3 consecutive years to see how they develop. Data about the assisted companies are obtained through questionnaire surveys. Data about the control group are obtained from the Tax Office and the Statistical Office.
Slovenia	n	y	?	
Spain	n	y	?	
Sweden	y	y	s	The Swedes use quantitative matched sample control group approaches, total population control group approaches, cost-benefit analyses, supported with more qualitative interviews. They use these methods to measure the effect of state aid on the survival and sales of firms, employment, regional development, etc. They also assess the influence of other variables, such as sector, firm size and region. Evaluations have yielded important insights into the effects of support. The Swedes do not use the outcomes of evaluation studies for re-focussing state aid objectives.
Turkey	n	n	?	
United Kingdom	y	y	s	Recently, evaluation of all DTI services acted as inputs for a new strategy to reduce the number of services to 10 – 14. All evaluations are outsourced and follow DTI evaluation guidelines. Where possible they use a set of standard indicators.

Q = questionnaire received from the country
- y: yes
- n: no

A = attended the Member State meeting on 25 May 2004
- y: yes
- n: no

E = perform state aid evaluation,
- a: all state aid schemes are evaluated
- s: some state aid schemes are evaluated
- n: no state aid scheme is evaluated
- ?: no information available

Conclusion

The Member States show a wide range of sophistication with regard to evaluation of state aid. Some Member States do not evaluate state aid to SMEs at all, other use intricate evaluation techniques. In some countries, evaluation of state aid is mandatory.

6.3 Evaluation methods used in evaluation literature

Literature

There are numerous publications on policy evaluation. Some of these publications deal with evaluation of state support for the private (SME) sector. Some good practices are described in the tables below. In addition to the overview below, Venetoklis (2000) gives an overview of 18 Finnish studies that evaluate business subsidies. Other publications of SME state aid evaluation are listed in the bibliography.

The evaluation practice

In practice, not all published evaluation studies can be earmarked as good practice. Many so-called evaluations apply too simple tools and techniques for data collection and analysis, not taking account of the aforementioned pitfalls, such as socially desirable answers of respondents in a questionnaire, deadweight (also known as *counterfactual*), additionality, or displacement. The publications summarised in the tables below are all good practices from literature, in the sense that they excel in at least one of the previously mentioned methodological issues. However, this does not mean that they incorporate all these issues, such as deadweight, displacement or additionality simultaneously.

In the remainder of this section we present a broad overview of the literature on evaluation of state aid to SMEs. We choose to present a wide overview, which implies that we cannot go in depth. However, in annex IV, we have selected three studies that are discussed in more detail.

Country	*the Netherlands*	*the Netherlands*	*the Netherlands*
Reference to the evaluation study	Brouwer et al. 2002	Vreeker et al. 2001	Algemene Rekenkamer 2001
Title of the evaluated aid scheme	WBSO (Law Supporting Research and Development)	Maastricht airport development plan	45 policy measures aimed at stimulating SMEs and entrepreneurship.
Aid instrument evaluated	Tax exemptions		Grants; Tax exemptions; Tax deferrals; Guarantees
Type of evaluation	Ex-nunc	Ex-ante	Ex-post
Annual expenditure	€ 337 million		over € 2 billion
Policy objective of the aid scheme: *First order effect*	To stimulate R&D activities through subsidising labour costs in the form of tax and social security reductions		Stimulate entrepreneurship, start-ups and expansion
Second order effect	Increase innovation / R&D. Increase innovativeness of the target company		Realising SME potential for employment growth
Third order effect	Stimulate innovativeness of the Dutch business sector		Increase employment
Evaluation method	Input-output analysis	Multi-criteria analysis	Meta-evaluation, concerning the quality of the evaluation procedures of the ministries involved.
Evaluated effects	To evaluate the effectiveness of the WBSO. The primary objective is to evaluate the first order effect, but also light is shed on the second order effect and the third order effect.		Confronting policy objectives and performance / realization of stated objectives, based on information supplied by ministries themselves. The (quality of this) information was an evaluation criterion as well.
Tools and techniques used for structuring the evaluation study		Combination of 3 methods: flag model (based on critical threshold values), regime analysis (an advanced pair-wise comparison method) and Saaty's analytical hierarchical process (a method whereby criteria are ranked on the basis of pair-wise comparisons).	No specification of tools used
Tools and techniques used for data collection	Secondary data Literature research Telephone questionnaire surveys Interviews	Combination of qualitative and quantitative data in an impact matrix of scores ranging 3-7	
Tools and techniques used for data analysis	Econometric models		
Tools and techniques used in the judgement phase	Cost-benefit analysis Control groups	1. impact matrix of qualitative and quantitative data, used as an input for 2. defining critical threshold values, which are 3. translated into judgments, ranging from 'no reason for specific concern' to 'stop further growth'.	

Country	*Finland*	*Finland*	*Finland*
Reference to the evaluation study	Venetoklis 1999	Venetoklis 1998	Kangasharju and Venetoklis 2002
Title of the evaluated aid scheme	Not one single scheme, business subsidies in Finland in general are evaluated	KTM	Various business subsidy schemes
Aid instrument evaluated	Grants	Grants	Grants
Type of evaluation	Ex-post	Ex-post	Ex-post
Annual expenditure	Not mentioned, but total is high, compared to other EFTA countries (1991-1992)	Not mentioned	Not mentioned
Policy objective of the aid scheme: *First order effect*		Not mentioned	Support hiring of employees
Second order effect		Not mentioned	Increase company workforce
Third order effect		Not mentioned	Increase employment
Evaluation method			
Evaluated effects	The financing process through which subsidies are distributed to enterprises	The study analyses the financing procedures through which state aid is distributed. It applies statistical analyses so that the factors that influence KTM decision-making are identified	The effects of labour subsidies on employment, investments and performance, controlled for firm size, region, etc. on the basis of a subsidised group and a control group.
Tools and techniques used for structuring the evaluation study	Causal model (but remains implicit)		Intervention Logic Analysis
Tools and techniques used for data collection	Secondary Application data	Secondary data Questionnaire surveys	Secondary data from the Tax Office
Tools and techniques used for data analysis	Bivariate and multivariate analysis Logistic regression	Bivariate and multivariate analysis Factor analysis Variance analysis Logistic regression	Regression analysis with control variables
Tools and techniques used in the judgement phase	Control groups		

Country	*United Kingdom*	*United Kingdom*	*United Kingdom*
Reference to the evaluation study	Wren and Storey (2002)	PACEC 2001	North et al., 2001
Title of the evaluated aid scheme	Consultancy Initiatives scheme of the UK Enterprise Initiative	Smart, including SPUR	Public sector support for innovating SMEs
Aid instrument evaluated	Subsidised advice for marketing	Grants	Smart scheme, Business Links, 'Teaching Company'-scheme, Business and innovation Centres
Type of evaluation	Ex-post	Ex-post	Ex-nunc
Annual expenditure	£ 275 million between 1994 and 1998	€ 15 million	Not mentioned
Policy objective of the aid scheme: *First order effect*	To improve the competitiveness of SMEs	Supporting near-market R&D projects by SMEs	To help overcome innovation barriers
Second order effect	Market failure of asymmetric information	Stimulate company innovation	Stimulate company innovation
Third order effect	Raise UK economic performance	Increase innovation / R&D Economic growth	Growth, increase productivity and creating jobs
Evaluation method	Statistical approach		Statistical comparison with control group
Evaluated effects	The direct effect of the support on the sales turnover of firms, employment, and survival	Effect of Smart on technological innovation, extent to which market failures have been addressed, value for money, proportion of successful projects, economic rationale for the scheme, reach of the scheme, additionality, etc.	The extent to which existing policy instruments are addressing the needs of firms and whether there are some gaps in provision which public support needs to address.
Tools and techniques used for structuring the evaluation study	*no specification of tools used*	Intervention Logic Analysis	Intervention Logic Analysis
Tools and techniques used for data collection	Primary data from questionnaire Survey and a follow-up telephone survey of non-respondents Secondary data from the DTI Offices	Secondary data Questionnaire surveys Interviews	Telephone surveys and face-to-face surveys
Tools and techniques used for data analysis	Bivariate probit model with sequential selection Parametric analysis Regression Chi-square ML probit equation	Coding and abstraction Univariate analysis Bivariate and multivariate analysis	Simple statistics such as cross tabs/frequency counts Control group
Tools and techniques used in the judgement phase	Control groups	Cost-benefit analysis Control groups	Comparing business' needs and government supply of innovation support.
Remarks	The study takes account of selection bias, whether through self-selection or agency selection.		

Country	*United Kingdom*	*United Kingdom*	*United Kingdom / Republic of Ireland*
Reference to the evaluation study	Fraser (2003)	Cowling and Mitchell	Rope, and Hewitt-Dunas (2001)
Title of the evaluated aid scheme	Investors in People	Small firm loan guarantee scheme (SFLGS)	Local Enterprise Development Unit (LEDU) and Small Business Programme (SBP)
Aid instrument evaluated	Advice, consultancy and other non-financial aid	Guarantees	Grants, sometimes augmented with loan guarantees and interest subsidies
Type of evaluation	Ex-nunc	Ex-post	Ex-post
Annual expenditure	30 million pounds	Not mentioned	Northern Ireland: 20 million pound, Republic of Ireland: unknown
Policy objective of the aid scheme: *First order effect*	To encourage employers to enhance their commitment to training and workforce development	Loan guarantee to banks for loans to small business	
Second order effect	Small business growth through a more skilled and productive workforce	Provide capital for growth; Stimulate SME growth rate	Support small business sector
Third order effect	Economic growth through a more skilled and productive workforce	Economic growth	Fighting unemployment, fighting market failures (access to finance).
Evaluation method	Econometric modelling and empirical testing		Econometric modelling, using Heckman and secondary data
Evaluated effects	The enhanced performance of participating small firms		Impact on assisted firm employment growth, profitability and turnover growth
Tools and techniques used for structuring the evaluation study	no specification of tools used		*no specification of tools used*
Tools and techniques used for data collection	Secondary data from the programme management database	Secondary data	Secondary data from the CAM project
Tools and techniques used for data analysis	Statistical: Multiple regression analysis Econometric model		Econometric modelling and testing the model on the basis of secondary data Chi-square tests
Tools and techniques used in the judgement phase	No specific techniques used	Comparing default specifications with actual results in a large loans database	Subjective judgement of the outcome of the econometric model

Country	*Belgium*	*Norway*	*USA*
Reference to the evaluation study	Clarysse et al (2004)	STEP 2004	Cooper 2003
Title of the evaluated aid scheme	IWT-Vlaanderen	EU 5th Framework programme	Small Business Innovation Research Programme (SBIR)
Aid instrument evaluated	Grants	Grants Advice, consultancy and other non-financial aid	
Type of evaluation	Ex-post	Ex-post	Ex-post
Annual expenditure		€ 55 million (€ 274 in 5 years, plus € 249 EU subsidies)	Ten federal agencies must designate 2,5% of their extramural R&D budget to small businesses, accounting for a total sum of $ 1,1 billion in 1999
Policy objective of the aid scheme: *First order effect*		Support R&D investments	Strengthen the role of small firms in federally funded R&D
Second order effect		Stimulate company innovation To promote R&D among SMEs in particular	Increase private sector commercialisation of innovations derived from federal R&D
Third order effect		Increase innovation / R&D Economic growth	Increase innovation / R&D Economic growth
Evaluation method			
Evaluated effects	The authors develop a survey tool to measure input, output and behavioural additionality of R&D grants to (SM and L) enterprises. The method can complement econometric evaluations to provide insight into the question why and how companies apply for R&D grants.	Norwegian participation in 5FP; economic return; effectiveness, measures as personnel competences; participants' valuation of the support systems; addistionality, synergy, participation barriers, impact on sectoral competitiveness.	Meeting federal research needs with small businesses / fostering commercialisation of federally funded research / supporting innovation by addressing a gap in early-stage financing
Tools and techniques used for structuring the evaluation study			
Tools and techniques used for data collection	Questionnaire	Secondary data (application data of the Norwegian participants in 5FP) Questionnaire surveys Interviews	Indicators from literature and other sources
Tools and techniques used for data analysis		Coding and abstraction Data matrices Frequency counts	
Tools and techniques used in the judgement phase		Limited and mainly qualitative benchmark of the performance of the Norwegian participation with other countries.	

