



**DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY
COMMITTEE ON INDUSTRY AND BUSINESS ENVIRONMENT**

Working Party on Small and Medium-Sized Enterprises and Entrepreneurship

EVALUATION OF SME POLICIES AND PROGRAMMES

THE OECD BOLOGNA PROCESS

2nd OECD Ministerial Conference on SMEs on "Promoting Entrepreneurship and Innovative SMEs in a Global Economy -- Towards a more Responsible and Inclusive Globalisation" jointly organised by the OECD and the Turkish Ministry of Industry and Trade, ISTANBUL, Turkey, 3-5 June 2004

JT00164693

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

FOREWORD

At the first OECD Conference of Ministers responsible for SMEs, hosted by the Italian government in Bologna, Italy, in June 2000, Ministers from nearly 50 member and non-member economies adopted the “Bologna Charter for SME Policies”. They envisaged the Bologna Conference as the start of a policy dialogue among OECD Member countries and non-Member economies and that it would be followed up by a continuous monitoring of progress with the implementation of the Bologna Charter. This dialogue and monitoring have become known as the “OECD Bologna Process”. The second OECD Conference of Ministers Responsible for SMEs, hosted by the Turkish Ministry for Industry and Trade, envisaged by Ministers at Bologna, provides an occasion to assess the impact on SMEs of new developments relating to globalisation.

This report is one of ten background reports prepared for the Istanbul Ministerial Conference, the theme of each of the ten reports being linked to one or more Workshops of the Ministerial Conference. Several earlier versions of the report were reviewed by the Working Party on SMEs and Entrepreneurship whose comments have been incorporated into the final version. Non member economies participating in the OECD Bologna Process have also had an opportunity to provide comments. This final report also sets out some policy messages and recommendations that have emerged from the preparatory work undertaken in the OECD Working Party for SMEs and Entrepreneurship. The wide variation in stages of economic development, institutional arrangements and political context across the economies participating in the Bologna Process, now more than 80, means that not all parts of specific policy recommendations are appropriate for all participants. The messages and recommendations outlined below provide material from which government may choose to draw in promoting innovative SMEs in the global economy. In broad terms, these policy messages and recommendations elaborate on the themes developed in the Bologna Charter. Ministers will consider these and other recommendations in their deliberations at the Istanbul Conference.

This report was prepared by Professor David Storey, Director, Small Business Centre, University of Warwick, United Kingdom, in close cooperation with the SME Unit of the OECD Secretariat.

This report is issued on the responsibility of the Secretary-General of the OECD. Views expressed are those of the authors and do not necessarily reflect those of the Organisation or its member governments.

This and other background documents prepared for the 2nd OECD Conference of Ministers Responsible for SMEs may be downloaded free of charge from the following websites:

SME Unit website: <http://www.oecd.org/sti/smes>

Conference website: <http://www.oecd-istanbul.sme2004.org>

Copyright OECD, 2004.

Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

TABLE OF CONTENTS

FOREWORD.....	2
EXECUTIVE SUMMARY	4
INTRODUCTION	7
BACKGROUND	8
EVALUATION OF SME POLICIES.....	10
The context for evaluation.....	10
The evaluation framework	15
The Six Steps Approach.....	16
Some modifications and limitations to the Six Steps approach.	23
Examples of Good Practice Evaluation.....	25
Lessons.....	30
(i) The link between the technical quality of evaluation techniques and the estimated programme .	32
(ii) The frequency and scale of the evaluation, related to the scale of the programme	32
(iii) Evaluation by internal or by external organisations?.....	32
(iv) Implementing the findings of the evaluation	33
(v) Significance for middle and lower income countries.....	33
CONCLUSIONS	35
ANNEX 1	37
ANNEX 2	38

Evaluation of SME Policies and Programmes

EXECUTIVE SUMMARY

This paper has provided an extensive review of the methods used to evaluate elements of SME policy. It draws upon experience primarily from OECD member countries but, where appropriate, has also made reference to non-member economies. The key conclusions are presented and summarised in the Box below.