Country	*Austria*	*Sweden*	*Central and Eastern Europe*[27]
Reference to the evaluation study	Gerhardter and Gruber 2001	Bager-Sjögren 2003	Bateman 2000
Title of the evaluated aid scheme		SME employee training in the frame of the EU Structural Funds Objective 4 in Sweden	Business Support Centres in Transition Economies
Aid instrument evaluated	Grants	Grants (50% co-funding of training)	Basic consultancy, business advice and support
Type of evaluation		Ex-post	Ex-nunc
Annual expenditure	Not mentioned	€ 50 million on average	Not mentioned, but in the magnitude of billions of euros
Policy objective of the aid scheme: *First order effect*	Stimulating self-supporting regional development	To pursue skill developments both in the labour force and in SMEs	To counter the rise of unemployment in the transition phase of the former communist countries in Central + Eastern European and act as the local engine for SME development
Second order effect		Facilitating the adaptation of workers of either sex to industrial changes and to changes in production systems	
Third order effect		Economic development	
Evaluation method		Quantitative impact assessment of the effect of the support on competitiveness, operationalised as productivity, gross profit margin and capital structure. Comparison of defined suitable treatment and control groups.	Process description
Evaluated effects	Innovativeness, impulses to regional development, impulses to create (specialized) consultancy services	Whether the objectives of the scheme have been fulfilled	The role of the Business Support Centres for SME development
Tools and techniques used for structuring the evaluation study		Econometric model	None
Tools and techniques used for data collection	Interviews Literature search	Secondary data based on annual financial reports; telephone survey among 1,500 randomly selected entrepreneurs (including control group), controlling for selection bias.	Purely qualitative: literature research and deduction
Tools and techniques used for data analysis	Document analysis	Regression analysis	Document analysis
Tools and techniques used in the judgement phase		Impact analysis	Qualitative cost-benefit analysis

[27] Albania, Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Poland, Romania, Slovakia and Slovenia.

Country	*European Union*	*European Union*	*European Union*
Reference to the evaluation study	Deloitte & Touche 2003	Röller et al, 2001	European Commission 1999
Title of the evaluated aid scheme	35 Financial assistance schemes for SMEs	Various aid measures	Structural Funds aid to SMEs
Aid instrument evaluated	Grants	All types of instruments	Grants; Equity participations; Guarantees; Advice, consultancy, training; and other non-financial aid
Type of evaluation	Ex-post	Ex-post	Ex-post
Annual expenditure		€ 3 billion	500 – 600 million ECU
Policy objective of the aid scheme: *First order effect*	Support R&D investments; increase the volume of external finance available; support SMEs with international trade development, partner search and joint venture investments; support SME vocational and management training; increase SME adoption of environmentally friendly technologies	Various	Specific objectives
Second order effect	Improve SMEs' employee quality, internationalisation, healthy financial structure, access to information and advice and environmentally friendly production methods	Various	
Third order effect		To stimulate economic performance, to remedy market failures	Regional assistance; combat unemployment; develop rural economies and support peripheral regions.
Evaluation method	Qualitative consultation of experts, programme managers and interest groups	Ratio approach, correlation approach and econometric approach	Combination of primary and secondary data
Evaluated effects	The reasons for being, relevance, coherence, effectiveness and overall efficiency of the 35 evaluated measures	SME share in total economic output	Jobs created and saved by Structural Fund intervention
Tools and techniques used for structuring the evaluation study	No explicit structuring with regard to expected causalities	No explicit structuring with regard to expected causalities	Conceptual- theoretical framework, based on SMEs' needs
Tools and techniques used for data collection	Desk research and Expert panels / DELPHI Consultation of programme managers and interest groups	Existing EUROSTAT data	Document analysis of existing evaluations, Interviews, Questionnaire[28], Telephone survey, Project reviews

[28] Questionnaire used mainly for user satisfaction measurement, hardly for the evaluation of impact of Structural Funds on SME development

Tools and techniques used for data analysis	None mentioned	Ratio analysis (SME share in the number of employees, or turnover, divided by the discounted amount of state aid), correlation between the SME share in the turnover and the discounted amount of state aid and following econometric model: ($P_i=\alpha_{0i}+\alpha_1 \log(A_i)+\alpha_2 \log(X_i)+\varepsilon_i$)	Frequency counts Estimate based on an elaborate calculation scheme
Tools and techniques used in judgement phase	None mentioned	Comparison of effectiveness of instruments.	Limited and mainly qualitative judgement

Note: The Analytical Hierarchical Programming (AHP)-method mentioned in the Dutch study by Vreeker et al (2001), is a method for assigning weights to aspects, criteria, phenomena, etc. The method is based on pair-wise comparisons of two items, whereby a respondent is asked to give his judgement about the relevance or preference of either of the two items, on a 19-point scale. Based on these pair-wise comparisons, the relative weights can be calculated for each of the aspects and the consistency of the respondents' answering behaviour can be judged (for a more detailed explanation see Saaty, 1980).

Table 4 Example of an AHP-questionnaire

Please tick your preference for the following items:

	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	
Coffee	o	o	●	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	Tea
Coffee	o	o	●	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
Chocolate																				
Tea	o	o	o	o	o	o	o	o	o	●	o	o	o	o	o	o	o	o	o	
Chocolate																				

The respondent above has a strong preference for coffee, and judges tea and chocolate equally desirable. His answering is consistent.

6.4 Conclusion

From the above overviews we can conclude that not many good practices use formal procedures for structuring the evaluation approach (or at least they do not document this in the final report). Nor is it always clear what the objectives (i.e. intended effects) of the evaluated support measure are.

The above overview demonstrates that, in practice, not one single standard method for comparison of the effectiveness of different state aid schemes is used. Only one publication (European Commission, 1999) quantifies the effects of such diverse measures as grants, training and consultancy services at one time. The conclusion is that there is not 'one best way to evaluate state aid to SMEs'. The experts that we have interviewed support this conclusion.

PART III:

SYNTHESIS

7 Selected methods

There are many techniques that can be used for evaluating the effectiveness of state aid. Based on our research, we arrive at a classification of 8 different 'basic' methods. This 'long list' comprises both qualitative and quantitative methods, and methods with and without control groups. The 8 basic methods reflect an increasing level of sophistication.

7.1 The 8 'basic' methods

As we have seen there are several techniques for data collection, data analysis, and for judgement of the facts found. Since the objective of this study is to arrive at a limited number of methods that best measure the effectiveness of state aid to SMEs, we have to draw up a long-list of methods that will be scored by experts in the field of evaluation research. This requires condensing the large number of the evaluation approaches found into a systematic classification. Synthesis of the existing evaluation studies results in the following overview:

		Data analysis	
		Qualitative	Quantitative
Data collection	Qualitative	– Interviews + coding and abstraction – Case studies + document analysis – Literature search + document analysis	
	Quantitative		– Existing statistics + statistical analysis (sometimes econometric analysis) – Questionnaire surveys + statistical analysis (sometimes econometric analysis) – Application data + limited statistical analysis

According to various sources (including the European Commission, 1997a), evaluations should not rely on one single data collection technique. When using various techniques side by side, the strengths of one technique can balance the weaknesses of another. The same applies to techniques for data analysis.

Venetoklis (2000) gives some additional recommendations with regard to the design of evaluation studies:

– Data regarding the development and operation of firms should be gathered directly from the firms at frequent time intervals. Financial statements as well as other, more detailed information, is welcome (e.g. balance sheet, profit & loss accounts, number of personnel, export as % of sales, R&D expenditures, etc.). This data should be gathered not only for recipient firms, but also for non-recipient firms and firms whose applications were rejected.
– Estimations of subsidy impacts should not be asked directly from the recipient firms only (neither from the non-recipients for that matter).
– Also, a control group of firms (based on the subsidised firms' industrial sector, geographical location, operating markets, size, etc.) should be created and monitored.
– Simple descriptive analyses that calculate differences of indicators between time periods in specific sub-groups of firms (i.e. recipients and non-recipients) may not completely isolate the net impact of the subsidies, but can give some indications on certain trends.
– Analysis of the effects of a support measure should focus not only on the size of the intended impact, but should also include other costs and benefits associated with the support scheme.

Based on the above analyses a long-list of existing and alternative methods is provided below. The composition of this list was based on frequently used combinations in SME state aid evaluation practice, literature from other evaluation fields (such as R&D, environment and infrastructure), and interviews with experts in the field of policy evaluation. Based on literature research, member state contributions and interviews, complemented with our own analyses, we arrive at the following eight 'basic' evaluation methods. The methods are presented in order of increasing sophistication:

1 Qualitative description of first order effects, based on literature, interviews and case studies

2 Ex-ante evaluation of the policy theory, i.e. scrutinising the intervention logic by assessing the expected working of mechanisms that the support measure puts into operation, based on literature (desk) research, case studies and/or experts

3 Quantitative evaluation of the reach of the programme, based on a statistical analysis of the assisted firms and a control group using existing statistical databases

4 Quantitative evaluation of the first order effects, based on a survey among assisted firms

5 Quantitative evaluation of the first and second order effects, based on a survey among assisted firms and a control group

6 Quantitative evaluation of the (first, second and) third order effects, based on an econometric model that uses a variety of qualitative and quantitative data sources, among which literature research, a survey of assisted firms and a control group, secondary data from the tax Office and/or Statistics Office, interviews and/or case studies

7 Ratio analysis (such as net average subsidy per job created), based on quantitative analyses of secondary data (e.g. application data and economic statistics)

8 Goal free programme evaluation; a combination of qualitative and quantitative approaches, that combines statistical analyses and case studies in order to determine to what extent a support programme[29] has been able to address the targeted market failures, incorporating such effects as displacement and additionality. The counterfactual situation is determined through the use of a control group, whereby the participant group is analysed for selection bias, for example using Heckman's 2-step procedure[30]. The control group could be constructed using the matched sample-approach.