First, that evaluation cannot take place adequately until the objectives and targets of SME policy are clearly specified. The paper suggests using the COTE framework for policy formulation. In this, the letter “C” stands for both policy clarity and policy coherence. ‘Clarity’, in this context, means that the policy is fully understood by those delivering it and those expected to benefit from it. ‘Coherence’, however, is a slightly more sophisticated concept. It refers to the need to ensure that all parts of government seek to collectively work in the interests of SMEs. Specifically SME policies are unlikely to be ‘coherent’ where they are delivered by many different government departments without co-ordination or where some departments focus upon regulation of SMEs and others focus upon ‘support’, without an adequate dialogue between them.

The letter “O” stands for ‘Objectives’ of policy. To ensure clarity the objectives of policy must be clearly stated; in practice such objectives might be to increase the number of firms, or to increase the formation rate of firms, or to increase the value added of firms. It might also include seeking to raise business ownership amongst groups defined in terms of ethnicity, gender, disadvantage or location.

The letter “T” stands for ‘Targets’ which need to be specified in a manner that enables an assessment to take place of whether or not they are achieved. An example of converting an ‘objective’ into a ‘target’ would be the objective of ‘to increase the formation rate of firms’. Here the target might be ‘to raise new firm formation by 10% over the next three years’. Only then is it possible to be clear whether or not the target is met.

The letter “E” stands for Evaluation, which is the prime focus of this paper. However the first key conclusion of the paper is that evaluation cannot be undertaken unless targets exist. This is because evaluation can only take place in a framework in which the expected policy impacts are clearly specified.

A second key conclusion of the paper is that Evaluation needs to become more central to the policy-making process. It should not be undertaken solely as a historic accounting exercise to determine whether public money has been spent wisely, although that role is of value. Instead of being “at the end of the line” evaluation should be used to inform current policy, so that the objectives and targets may be modified in the light of evidence of policy effectiveness. Hence considerations of how policy is evaluated should therefore be incorporated into policy formulation when new ideas are being developed. They could even influence the choices made by governments about how best to engage with SMEs. To achieve this requires a balance to be struck between the need for simplicity and the need for accuracy.

Ensuring that evaluation is given a higher priority means that all, rather than merely some, programmes are assessed. The paper however recognises there are conflicting arguments about the extent

to which all programmes should be assessed with the same level of sophistication. The key argument in favour of a consistent level of sophistication is that this enables valid comparisons to be undertaken between one programme and another. The argument against is that it is inappropriate to have similar evaluation budgets for programmes of widely differing scales, especially when this imposes information collection costs on SMEs themselves.

Nevertheless the paper makes it clear that it is now technically possible to provide accurate measures of programme impact. It shows, through its “Six Steps to Heaven” approach that, although there are examples of the most sophisticated approach being used, these remain the exception rather than the rule. The cause for concern is that the less sophisticated approaches tend to provide misleading answers, most frequently over-estimating policy impact. The third key conclusion of the paper is that, where appropriate, the most sophisticated – Step 5 or Step 6 – approaches are to be used. The key limitation of the “Six Steps” approach – that it is appropriately applied to assessing the impact of individual programmes and policies rather than that of SME policy in total – is also recognised.