Explanation

The eight 'basic' methods listed above are explained in more detail below.

29 A support programme consists of a pallet of support measures. Programme evaluation does not look at the effects of only one support measure, but assesses the individual, interaction and collective effects of an integral set of policy measures (the policy programme).

30 The Heckman 2-step adjustment procedure formulates a single equation to explain the selection procedure and then, formulates a second equation that explains performance change on the basis of the factors included in the selection equation (also see: Storey, 2002). For a more detailed explanation see annex III.

1	Qualitative description of first order effects.	This comprises the first 3 steps of Storey's 'six steps to heaven'. Can be based on interviews among recipients, enquiring after their view of the impact of the support.
2	Ex-ante evaluation of the policy theory.	The policy theory is the set of assumptions that the policy makers have, and on which they base their expectations of the working of the aid measure. This method evaluates the effects prior to the implementation of the state aid measure by testing these assumptions.
3	Quantitative evaluation of the reach of the state aid.	This method does not collect new (primary) data, but bases its analysis solely on existing (secondary) data such as national statistics.
4	Quantitative evaluation of the first order effects, based on a survey among assisted firms.	Zero-order effects: has the state aid really been distributed / provided, First-order effects: has the stimulated/financed activity really led to an implementation,
5	Quantitative evaluation of the first and second order effects, based on a survey and use of a control group.	Second order effects include effects of the support on the (behaviour of) company. Second-order effects measure whether the implementation led to a better competitive position, *e.g*. more innovative products, more employment, more exports, etc. This method builds on the previous method. It includes the same items as method 4, plus an evaluation of the effects of the aid on the assisted company. Further, a control group is used.
6	Quantitative evaluation of the (first, second and) third order effects, based on an econometric model and a variety of qualitative and quantitative data sources.	Third-order effects measure the overall economic effect of a state aid measure (e.g. what the contribution of one euro state aid has been to the competitiveness of the Member State). This method builds on method 5. The third order effect is added, and more sources are used. Qualitative information is gathered to supplement and interpret the quantitative insights (why?, how?, etc.).
7	Ratio analysis, based on quantitative analyses of secondary data.	This approach calculates ratios, for example the amount of subsidies per extra job created or the leverage in percents of the aid. Drawback of this approach: it is difficult to link ratio's to the targeted market failure.
8	Goal free programme evaluation.	A support programme consists of a pallet of support measures. Programme evaluation does not look at the effects of only one support measure, but assesses the individual, interaction and collective effects of an integral set of support measures (the support programme). This approach builds on method 6, adding the interaction element with other state aid programmes. The selection bias is also studied, and researchers are more open for other effects than the intended effects only (such as side effects). The Heckman 2-step adjustment procedure could be used to control for sample selection bias. This procedure formulates a single equation to explain the selection procedure and then formulates a second equation that explains performance change based on the factors included in the selection equation (also see: annex III).

It is important to note that there are certain linkages between the proposed methods. The most important notion is that methods 4, 5, 6 and 8 build on each other. In other words, method 6 contains all elements of method 5, plus some extra features (i.e. the assessment of the third order effects and the consultation of more data sources). In this way, the four methods represent a modular approach, where an evaluation can start at the level of method 4, and be extended up to the level of method 8. This notion is especially important in the view of making (international) comparisons of the effectiveness of various state aid types possible: since method 8 also contains all the elements of method 5, the results of both types of evaluation studies can be compared, at the level of method 5.

7.2 Clusters of evaluation methods

According to their characteristics, the evaluation methods can be grouped into three main clusters:

1	Qualitative methods	- Method 1 - Method 2
2	Quantitative methods without a control group	- Method 3 - Method 4 - Method 7
3	Quantitative methods with a control group	- Method 5 (quantitative only) - Method 6 (including also qualitative information) - Method 8 (including qualitative information and interaction effects)

In general, qualitative methods are simplest and cheapest, but have limited explanatory power. Quantitative methods that use a control group are best suited for measuring effectiveness, but are also more expensive and complex (see figure 13).

Figure 12 Trade-offs in evaluation designs

Source: EIM, based on OECD, 2004

8 Assessing evaluation methods

To condense our selection to a limited number of methods that best measure the effectiveness of state aid we applied a multi-criteria analysis, which was operationalised in the form of a questionnaire. The questionnaire was completed by a number of leading scientists, all experts in the field of evaluation state aid to SME. Based on the results we selected three methods that are best for measuring the effectiveness of state aid to SMEs. These methods are best for quantifying the effects of state aid, incorporating such effects as deadweight, additionality and displacement.

8.1 Criteria

As previously stated the 8 selected 'basic' methods differ in their level of sophistication. In general, method 1 is the least sophisticated approach for determining the effectiveness of state aid to SMEs, while method 8 is the one with the highest sophistication. However, the complexity and costs increase too (such a trade-off is illustrated in figure 13).

In particular, selecting methods that are to be applied everywhere, under each and every circumstance, and still produce valuable and reliable results, requires a trade-off of the various aspects of these methods, such as soundness, cost, data availability and applicability under all circumstances. Thus, the methods have to be compared for various characteristics.

An appropriate scientific method for comparing the characteristics of various items is a multi-criteria analysis. For the purpose of this study, we designed a multi-criteria analysis with which the methods were assessed.

In close consultation with the principal of the study, 13 criteria were identified against which the methods were to be judged. Two types of criteria were selected: *general* and *specific* criteria. General criteria are the requirements that can be applied to evaluation in general (among others, see Prince 1998 and Algemene Rekenkamer 2001). Specific criteria are the criteria that the principal finds important and that were raised during the Member State meeting[31] (among others, see the Tender Specifications). The criteria are:

1 Whether or not the method will **quantify** the effect of state aid.
2 Whether or not the method can **prove the causal relations** between the state aid and the effects.
3 Whether or not the method is **objective/neutral.**
4 Whether or not the method is **generally applicable.**
5 Whether or not the method is **transparent.**
6 Whether or not the method is **cost-effective.**
7 Whether or not the method is **time-effective.**
8 Whether or not the **required data** for this method are available or can be gathered easily.
9 Whether or not the method can take account of **deadweight.**
10 Whether or not the method can take account of **displacement.**
11 Whether or not the method can take account of **additionality.**

31 One criterion, namely whether the method can control for whether the support is direct or indirect, is left out of consideration here, since this involves comparing the effectiveness of several schemes, some direct and other indirect.

12 Whether or not the method can control for **company-specific background characteristics** (such as the beneficiary's stage in the life cycle, company size, sector, etc.).

13 Whether or not the method can control for **locational and other external factors** (such as is the company is located in a region eligible for regional state aid, the level of regulatory, technical and/or administrative burdens, the overall fiscal burden (corporate or personal income tax), the business cycle, etc.).

Explanation

Below, we further explain the selection criteria:

Whether or not the method:

1 will **quantify** the effect of state aid	It seems important to quantify the effects, in order to arrive at a sound judgement of the effectiveness of a state aid measure.
2 will **prove the causal relations** between the state aid and the effects	Causality is very important for proving that that measured effects are initiated by the state aid.
3 is **objective/neutral**	It is important that the evaluation scheme does not focus on one side of the story, or includes only the positive effects. The evaluation should give a balanced view of both the positive and negative (side) effects of the state aid.
4 is **generally applicable**	General applicability means that the method must be applicable to all forms of state aid, from tax reductions to equity participation and from subsidies to soft loans.
5 is **transparent**	Whether or not the reader of the evaluation report easily understands how the evaluator arrived at his conclusions, and that the study could be repeated quite easily.
6 is **cost effective**	From the policy maker's point of view, it is important that the costs of the evaluation study are in line with the magnitude of the state aid instrument. A general rule of thumb for considering the reasonable costs of evaluation is somewhere between 2% and 10% of the overall project budget. The method should be affordable, yet achieve good results. This criterion should judge the costs of the method relative to the costs of the other methods. Combined with the scores of the other criteria, this criterion gives input for the cost-effectiveness ratio of the method.
7 is **time effective**	From the policy maker's point of view, it may be important that the results of the study become available within a reasonable time period. This criterion should judge the lead-time of the method, relative to the lead-time of the other methods. Combined with the scores of the other criteria, this criterion gives input for the time-effectiveness ratio of the method.

8	**required data** are available or can be gathered easily	Data availability is a major concern. The European Commission made clear that the recommended evaluation method (or methods) should be applicable in all 25 EU Member States. The availability of secondary data (such as statistics) in particular is a potential problem. Especially control variables and internationally standardised data will be difficult to gather. Alternatively, primary data can be gathered using case studies, analysing application data, conducting questionnaires or having interviews. Since these primary data can be collected more easily, more consistently, and can be tailored to the specific needs of the evaluator, primary data may be preferred and prevalent to existing data.
9	can take account of **deadweight**	Deadweight is the autonomous development. State aid evaluations always have to address the question whether the effects that they observe would also have taken place without the state aid measure. For example: would these jobs also have been created if the company had not been supported?
10	can take account of **displacement**	If jobs have been created, the question is whether they have replaced jobs at companies that did not receive assistance. This effect is called displacement. Can the method measure displacement?
11	can take account of **additionality**	An important question for policy makers is, whether state aid really affected the SMEs (if it has an incentive effect). Does the aid induce additional investments from entrepreneurs that would not have been made without the aid? This is called additionality.
12	can control for **company-specific background characteristics**	Company specific characteristics, such as the beneficiary's stage in the life cycle, the recipient's size or the sector in which it is active may influence the effectiveness of state aid. Can the method demonstrate this difference?
13	can control for **locational and other external factors**	Besides company specific factors, locational factors and other external factors could also influence aid effectiveness. These include, among others, whether or not the company is located in a region that receives regional aid, the level of regulatory, technical and administrative barriers, and the overall fiscal burden in the Member State. Can the method take such differences into account?

8.2 Results of the assessment

Relative weights of the criteria

Although all the 13 criteria identified were relevant, they were not equally important. Some criteria were more important than others. During our consultation, we asked the experts to give their judgement of the relative importance of the criteria, by distributing a total of 100 (percentage) points over the 13 criteria. The average results of the weighting exercise showed that quantifyability, data availability, causality and control variables are judged as being the most important criteria. Together, these made up 50% of the final score of the method. Objectivity, general applicability, cost effectiveness, deadweight,

and additionality were judged by the experts as being of moderate importance. Displacement, transparency and time effectiveness were judged to be the least important criteria (see table 5).