Key policy recommendations

- **Apply the "C.O.T.E." framework to SME policy:**
 - **Clarity and Coherence** of SME policies require a clear rationale for policy intervention and statement of purpose. The various parts of government interacting with SMEs to facilitate their development should ensure that their efforts are consistent and co-ordinated.
 - **Objectives** of SME policies should be clearly specified. Examples would include the creation of new firms, the growth of existing firms or promoting enterprise among target groups in the population.
 - **Targets** should be specified in measurable ways to facilitate evaluation of the extent to which objectives should have been achieved.
 - **Evaluation** of policy, which must be based on clear policy targets, should be the most important test of its effectiveness.
- **Develop an “evaluation culture: by making evaluation of programmes central to the policy process.** Evaluation should not be simply undertaken as a historical accounting exercise but should be used to inform current policy, the objectives and targets of which may be modified in the light of the results of evaluation.
- **Integrate the methodology of, and budget for, evaluation of programmes as part of the legislative process.** These should not be delayed until after the programme has been implemented.
- **Ensure that all, rather than merely some selected, programmes are subject to evaluation.**
- **Major evaluations should be undertaken with independent but informed assessors.** Nevertheless, the involvement of the programme teams, both policy makers and deliverers, is vital.
- **Evaluate all programmes using the most sophisticated techniques that are feasible taking into account the need to be cost-effective.** In practice, more approximate approaches will be necessary in assessing smaller programmes.

In undertaking these evaluations it is necessary to bear in mind some potential conflicts. The first is that ideally those undertaking the evaluations need to be independent of those responsible for the programmes, as in any audit role. But, 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. This may be more difficult to achieve if the evaluator is viewed as an “outsider”. On balance, the paper concludes that the independence of the evaluator is of the greater importance, but this is a close call.

Finally the paper concludes that OECD itself could play a greater role in disseminating “good practice” in SME policy evaluation techniques.

Evaluation of SME Policies and Programmes

INTRODUCTION

The “cross-cutting” theme of this paper is the evaluation of SME policies and programmes. Ministers are asked to consider the form and nature of SME and Entrepreneurship policy evaluation deemed to be best practice. It is also linked to a second theme – that of ensuring the coherence of SME policy. Coherence is achieved when the total public policy package, when viewed holistically, achieves the overall objectives of policy. The elements of the package need to be mutually reinforcing rather than conflicting. A key test is whether it is possible to reallocate funds in a manner which will yield greater/additional benefits to the economy. If reallocation is not possible then policy may be deemed to be coherent. [OECD 2003]¹

In practice the commitment of policy makers and professionals to policy evaluation varies both between countries and even between programmes within the same country. This is reflected in the variation in sophistication of the evaluation techniques employed. This makes it difficult, if different techniques are used, to compare the impact of similar policies within countries. A key role that OECD can therefore play is to seek to harmonise evaluation procedures, thus making it easier to compare policy effectiveness. Nevertheless there are some generic issues common to all countries. These are the need for clarity in the objectives of programmes, the need to set-aside budgets to conduct evaluations, the need for these evaluations to enter the analytical domain and the need for the evaluations to be of a high technical quality.

The key conclusions of the paper are:

- That policies and programmes should have objectives and targets that are, in principle, capable of measurement
- That the methodology of, and budget for, the evaluation should be identified as part of the legislative process and not be delayed until the programme has been implemented.
- That all, rather than merely some, programmes are evaluated.
- That large programmes should use the most sophisticated evaluation methodology available, with this sophistication being ‘scaled down’ appropriately for smaller programmes.

The paper also raises the following issues:

- Who is to undertake the evaluations? Should evaluation be undertaken by programme deliverers, by independent public officials, or only by outside organisations?
- Whether OECD could act as a bank of evaluation expertise, to be drawn upon appropriately both by member and non-member states.

¹ In their review of Policy Coherence in the context of Global Development, OECD (2003) say “there is little point in providing development aid to improve a country’s ability to engage in trade if the donor countries then maintain trade barriers to keep the developing country’s goods out”.

BACKGROUND

OECD has already undertaken considerable work in policy evaluation. For example, OECD (1997) provided five recommendations regarding evaluation:

- The evaluation methodology should be designed at the same time as the policy programme so that the necessary data and acceptance by the relevant parties of the evaluation procedures and criteria can be secured.
- There is a need to ensure the take-up of the evaluation results at the highest possible level of policy making, encouraging public discussion of evaluation results and ensuring a response to the evaluation results from policy makers.
- Evaluation should be user-oriented, serving the information needs of the different programme participants and clients.
- Evaluations should use a combination of methods to satisfy different information needs.
- Evaluations are still frequently inconclusive due to inadequate estimation of the programme impacts.