Table 5 Relative weights of the criteria (in %), as judged by the evaluation experts

Whether or not the method:	*Relative weight*
1 will **quantify** the effect of state aid	10,375
2 will **prove the causal relations** between the state aid and the effects	9,75
3 is **objective/neutral**	7
4 is **generally applicable**	7,4375
5 is **transparent**	5,8
6 is **cost effective**	6,5
7 is **time effective**	4,875
8 **required data** are available or can be gathered easily	12,625
9 can take account of **deadweight**	6,4625
10 can take account of **displacement**	5,2125
11 can take account of **additionality**	6,3375
12 can control for **company-specific background characteristics**	9
13 can control for **locational and other external factors**	8,625
Total	**100**

Source: EIM, based on expert judgements, 2004

Scores of the individual methods

After assigning weights to the criteria, the individual methods were scored on each of the criteria. The respondents were asked to assign a score between 0 and 10 (both included, where 0 means that the method scores very low on a certain criterion, and 10 means that the method scores very high). The unweighted average scores are provided in the table below. The table shows that the more sophisticated methods are better at quantifying the effects of state aid, but are less time and cost effective than the less comprehensive methods.

Table 6 Average scores of the methods, as judged by the evaluation experts

Whether or not the method:	*Method 1*	*Method 2*	*Method 3*	*Method 4*	*Method 5*	*Method 6*	*Method 7*	*Method 8*
1 will **quantify** the effect of state aid	1,9	2,6	6,5	6,3	7,8	8,6	6,9	8,4
2 will **prove the causal relations** between the state aid and the effects	2,5	3,0	4,3	3,1	6,1	7,5	4,8	8,3
3 is **objective/neutral**	5,1	5,5	7,6	5,4	7,5	7,8	6,6	8,0
4 is **generally applicable**	8,6	7,1	5,5	8,3	8,0	6,0	6,1	7,5
5 is **transparent**	6,0	5,3	6,5	7,3	8,3	6,1	7,0	7,6
6 is **cost effective**	6,8	6,4	6,4	6,1	5,1	3,6	6,5	2,5
7 is **time effective**	6,5	6,4	5,5	6,0	5,4	3,1	6,4	2,9
8 **required data** are available or can be gathered easily	7,8	7,3	4,6	6,3	5,9	5,5	5,3	5,1
9 can take account of **deadweight**	5,3	3,5	6,4	3,6	7,4	7,8	3,4	8,1
10 can take account of **displacement**	3,3	2,9	3,4	1,8	3,4	6,1	2,5	7,1
11 can take account of **additionality**	4,3	3,5	5,5	3,5	6,9	8,1	3,0	8,1
12 can control for **company-specific background characteristics**	4,0	2,6	6,9	6,4	7,8	7,6	4,4	8,6
13 can control for **locational and other external factors**	4,3	2,8	6,3	6,3	7,3	7,6	3,1	8,5

Source: EIM, based on expert judgements, 2004

Judgement criterion: sum of the weighted scores

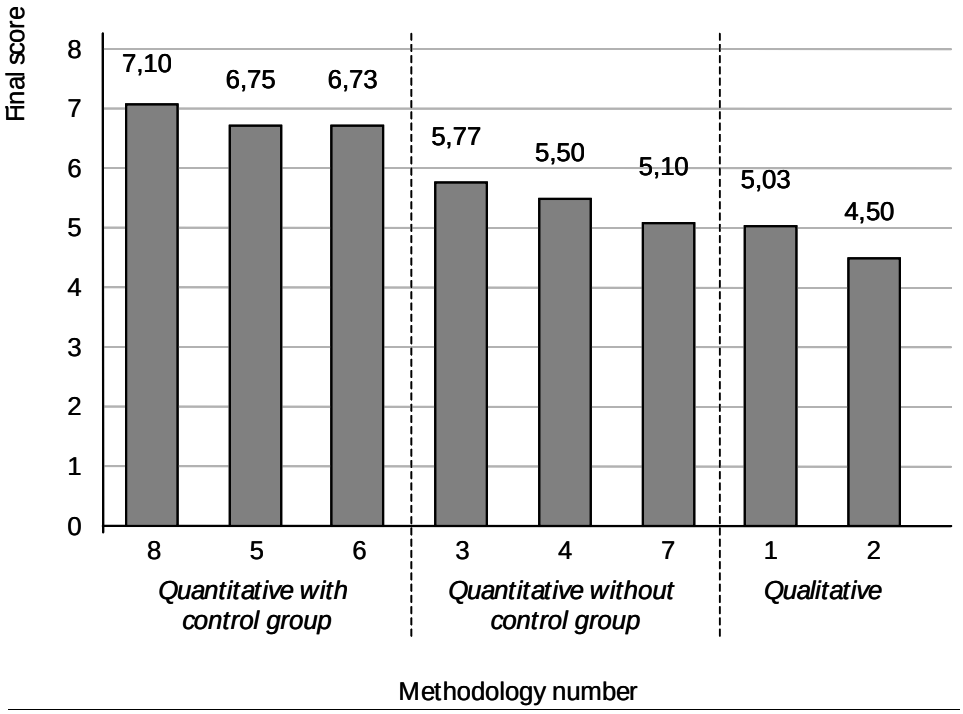
The aim of this study was to select a limited number of methods that are best for evaluating state aid effectiveness. We needed clear judgement criterion to select these methods. and for this purpose we calculated a final score for each of the methods. This final score is based on the sum of the weighted criteria scores of the methods. The final score is calculated with the following equation:

$$FS_i = \sum_{j=1}^{13} w_j c_{ij}$$

where FS_i is the final score for method i, c_{ij} is method i's average score on criterion j, w_j is the relative weight of criterion j. Since there are 13 criteria, j ranges from 1 through 13. The methods with the highest final score (the sum of the weighted scores on each criterion) will be selected.

The final scores for each of the methods are presented in figure 14:

Figure 13 Graphical presentation of the scores (in descending order)



Source: EIM, based on expert judgements, 2004

The ranking shows that the experts judged the three quantitative methods with control groups[32] (method 5, 6 and 8), as the best methods for assessing the effectiveness of state aid to SMEs. This is because of their ability to quantify the effects, their objectivity and their ability to prove causal relationships. These methods are best able to demonstrate such effects as the deadweight, displacement and additionality of state aid.

The quantitative methods without control groups (methods 3, 4 and 7) are valued less appropriate for measuring the effectiveness of state aid. According to their ranking, the qualitative methods (1 and 2) are least suited for a thorough evaluation of the effectiveness (for a more detailed explanation, see table 7 below). Although they score highest, on the availability of data, it is not surprising that they are least suitable for quantifying the effects.

32 See section 7.2 and Figure 12 in particular for the classification of the methods.

Table 7 Relative weighted average scorings of the methods, as judged by the evaluation experts

Whether or not the method:	*Method 8*	*Method 5*	*Method 6*	*Method 3*	*Method 4*	*Method 7*	*Method 1*	*Method 2*
will **quantify** the effect of state aid	0,87	0,80	0,89	0,67	0,65	0,71	0,19	0,27
will **prove the causal relations** between the state aid and the effects	0,80	0,60	0,73	0,41	0,30	0,46	0,24	0,29
is **objective/neutral**	0,56	0,53	0,54	0,53	0,38	0,46	0,36	0,39
is **generally applicable**	0,56	0,60	0,45	0,41	0,61	0,46	0,64	0,53
is **transparent**	0,44	0,48	0,36	0,38	0,42	0,41	0,35	0,30
is **cost effective**	0,16	0,33	0,24	0,41	0,40	0,42	0,44	0,41
is **time effective**	0,14	0,26	0,15	0,27	0,29	0,31	0,32	0,31
required data are available or can be gathered easily	0,65	0,74	0,69	0,58	0,79	0,66	0,98	0,92
can take account of **deadweight**	0,53	0,48	0,50	0,41	0,23	0,22	0,34	0,23
can take account of **displacement**	0,37	0,18	0,32	0,18	0,09	0,13	0,17	0,15
can take account of **additionality**	0,51	0,44	0,51	0,35	0,22	0,19	0,27	0,22
can control for **company-specific background characteristics**	0,78	0,70	0,69	0,62	0,57	0,39	0,36	0,24
can control for **locational and other external factors**	0,73	0,63	0,66	0,54	0,54	0,27	0,37	0,24
Total score	**7,10**	**6,75**	**6,73**	**5,77**	**5,50**	**5,10**	**5,03**	**4,50**

Source: EIM, based on expert judgements, 2004

8.3 Discussion

Method 1 Qualitative description of first order effects

The first evaluation method is based on qualitative research techniques. The strength of qualitative techniques is that they describe the processes behind the results. However, its weakness is that it is unable to tell *to what extent* a state aid measure achieves its goals. Another weakness is that this evaluation method focuses on the first order effects only. First order effects (such as an investment, training or advice), are usually only the means through which state aid goals (firm survival, more innovation, economic growth) should be achieved.

In general, experts judge this method as being widely applicable, relatively cheap and quick. Also, the necessary information is available or can be collected easily. However, the method is not very neutral, does not allow for quantifying the state aid effects, and is unsuitable for proving causal relations between the aid measure and the effects. Also, this method cannot take account of side effects (such as additionality) and background characteristics. As such, this method fails some of the most important conditions for measuring the effectiveness of state aid to SMEs.

Experts' views: "*This method will actually tell you very little about the impact of state aid on SMEs*."
"*Too much EU evaluation is based on this kind of method, but it is really not sufficiently rigorous to give worthwhile results.*"
"This method reveals the process rather than the effectiveness"

Method 2 Ex-ante evaluation of the policy theory

Ex-ante evaluation, which constitutes the second evaluation method, is a good instrument for scrutinising a state aid policy prior to its implementation. Ex-ante evaluations judge the value, effectiveness and efficiency by assessing the *expected* effects of the policy. Therefore, ex-ante evaluations can be of use for optimising policies before implementation.

Experts judge this method as widely applicable and the required data (if any) can be found without many problems. It is also quite cost and time effective. However, the same reserves apply as to method 1, namely that the method is not very neutral, does not allow for quantifying the state aid effects, and is unsuitable for proving causal relations between the aid measure and the effects. Also, this method cannot take account of side effects (such as additionality) and background characteristics.

Experts' views: "*Again, this method is often used - but the ex-ante study is rarely followed up by rigorous ex-post evaluation on the actual effects of the state aid."*
"Although suited before introducing the state aid measure, it is less suited for measuring the effects really taken place after implementing the measure"

Method 3 Quantitative evaluation of the reach of the state aid

The third method does not challenge the effects that the scheme invokes, but instead it checks whether or not all the companies that are entitled to support, have received this assistance. As such, this method does not really evaluate the effectiveness: it can neither tell what the results of the support measure are, nor how these results are achieved. The method may arrive at interesting additional information and may yield insights that can explain the support measure's effects, but this method does not evaluate the effectiveness itself.

Experts judge this method as being more neutral and transparent than methods 1 and 2, and also being better able to take account of control variables. However, three of the most important characteristics are judged with more reservations: quantifyability is moderate and causality and data availability are low.

Experts' views: "*The quantification is dependent on the kind of variables in the statistical databases and the way the data are stocked in the database (do they enable statistical analysis)*."
"*The weakness with this method is that it fails to provide the 'reality check'. It does not gather insights on the performance of the scheme based on its actual performance on the ground*."

Method 4 Quantitative evaluation of the first order effects, based on a survey among assisted firms

The fourth method does actually look at the effects that a support measure discharges into. The method is based on a survey that consults the assisted firms, in order to find out whether or not the support measure contributed to performing some desired activity (e.g. training employees or making a certain investment). The strengths of this approach are its simplicity and efficiency. The method scores highest on data availability since the data can be gathered quite easily by performing a survey among supported enterprises. However, the weaknesses include the lack of a benchmark or counterfactual (no control group), the relative subjectivity and the negligence concerning the second and third order effects.

Experts judge this method as transparent and generally applicable, medium cost and time effective and somewhat able to quantify. The method is relatively bad at taking account of important issues as displacement and additionality and cannot convincingly prove causal relations between the support and observed 'effects'.

Experts' views: "*Is not very useful because there is no control group*."
"*(…) A postal survey is likely to achieve very low response rates which would undermine the value of the method. Telephone or face to face interviews would be more likely to achieve higher response rates, but would be more expensive*."

Method 5 Quantitative evaluation of the first and second order effects, based on a survey and use of a control group

This method seems to measure what we want to know in a relatively effective manner. The second order effects (such as company survival, turnover growth, increased innovations, etc.) are taken into account, and these second order effects are usually the intended effects of the support measure. Also the approach (a survey with a control group) allows for quantifying the effects and is a relatively efficient way of collecting reliable data. Since new data are gathered, surveying allows for including control variables and background variables.

Experts judge this method as being good at quantifying the effects of a measure, as neutral, transparent and generally applicable. Also, if performed well, this approach allows for taking into account such issues as deadweight and additionality. However, the method is a little less efficient than the previous methods and also is moderate in taking displacement into account.

The experts' previous comments on survey method apply again here.

Method 6 Quantitative evaluation of the (first, second and) third order effects, based on an econometric model and a variety of qualitative and quantitative data sources

Method 6 comprises an econometric model based on a variety of sources. This has some major advantages: subjectivity of interpretation of the facts can be eliminated due to the multiple sources of information (this is called inter-subjective observation), taking account of other effects than the ones usually associated with the support measure, and paying attention to the (long term) ultimate benefits of a support measure. Also, the econometric model allows for 'fine tuning' in the sense that the evaluator can look what happens if certain variables of his model are altered. A drawback of this approach is that it requires detailed knowledge of econometrics, is not very transparent and less effective than the previous measures.

Experts judge this method as the one that is best able to quantify the state aid effects (also the unintended effects), as being objective and able to prove causality, and able to take account of deadweight, displacement, additionality and background characteristics as control variables. However, the method is quite laborious, and thus inefficient with respect to time and costs, and not so transparent.

Experts' views: "*This method is a quite heavy approach for measuring the effectiveness and places high requirements on data availability and economic skills. For most measures this might be a too heavy method*."
The experts' previous comments on survey method apply here too.

Method 7 Ratio analysis, based on quantitative analyses of secondary data

Ratio analysis is an interesting method, since it makes very clear some essential effects of a support measure. The strength of ratio analysis is that the effectiveness of a support measure can be expressed in one or a few single figures. The reverse to the medal is that ratio analysis does not allow for any nuance or shade. Another drawback of this approach is that it uses existing data sources only. Thus, the evaluator does not have control over the quality and level of detail of his inputs.

Experts value this method as transparent and being able to quantify state aid effects, yet they find it unfit to take account of relevant factors such as deadweight and additionality and judge it as moderate at demonstrating causality.

Experts' views: "*The quality of the evaluation depends on the quality of the secondary data. In addition, the results of this method can lead to the conclusion that the measure is not effective, because the macro-objectives do not correspond with the objectives of the entrepreneur*."
"*The weakness of this method, as with method 3 is the lack of the input from the on the ground real experience of the scheme. It is dubious if the data required for such analysis would be readily available (and in any case it would probably be incomparable across schemes).*"

Method 8 Goal free programme evaluation

Goal free programme evaluation is clearly the 'top of the bill' as concerns evaluation of a state aid programme. Its main advantages are that it also considers interaction effects between various state aid measures, and that it considers all types of effects and not only intended effects. Another advantage is that this approach does not only answer the question *how* effective a state aid measure or programme is, but also *why* it is so (in)effective. Disadvantages of this approach include the laboriousness and data constraints, for this intricate method demands high data quality. And it is clear that this method asks for a relatively long lead-time with corresponding high costs.

Experts view this method as the best method, since it can quantify the state aid effects, will prove causal relations, is transparent, objective and generally applicable, takes account of such phenomena as deadweight, displacement and additionality, and also allows for the use of various control variables. From an expert's point of view, this evaluation method is 'heaven' to speak with David Storey. However, compared to the other methods, this approach is relatively expensive and lengthy.

Experts' views: "*Attention should be paid to statistical fallacies (e.g. regression to the mean).*"
"*This method is probably the "ideal", but is expensive and time consuming.*"

Comparing the methods

It is clear that all methods have advantages and disadvantages. However, as shown in section 8.2, the experts have consistently given more weight to certain aspects of the methods, than to other aspects. As a result, methods that are best at quantifying the effects, incorporating data availability, causality and control variables are the ones most suited for evaluating state aid effectiveness. Table 7 shows that methods 5, 6 and 8 meet the requirements with regard to quantifying, causality and objectivity best, and additionally can also take aspects such as deadweight, additionality, displacement and control variables into account.

8.4 Conclusion

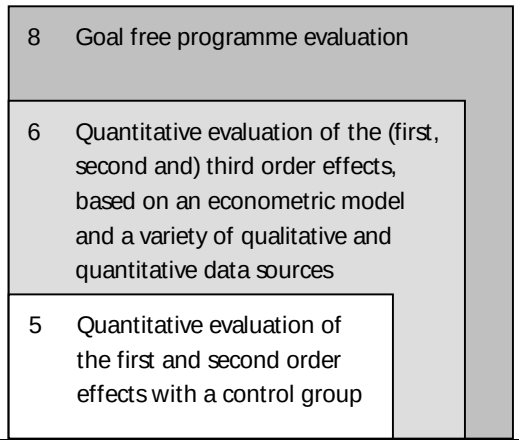
The results of the expert consultation show that there are three methods that are most suited for evaluating the effectiveness of state aid to SMEs (also see figure 14 and table 7). These methods are:

5 Quantitative evaluation of the first and second order effects.
6 Quantitative evaluation of the (first, second and) third order effects, based on an econometric model and a variety of qualitative and quantitative data sources.
8 Goal free programme evaluation.

The characteristic that sets these methods apart from the other approaches is the fact that methods 5, 6 and 8 will prove best the *causal relationships* between the support measure and the observed *quantified* effects, taking account of *deadweight*, *additionality*, *displacement* and *control variables*. The other methods are less suited for evaluating the effectiveness of state aid to SMEs because they score lower on these important requirements. It is hence that we recommend methods 5, 6 and 8 as the best methods for evaluation state aid to SMEs.

It has already been explained that methods 5, 6 and 8 build on each other. Method 8 contains all the elements of method 5 and 6, and method 6 contains the elements of method 5 (see figure 15). So, if two state aid measures have been evaluated using different methods, e.g. one with method 5 and the other with method 8, then the results of the evaluation studies can be compared, at the level of method 5.

Figure 14 graphical representation of the links between the selected methods



Each of the methods has advantages and drawbacks. In general, the simplest method (method 5) is the one most easily applicable and provides the required results (a measure of the support measure's effectiveness), but no insight in the mechanisms with which these results are achieved. The most intricate method (method 8) yields additional results such as reasons and suggestions for improvements, but is more expensive and difficult to perform. Annex IV provides one example from recent evaluation practice for each of the three recommended methods.

However, if the causality and side effects of the state aid measure are given less weight and the emphasis of the evaluation is more on improving the process than on measuring the effectiveness, than methodology 4 or 3 might be used.

Since the results of the selected methods are always comparable, we suggest that policy makers should choose the one method that best suits their circumstances, requirements and budget. This approach leaves policy makers enough room to tailor an evaluation to their specific circumstances and requirements, while guaranteeing that the minimum acceptable quality level is reached. For example, countries that have limited experience with evaluation or where existing data sources are scarce could

use method 5. As they gain more evaluation experience and evaluation is integrated into their policy processes they could ‘upgrade’ towards method 6. Ultimately, after building up considerable evaluation experience and developing longitudinal economic statistics, they could apply method 8.

9 Conclusions

Based on the information gained from our study, we are able to draw the following conclusions:

- Annually, a substantial percentage of the total state aid is allotted to SMEs.
- Evaluation of SME state aid is necessary both for public accountability, and for improving state aid effectiveness.
- Only a few of the EU-Member States systematically evaluate the effectiveness of state aid to SMEs. Most Member States evaluate only some of the state aid instruments or programmes. There are also Member States that do not evaluate at all.
- Effectiveness comprises more than merely achieving the policy goals. When evaluating the effectiveness, deadweight, additionality, and displacement are important aspects that should also be taken into account.
- The state aid evaluation process consists of four phases that are all influenced by the research question. These four phases are evaluation design, data collection, data analysis and judgment.
- A method for state aid effectiveness evaluation consists of a combination of data collection and data analysis methods.
- Eight basic types of state aid effectiveness evaluation methods can be distinguished, ranging from quite simple methods to intricate, comprehensive and expensive approaches. Each of the approaches has advantages and disadvantages.
- These eight basic types can be classified in three clusters: qualitative methods; quantitative methods without a control group; and quantitative methods with a control group.
- Only methods with a control group are able to address displacement, additionality and causality.
- The question which of the methods is the most suitable for evaluating the effectiveness of state aid to SMEs is answered in this study by using a multi-criteria approach that combines general and specific criteria. Experts scored the eight 'basic' methods against 13 criteria.
- Based on our research, we conclude that experts in the field of evaluation of SME policies judge the three methods in the cluster 'quantitative with control group' as the best for evaluating state aid effectiveness. These are methods 5, 6 and 8.
- Since these three methods build on each other, and since the international variation in evaluation experience and the quality of existing data sources varies, we recommend the Member States to start with the simplest of these three methods (method 5) and build towards applying methods 6 and 8 at later stages of their development.

10 Recommendations

The most important recommendation of our research is that national governments should use either method 5, 6 or 8 for evaluating the effectiveness of state aid to SMEs. Other methods are less suited, since they are not able to fully demonstrate the effects of state aid. National governments should select the one method that best fits their circumstances. Due to the variation in experience and knowledge of evaluation studies, method 5 is recommended for international comparisons.

In addition to this recommendation, the following suggestions follow from the current study:

– Politicians and their civil servants should start thinking about how their policies will be evaluated at the earliest stages of the policy process, e.g. during the policy design phase.

– Public policy evaluation should ideally pay attention to deadweight, additionality, displacement and causality. In the evaluation studies that we saw, far too often, no attention was paid to one or more of these items.

– Before implementing a policy process, the situation-before-policy-implementation should be documented thoroughly. This so-called zero-measurement makes measuring policy effects easier.

– It would be beneficial to set up a EU-wide network of national (state aid) evaluation experts, where the experts can share experiences in order to improve their profession. This to improve the (cross-border) learning effects and to co-ordinate evaluation goals and methods used.

Bibliography

Abma, T.A. and M. Noordegraaf (2003), 'Public managers amidst ambiguity; Towards a typology of evaluative practices in public management' in: Evaluation, Vol. 9(3), pp. 285-306

Algemene Rekenkamer (2001), Stimulering van midden- en kleinbedrijf en ondernemerschap (stimulating SMEs and entrepreneurship), nr. 28115, The Hague

Arup Economics & Planning (2000), Evaluation of Regional Selective Assistance 1991–1995, Arup Economics & Planning, London

Auderetsch, D.B. (2002), Entrepreneurship: A survey of the literature, Centre for Economic Policy Research, London

Audretsch, David B. (2003), 'Standing on the shoulders of midgets: the US Small Business Research Program (SBIR)', in: Small Business Economics, Vol. 20, pp. 129-135

Babbie, E. (2004), The practice of social research, Belmont, Wadsworth

Bager-Sjögren, L. (2003), On-the-Job training and Small Business Competitiveness – EU-structural fund impact in Sweden, Swedish Institute of Growth Policy Studies, Paper presented at ICSB conference 2003

Bateman, M. (2000), 'Neo-Liberalism, SME development and the role of business support centres in the transition economies of Central and Eastern Europe', in: Small Business Economics, Vol. 14, pp. 275-298

Bekkers, V. and A.B. Ringeling (eds.) (2003), Vragen over beleid; perspectieven over waardering, Lemma, Utrecht

Bemelmans–Videc, M.L., E. Vedung and R.C. Rist (2003), Carrots, sticks and sermons, policy; instruments and their evaluation, Transaction Publishers, London

Benett, J. (2003), Evaluation methods in research, Contunuum, London

Besley, T. and P. Seabright (1999), The effects and policy implications of state aids to industry: An economic analysis; report to DG-III of the European Commission

Bondonia, D. (2002), 'Evaluating decentralised policies; A method to compare the performance of economic development programmes across different regions or states', in: Evaluation Vol. 8(1), pp. 101-124

Bressers, J.Th.A. and A. Hoogerwerf (eds.) (1995), Beleidsevaluatie, Samsom, Alphen aan den Rijn

Brofoss et al. (2004), Evaluation of Norway's participation in the EU's 5th Framework Programme, STEP Senter for Innovasjonsforskning, Oslo, Norway

Brouwer, E.B., P. den Hertog, T. Poot, and J. Segers (2002), WBSO nader beschouwd; onderzoek naar de effeciviteit van de WBSO, PriceWaterhouseCoopers

Brugge, F. van der, and P. van Run (2004), ''Verantwoording van beleid én leerinstrument'; Evaluatieonderzoek in de Europese Commissie', in: Basis, Vol. 2004(2), pp. 28-30

Bulder, B. (2004) 'Effectief evalueren; zeven adviezen voor de strategie van evaluatieonderzoek', in: BASIS, periodiek voor beleids- en consumentenonderzoek 2004, evaluatie-special, pp 30-32.

Clarysse, B., V. Bilsen, G. Steurs and J. Larosse (2004), Measuring Additionality of R&D Subsidies with Surveys: Towards an Evaluation Method for IWT-Flanders, Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen

Commission of the European Communities (1992), Evaluation of economic effects, Luxembourg

Commission of the European Communities (2001), Ninth survey on State Aid in the European Union, Brussels, 18.7.2001

Commission of the European Communities (2003), Creating an entrepreneurial Europe; The activities of the European Union for small and medium-sized enterprises (SMEs), Commission Staff Working Paper, Brussels

Commission of the European Communities (2003a), State Aid Scoreboard, autumn 2003 update, Brussels

Commission of the European Communities (2004), State Aid Scoreboard, spring 2004 update, Brussels

Cooper, R.S. (2003),' Purpose and Performance of the Small Business innovation research (SBIR) Program' in: Small Business Economics, Vol. 20, pp. 137-151

Cowling, M. and P. Mitchell (2003), 'Is the Small Firm Loan Guarantee Scheme Hazardous for Banks or Helpful to Small Business?' in: Small Business Economics, Vol. 21, pp. 63-71

Deloitte & Touche (2003), Strategic evaluation of financial assistance schemes to SMEs, Final report

Demarteau, M. (2002), A 'Theoretical framework and grid for analysis of programme-evaluation practices' in: Evaluation, Vol. 8(4), pp. 454-473

Department of Trade and Industry, Guidance on preparing evaluation plans, DTI, London

Diez, M.A. (2002), 'Evaluating new regional policies. Reviewing the theory and practice', in: Evaluation, Vol. 8, pp. 285-305

Dunn, W.N. (2004), Public policy analysis, an introduction, Prentice Hall, Englewood Cliffs

Ederveen, S., J. Gorter, R. de Mooij and R. Nahuis (2002), Funds and games; the economics of European cohesion policy, CPB Netherlands' Bureau for Economic Policy Analysis, the Hague

EIM (2002), Effective knowledge transfer to SMEs, Zoetermeer, the Netherlands

EIM (2003), Barriers to Entry; Differences in barriers to entry for SMEs and large enterprises, EIM, Zoetermeer, the Netherlands

European Commisison (1995), Identifying the Structural Effects of Community Interventions, MEANS handbook No.2, Luxembourg

European Commisison (1996), Quality assessment of evaluation reports: a framework, MEANS handbook No.5, Luxembourg

European Commission (1997), Competition law in the European Communities; Volume IIB, Explanation of the rules applicable to State aid, Brussels/Luxembourg

European Commission (1997a), Evaluating EU Expenditure programmes: A Guide; Ex-post and intermediate Evaluation

European Commission (1999), Thematic evaluation of structural fund impacts on SMEs

European Commission (1999a), the Collection: Evaluating Socio-Economic Programmes, Luxembourg: Office for Official Publications of the European Communities

European Commission (2000), Competition policy in Europe and the citizen, Luxembourg, Office for Official Publications of the European Communities

European Commission (2002), A study of Business Support Services and Market Failure

European Commission (2002a), European Competitiveness Report 2002, Luxembourg, Office for Official Publications of the European Communities

European Commission (2003), Evaluating EU Activities; A practical guide for the Commission services, European Commission, Brussels

European Commission (2003a), General Invitation to tender No ENTR/03/79; Review of methods to measure the effectiveness of state aid to Small and Medium-Sized Enterprises (SMEs), European Commission, Brussels

European Commission (2003b), Green Paper; Entrepreneurship in Europe, European Commission, Brussels

European Institute for Public Administration (2002), Summary of Proceedings of the seminar 'Towards Efficient and Effective State Aid', Maastricht, the Netherlands

Fraser, S. (2003), The impact of Investors in People on small business growth: who benifits?, in: Environment and Planning C: Government and Policy, Vol. 21, pp. 793-812

Fundación de Economía Analítica (1998), State aid and convergence in the European Union

Gerhardter, G. and M. Gruber (2001), Regionalförderung als Lernprozess; Evaluierung der Förderungen des Bundeskanzleramtes für eigenständige Regionalentwicklung, Schriften zur Regionalpolitik und Raumordnung nr. 32, Bundeskanzlerambt, Vienna

Hart, H. 't, J. van Dijk, M. de Goede, W. Jansen and J. Teunissen (1998), Onderzoeksmethoden, Boom, Meppel

Henry, G.T. (2002), 'Choosing criteria to judge program success; A values inquiry' in: Evaluation, Vol 8(2), pp. 182-204

Hill, R.C., L.C. Adkins and K.A. Bender (2003) Test statistics and critical values, (see http://www.bus.lsu.edu/academics/economics/faculty/chill/personal/heckit)

Hornschild, K. and F. Meyer-Krahmer (1992), Evaluation of Economic Effects: Relevance and impacts of EC-Programmes promoting Industrial R&D with special emphasis on Small and Medium sized Enterprises (Pilot methodological study), Commission of the European Communities, Luxembourg

Hufen, J.A.M. and A.B. Ringeling (1990), Beleidsnetwerken; Overheids-, semi-overheids- en particuliere organisaties in wisselwerking, VUGA, 's Gravenhage

Innovation Development and Planning Ltd. (1998), Evaluation of Policy measures for Creation and Development of Small and medium-sized Enterprises, London

Innovation Development Planning Group (1988), Evaluation of policy measures of the creation and development of small and medium-sized enterprises, Synthesis report, London

Jansen Schoonhoven, P. and M. Vos (1998), "Ex-ante en Ex-post Evaluatie-Onderzoek: verschillen, overeenkomsten en afstemmingsmogelijkheden verkend" in: J.C. Hellendoorn et al. (eds.), Gewikt en Gewogen; Vijventwintig jaar Beleidsanalyse, Sdu Uitgevers, Den Haag, pp. 347-366

Jansen, R. (1984) Evaluatiemethoden ten behoeve van milieubeleid en –beheer, Economisch en Sociaal Instituut der Vrije Universiteit, Amsterdam, the Netherlands

Kangasharju, A., and T. Venetoklis (2002), Business subsidies and employment of firms; Overall evaluation and regional extensions, VATT, Helsinki

Lach, S. and R.M. Sauer (2001), R&D, Subsidies and productivity, Working Paper, Hebrew University of Jerusalem, Department of Economics

Lang, J. (2001), 'Improving structural policy under conditions of hybrid governance; Multi-actor constellations, integrated feedback instruments and the use of evaluation results' in: Evaluation, Vol. 7(1), pp. 7-23

Lenihan, H., M. Hart and S. Roper (2003), Evaluation the net additionality of industrial development assistance in Ireland, Paper presented at the European Regional Science Association Conference in Jyväskylä, Finland

Levin-Rozalis, M. (2000), 'Abduction: A logical criterion for programme and project evaluation' in: Evaluation, Vol. 6(4), pp. 415-432

Licht, G. (2003), The role of additionality in evaluation of public R&D programmes, presentation for the 11th annual TAFTIE Seminar, held in Vienna on 24/25 November 2003

Maddala, G.S. (1983), Limited-dependent and qualitative variables in econometrics, (Econometric Society Monographs), Cambridge: Cambridge University Press

Maes, J., and L. Sels (1999), De evaluatie van opleidingseffecten; Een vergelijking van evaluatiemethoden, research Report 9919, Katholieke Universiteit Leuven

Majone, G. (1996), Regulating Europe, Routledge, London

Martin, S. (2001), Industrial Organisation; A European perspective, Oxford University Press, Oxford

Mayne, J., et al. (1992), Advancing Public Policy Evaluation, North-Holland, Amsterdam

MEET (2000), Monitoring and effect measurement of the technology instruments, Dutch Ministry of Economic Affairs, The Hague

Meeusen, W., and W. Janssens (2000), 'Additionaliteit'– versus 'substitutie'-effecten van overheidssteun aan O&O in bedrijven in Vlaanderen: een econometrische analyse aangevuld met de resultaten van een kwalitatieve bevraging, IWT-Observatorium, Brussels

Ministry of Economic Affairs (2002), Manual public policy evaluation & performance data, The Hague

Muizer, A.P., and Y.M. Prince (2000), Zero-measurement of the Syntens innovation centres - with regard to an evaluation in the future, EIM, Zoetermeer

Muizer, A.P., Y.M. Prince, M.B.M. Lankhuizen (2003), Evaluatie SKO- en SBO-regeling; Eindrapportage, EIM, Zoetermeer

Nagel, S.S. (ed.) (2002), Handbook of public policy evaluation, Sage Publications, Thousand Oaks

Neaman, W.L. (2000), Social research methods; Qualitative and quantitative approaches, Allyn and Bacon, Boston

North, D., D. Smallbone and I. Vickers (2001), 'Public Sector Support for Innovating SMEs', in: Small Business Economics, Vol. 16, pp. 303-317

OECD (1998), Industrial performance and competitiveness in an era of globalisation and technological change, Paris

OECD (1999), Improving evaluation practices, Paris

OECD (2002), Competition and efficiency in publicly funded services, Paris

OECD (2004), Evaluation of SME policies and programmes, Paris

Österreichisches Institut für Raumplanung (ÖIR) (2001), 6 Jahre Program-Evaluation -eine Bilanz. Endbericht, Vienna

PACEC (2001), Evaluation of Smart (including SPUR) 2001, Final report and Appendices, DTI Evaluation Report Series, no. 3, Department of Trade and Industry, London

Pawson, R. and N. Tilley (1997), Realistic evaluation, Sage, London

Peters, L. (1998), Diffusie van belichaamde technologie in Vlaanderen: Een empirisch onderzoek op basis van input/output-gegevens, Vlaams Technologie Observatorium, Brussels

Porter, M.E. (2000, Creating innovative capacity, The 1999 Gilbert memorial lecture, January 2000, in: Management Services, pp.8-12

Prince, Y.M. (1998), Analysing the Impact of Government Programmes, Orientation Report, EIM, Zoetermeer, the Netherlands

Prince, Y.M. (2000), Effect measurement of knowledge transfer instruments, EIM, Zoetermeer

Proceedings of the Seminar: Towards efficient and effective state aid, 20 November 2002

Quinn Patton, M. (2001), Qualitative research & evaluation methods, Sage Publications, Thousand Oaks

Röller, L.H., H.W. Friederiszick and D.J. Neven (2001), Evaluation of the effectiveness of state aid as a policy instrument, Final report, Berlin / Geneva

Roper, S. and N. Hewitt-Dunas (2001), Grant assistance and small firm devenlopemnt in Northern Ireland and the Republic of Ireland

Rossi, P.H., H.E. Freeman and M.W. Lipsey (1999), Evaluation; A systematic approach, Sage Publications, Thousand Oaks

Saaty, T.L. (1980), The Analytic Hierarchy Process, McGraw Hill, New York

Sander, C. (2000), Bridging the Gaps of Market Failure and Asymmetries in Microfinance and Enterprise Financing, Austrian Development Cooperation, Kampala

Sheperd, W.G. (1997), The economics of Industrial Organisation; Analysis, markets, policies, Prentice-Hall, Upper Saddle River, New Jersey

Siegel, D.S., C. Wessner, M. Binks and A. Lockett (2003), 'Policies promoting innovation in small firms: Evidence from the U.S. and U.K.,' in: Small Business Economics, Vol. 20, pp. 121-127

Small Business Service (2002), Cross cutting review of government services for small businesses, London

STEP (2004), Evaluation of Norway's participation in the EU's 5th Framework programme, STEP Senter for innovatsjonsforskning, Oslo, Norway

Storey, D. (2002), 'Six Steps to Heaven: Evaluating the Impact of Public Policies to Support Small Business in Developed Economies', in: International Journal of Entrepreneurship Education, Vol. 1, No. 2, pp. 181-202

Turok, I. (1991) Policy evaluation as science: a critical assessment, in: Applied Economics, Vol. 23, pp. 1543-1550

Uusikylä, P. and P. Virtanen (2000), 'Meta-evaluation as a tool for learning; a case study of the European Structural Fund evaluations in Finland' in: Evaluation Vol. 6(1), pp. 50-65

Vedung, E. (2000), Public Policy and Programme Evaluation, Transaction Publishers, London

Venetoklis, T. (1998), Evaluation and monitoring of business aid in Finland; Applicant enterprises projects and distributors of aid in industrially declining regions, a quantitative approach, VATT, Helsinki

Venetoklis, T. (1999), Process evaluation of business subsidies in Finland; A quantitative approach, Government institute for Economic Research, Helsinki

Venetoklis, T. (2000), Methods applied in evaluating business subsidy programs: a survey, VATT-discussion papers no. 236, Helsinki

Venetoklis, T. (2002), Public policy evaluation: introduction to quantitative methods, Government Institute for Economic research (VATT), Helsinki

Verhoeven, N. (2004), Wat is onderzoek? Praktijkboek methoden en technieken voor het hoger beroepsonderwijs, Boom, Amsterdam

Verschuren, P.J.M. (1998), "Wetenschappelijke beleidsevaluatie door toepassing van structurele modellen" in: J.C. Hellendoorn et al. (eds.), Gewikt en Gewogen; Vijfentwintig jaar Beleidsanalyse, Sdu Uitgevers, Den Haag, pp. 277-298

Vreeker, R., P. Nijkamp and C. ter Welle (2001), A Multicriteria Decision Support Method for evaluation airport expansion plans, Tinbergen Institute Discussion Paper

Wren, C. and D.J. Storey (2002), Evaluating the effect of soft business support upon small firm performance, in: Oxford Economic Papers, Volume 54, pp. 334-365

Ziegelaar, A. (2004), “Licht in de zwarte doos; Naar een evenwichtige methodologie van beleidsevaluatie”, in: BASIS, periodiek voor beleids- en consumentenonderzoek 2004, evaluatie-special, pp. 16-20

Annex I Overview of methods for evaluating alternative environmental policy instruments

Continuous quantitative evaluation methods

- Explicit utility functions
- Expected utility functions
- Restricted utility and satisficing analysis
- Hierarchical decision structures
- Goal programming
- Game theoretical methods
- Ideal point method
- Interactive decision models

Discrete quantitative monetary evaluation methods

- Cost/benefit analysis
- Cost effectiveness analysis
- Planning Balance Sheet method
- Shade projection method

Discrete quantitative non-monetary evaluation methods

- Goal Achievement matrix
- Expected value method
- Discrepancy analysis
- Concordance analysis
- Graphic representation

Discrete quantitative numeric evaluation methods

- Ordinal concordance analysis
- Frequency method
- Lexicographic ordering
- Permutation method
- Eigenvalue method
- Regime method
- Multi-dimensional method
- Metagame method
- Mixed data method (evamix)
- Trichotomic choice method

Discrete qualitative non-numeric evaluation methods

- Score card method
- Key factor method

Source: Jansen ,1984

Annex II Overview of methods for evaluating state aid effectiveness

Table 8 Overview of methods for evaluating state aid effectiveness

Stage in the policy process	*Research methods / data collection*	*Ex-ante*	*Ex-nunc*	*Ex-post*
Fixing the agenda	– Monitor study	X	X	X
	– Scenario studies	X		
Problem diagnosis	– Case studies	X		
	– International comparison	X		
	– Regional comparison	X		
Policy design	– Field or policy experiment	X		
	– Ex-ante cost-benefit analysis	X		
	– Policy theory analysis	X	X	X
	– Pilot	X		
	– Simulation and gaming	X		
	– International comparison		X	X
	– Regional comparison	X		
Implementation	– Monitor study	X	X	X
	– Primary behaviour measurement at the target group		X	
	– Case studies		X	
	– Process evaluation		X	X
Execution of the scheme	– Behaviour measurement at target group		X	X
	– Monitor study	X	X	X
	– Case studies		X	X
	– Customer satisfaction study		X	X
	– Mystery clients study		X	
	– Document analysis		X	X
	– Cost-benefit analysis	X	X	X
	– Cost effectiveness study		X	X
	– SWOT analysis		X	X

Source: EIM, based on the interview with Professor Peter van Hoesel

Monitor studies are studies that follow a group of subjects for a longer period, analysing how certain characteristics (such as behaviour, etc.) change through time.
Scenario studies are methods for forecasting what would happen if a policy were implemented. Scenario studies often develop several possible scenarios, based on different assumptions.

Case studies are in-depth analyses of a limited number of subjects.
Regional and international comparisons use other regions or countries as a benchmark for judging results of the investigated region or country.
Field or policy experiments analyse the differences in behaviour between a targeted group (e.g. aid recipients) and a control group (e.g. non-recipients with same characteristics).
Cost-benefit analyses judge the ratio between the costs and the achievements of a policy measure.
Policy theory analysis judges the (academic) validity of the assumptions and relations that policy makers have made, prior to the implementation of a policy measure.
A pilot is a small-scale trial of a policy measure, to see what effect it has on the selected test subjects.
Simulation and gaming involve techniques for modelling or imitating the (expected) behaviour of the target group and simulating how the target group would react to (alternative) policy measures.
Behaviour measurement involves analysis of the behaviour of the target group after the implementation of the policy measure. Primary behaviour measurement takes place in an early stage of the implementation of the policy measure.
Process evaluation is concerned with the policy process. It does not analyse whether the *state aid measure* was effective, but it shows the effectiveness of the *policy process*
A customer satisfaction study is an analysis of the satisfaction of assisted firms. This type of study is subjective by nature. Assisted firms judge the utility of the aid against their own (subjective) expectations and standards for the added value of the aid.
Mystery clients are 'under cover' evaluators that pretend to be clients, and judge the deliverer's performance from a clients' perspective.
Document analysis involves desk research of various types of secondary data (documents), such as application forms and progress reports of the target groups.
Cost effectiveness study is an extended version of a cost-benefit analysis, where the benefits are extended to the total of effects (including positive and negative side effects).
SWOT analysis is an analysis of the strengths, weaknesses, opportunities and threats, where the total advantages and disadvantages of a policy measure are weighted (e.g. against each other).

Annex III Description of sample selection bias and Heckman's two-step model

Sample selection

Storey (2002) shows that public policy evaluations invariably have to overcome the *sample selection bias* problem. Sample selection occurs when assisted firms are not representative for the total population of firms. The assisted firms are 'selected' from the total population on the basis of self-selection or agency selection. As a result, the firms that receive support differ systematically from the unassisted firms. Selection depends on some (latent) selection variable (i.e. only the firms with certain common characteristics will participate). If the evaluation criterion is somehow associated with this selection variable, than the outcomes of the evaluation analyses may be biased. This bias, which is due to the sample selection phenomenon, is called the *sample selection bias*.

An example

Let us illustrate the sample selection bias with the use of a simplified example. In our example, firms can apply for state aid in the form of free consultancy services. The aim of the scheme is to boost the national economic development through accelerating the growth rate of private firms. The scheme's quantified target (i.e. the evaluation criterion) is the annual growth rate of the assisted firms. If the annual growth rate is at least 5 percentage points higher than non-assisted firms, the state aid measure is successful.

After one year, our state aid measure is being evaluated. For this evaluation, the evaluator wants to compare the growth rate of the assisted firms with the growth rate of a control group of non-assisted firms. The evaluator has compiled the following data set:

Firm	*Does the firm have a growth objective?*	*Has the firm been assisted?*	*Annual firm growth*
1	yes	yes	14%
2	yes	yes	15%
3	yes	yes	18%
4	yes	yes	13%
5	yes	yes	16%
6	yes	yes	12%
7	yes	yes	17%
8	yes	no	12%
9	yes	no	15%
10	yes	no	13%
11	yes	no	14%
12	yes	no	13%
13	yes	no	12%
14	yes	no	12%
15	no	no	5%
16	no	no	4%
17	no	no	2%
18	no	no	4%
19	no	no	8%
20	no	no	3%

The average growth rate of the assisted firms (firms 1-7) is 15%, whereas the average growth rate of the non-assisted firms (firms 8-20) is 9%. The evaluator concludes that the state aid scheme is successful.

The selection bias problem

At first glance, this seems a sound approach. The evaluation is based on a simple but effective statistical technique, it is robust, cost-effective and transparent. However, there may be a problem with sample selection. If there is sample selection, the evaluation results will be heavily biased by this selection. For example, only firms with growth-objectives may want to participate. Obviously, the growth rate depends on whether or not firms have a growth objective.

If we control for this sample selection bias, comparing the growth rates of the assisted firms with the growth rates of only the non-assisted firms that have a growth objective, than the evaluation results are indeed very different. The growth rate of the assisted firms *with* a growth objective (firms 1-7) is still 15%, but the growth rate of the unassisted firms *with* a growth objective (firms 8-14) is 13%. Taking this sample selection into account, the evaluator will arrive at a much different conclusion. The difference between the assisted firms and the control group is so small that we can no longer conclude that the state aid scheme is a success!

The Heckit model

In our simplified example, we have used an *observed* selection variable (growth objective) and a very simple evaluation criterion (annual growth rate). However, in practice sample selection often depends on a number of *unobservable* (latent) selection variables which makes it much more difficult to control for sample selection. Also, we have used a discrete selection variable (yes/no) while in practice selection variables may be of another type (e.g. scale variables). Third, evaluations often use regression

models instead of simple bi-variate comparison of averages. To control for sample selection in linear regression models, Heckman developed a two-step estimation procedure. This procedure is sometimes called the Heckit model. The value of the Heckit procedure is that it can control for sample selection when the effectiveness of a policy measure is being assessed.
The Heckit two-step estimation procedure is easy to implement. It comprises formulating a probit estimation of the selection equation followed by least squares estimation of an augmented regression. The selection equation is used to estimate the value of the unobservable selection parameters, based on the question whether or not the firm was (self-)selected for assistance. The second equation explains the performance change (e.g. firm growth, level of education, survival rate, etc.), taking account of the factors of the selection equation.

Sources: Storey (2002), Hill et al. (2003), Maddala (1983)

Annex IV Discussion of the selected evaluation studies

In this annex we discuss three good practices of evaluation studies from the recent past (all three were published in or after 2002). We choose studies representing each of the recommended evaluation methods (methods 5, 6 and 8).

Study 1: evaluation of Subsidy Knowledge transfer Enterprises (following method 5)

In 2001, the Dutch Subsidy Knowledge transfer Enterprises (SKO) was implemented in order to stimulate technology diffusion among SMEs. The SKO intends to stimulate the application of existing technologies in enterprises that are unfamiliar with the technology. With the implementation of the support policy, a quantitative objective was not formulated. The SKO is targeted at technology-following SMEs (i.e. SMEs that do not actively develop technologies themselves). Between May 2001 and December 2002, some € 8 million of subsidies was paid out.

In 2003, the (intermediary) effects of the SKO were evaluated. The evaluators formulated a series of research questions, among which:

- How effective is the SKO?
- What is the effect of the SKO on innovation and business performance (second order effects)
- How large is the additionality of the SKO?

Although the SKO-policy did not have quantified objectives, the researchers have tried to quantify the effects anyhow. This was done by conducting a telephone survey among 244 beneficiaries, 49 rejected applicants and a control group of 101 non-beneficiaries. The data from the survey were analysed with very simple statistical tools: frequencies and cross-tabs. These analyses resulted in such outcomes as: with 21% of the beneficiaries, implementation of the new technology resulted in additional turnover, compared to 37% of rejected applicants.

For easy reading, the results were displayed in tables or in graphs. In addition to the survey, in-depth interviews were performed with key persons, among which staff members of employers' organisations and employees of the agency that executed the SKO-scheme.

On the basis of the survey results, the evaluators were able to quantify the first and second order effects of the SKO measure. It appeared that there were displacement effects (i.e. 65% of the firms whose application was rejected performed the activities anyhow). Also, the additionality of the measure was established.

Source: Muizer et al., 2003

Study 2: evaluation of the WBSO (following method 6)

The Dutch Law Stimulating Research & Development (WBSO), enacted in 1994, intends to reduce the wage-costs of R&D-workers through exemptions for tax and social contributions. The underlying policy objective of the measure is to increase the level of private R&D. Between 1994 and 2001, as many as 73,145 requests were made by 24,754 firms. In 2001 alone, the WBSO budget amounted € 337 million.

In 2002, the WBSO was evaluated. The central question of the evaluation was if, and to what extent, the WBSO leads to increased private R&D activities (first order effects). Also, the evaluation estimates the effects of the WBSO on the innovativeness (second order effect) and the business performance (third order effects) of the beneficiaries. The evaluation includes a study of the *reach* of the WBSO with respect to the target group, an overview of the beneficiaries' perception of the execution of the scheme, and some recommendations for possible improvements. Besides, the evaluation investigates whether the economic circumstances that gave rise to the aid scheme, had changed since its implementation.

Since the evaluation objective includes both quantitative aspects (effects, reach) and qualitative aspects (perception, improvement), the evaluators applied a combined evaluation approach. They collected existing statistical data and collected new statistics (using a telephone questionnaire survey), but they also conducted semi-structured interviews.

For estimating the effect of the WBSO on the R&D activities (first order effect), the evaluators specified an OLS-model and a tobit-model where the R&D labour costs are explained based on the R&D labour costs in the previous year, the amount of WBSO benefits and other exogenous variables. The necessary data were collected during a survey. The survey was a telephone survey among a sample of beneficiaries and a control group of mirror firms, which were selected on the basis of some characteristics of the beneficiaries.

The effects of the WBSO on the innovativeness (second order effects) were estimated by the same econometric models, but now with the turnover of new products as a percentage of the total turnover, as dependent variable.

The evaluators tried to explain the effects of the WBSO on the business performance (third order effects), using the same mode again, but now with the 'gross value added' as dependent variable. However, on the basis of the available data they could not reach a high enough explaining factor (the outcomes of the tests were not significant), due to the many explaining variables, of which only a few were available in the database.

The quantitative data were assessed with econometric analyses. In their analyses, the evaluators included the concepts of substitution and additionality.

Source: Brouwer et al., 2002

Study 3: evaluation of the Consultancy Initiatives Scheme (following method 8)

So far the authors know, a good example of method 8 is not (yet) available. A few studies include some different aspects of method 8 (for example Barger-Sjögren (2003), Gerhardter and Gruber (2001), Röller and Friedriszick (2001), European Commission (1999), Vebnetoklis (1999), Arup Economics & Planning (2000), PACEC (2001)), however, none fully cover all aspects of method 8. Some use existing statistics only, others cover a single aid scheme, etc. Below, we therefore describe a study that comes quite close to applying method 8. We selected an evaluation study that does apply the right techniques and uses all the required data from various sources that are contained in method 8, but it does not investigate a full range of measures forming a programme. Rather it applies method 8 to a single state aid measure.

The 1988 Enterprise Initiative is a set of policy measures that intends to raise the United Kingdom's economic performance. The Consultancy Initiatives Scheme-programme was a key component of the Enterprise Initiative. The objective of the programme was to 'improve the competitiveness of small- and medium sized enterprises by improving the quality of management through subsidised consultancy in key strategic functions'. The programme consisted of various types of consultancy support. Through its objective, the programme targeted the market failure of asymmetric information. The Consultancy Initiatives Scheme-programme was available from January 1988 through September 1994. Participating companies received a grant for between 5 and 15 days of the consultant's time. The grant was in the form of a subsidy for 50% or two thirds of the approved daily rates. During the scheme's period 114,400 projects had been approved, involving £ 275 million of public subsidies.

The evaluators intended to examine the direct effects of the support on sales turnover growth, employment growth, and firm survival, using an econometric approach. The empirical research was based on a sample of 2,840 assisted firms and a control group of 1,486 unassisted firms. For collecting the necessary data, the evaluators studied the record cards of all sample firms, they conducted a written questionnaire survey among all the 4,326 firms in the sample, and they held a telephone survey among non-respondents. Of the 4,326 sample firms, 64.7% had survived until 1996 and from these survivors, 1,136 usable responses were obtained.

However, from previous research it became eminent that most projects were undertaken for reasons of expansion, such as developing new products or exporting as a response to a rapidly changing market. This fact heavily influences the outcome of any growth-analysis, since participating firms are likely to grow a lot faster than any control group, but this growth is not due to the support provided. Besides, the government had set explicit criteria for participation in the scheme. Thus, self-selection and agency-selection play a major role in the evaluation of the scheme.

The evaluators include the selection problem in their econometric analysis, by specifying different models for the assisted firms and a control group[33]. They analyse whether sample selection occurred using a bivariate probit method with sequential selection. They include the identified selection effects in the regressions.

For evaluating the effects of the Consultancy Initiatives Scheme on firm survival and firm growth (which constitute the second order effect), the evaluators applied parametric analysis with regression

33 The control group consisted of firms that applied for support, but did not pass the final selection stage (i.e. they did not agree to the consultant's terms of reference). Such a control group can be assumed to have characteristics that are similar to the characteristics of the supported firms.

analysis and chi-square tests, and Maximum Likelihood probit equations with regression analysis and chi-square respectively.

The evaluation concluded that the consultancy advice provided toward marketing does not have an impact on the smallest firms. According to the evaluators, the scheme is most effective with SMEs that are neither too small nor too large in size, raising their survival rate by 4% and the growth rate of surviving firms by up to 10% per year. They find that the scheme has a 'substantial' impact, but that the effects should be viewed in the light of possible displacement effects.

Source: Wren and Storey (2002)