The above recommendations were initially formulated to address issues of policy evaluation in innovation and technology. The SME Working Party, [OECD (2000)] modified this framework to address the specific issues of SMEs and Entrepreneurship. They identified seven headings under which policies in this area can be assessed. These are:

- Rationale;
- Additionality;
- Appropriateness;
- Superiority;
- Systemic Efficiency;
- Own Efficiency;
- Adaptive Efficiency.

Each is defined below:

- *Rationale*: The justification for the policy.

- *Additionality*: The net added value of the programme.
- *Appropriateness*: The extent to which the programme addresses a clearly identified market failure.
- *Superiority*: Whether the programme is more effective than other possible programmes addressing the same goals.
- *Systemic Efficiency*: The extent to which the programme interacts positively and negatively with other government actions.
- *Own Efficiency*: Whether the programme is cost effective in achieving its objectives.
- *Adaptive Efficiency*: The extent to which evaluations lead to the implementation of programme change.

The above taxonomy emphasises that evaluation of programmes is not to be viewed as “the end of the line”. Instead, evaluations lead to adjustments and amendments to programmes, which are then amended further in the light of experience. These amendments may be in the form of different targets and objectives, a focus on different client groups, different methods of achieving objectives and different resourcing levels. The central challenge has to be to encourage policy makers and professionals away from viewing evaluations as a technique for allocating blame and closing down ineffective programmes – although this may be necessary in some instances. Instead the prime purpose of evaluation has to be seen to be facilitating better service delivery.

A prerequisite of successful evaluations is the need to specify clear objectives and targets for policies. Because they are often expressed in “visionary” terms in order to generate political support, the objectives can be opaque, and so require interpretation on the part of the evaluator.²

For evaluation to play a key role requires the commitment of professionals and policy makers. Prior OECD work has emphasised the importance of evaluation in identifying the principles that need to guide policy. It also stressed that SME policies should be regularly evaluated for effectiveness and efficiency.

² In Lisbon the EC defined its objectives as the Union aims by 2010 “to become the most competitive and dynamic based economy in the world, capable of sustaining growth with more and better jobs and greater social cohesion” EU Green Paper “*Entrepreneurship in Europe*” 2003.

The UK had earlier announced an equally grand objective “to make the UK the best place in the world to start and grow a business” UK Small Business Service, 2000.

EVALUATION OF SME POLICIES

The report will be in five sections:

- The context for evaluation.
- The evaluation framework.
- The Six Steps framework.
- Some limitations of the Six Steps approach.
- Good practice evaluation.
- The lessons.

The context for evaluation

In their review of policy evaluation in innovation and technology, Papaconstantinou and Polt (1997) provide a very helpful definition of evaluation. They say:

"Evaluation refers to a process that seeks to determine as systematically and objectively as possible the relevance, efficiency and effect of an activity in terms of its objectives, including the analysis of the implementation and administrative management of such activities".

Several words or phrases in this definition merit strong emphasis. The first key-word is “*process*”. A central theme of this report is that evaluation is not a “once-off” activity, undertaken once a particular programme has been completed. Instead it is an integral element of a *process* of improved policy or service delivery.

To illustrate this, Figure I shows a highly simplified way in which policy is delivered. It assumes a linear process, beginning with a “Vision Statement”, resulting in the formulation of legislation. The second and third stages, shown as Targets and Measures, occur when legislation is “interpreted” by public servants, to make it applicable to the practical circumstances they face. Examples of such Objectives and Targets are those of the UK Small Business Service in 2000/1. Box 1 shows the Objectives of the Service and Box 2 shows the Targets.

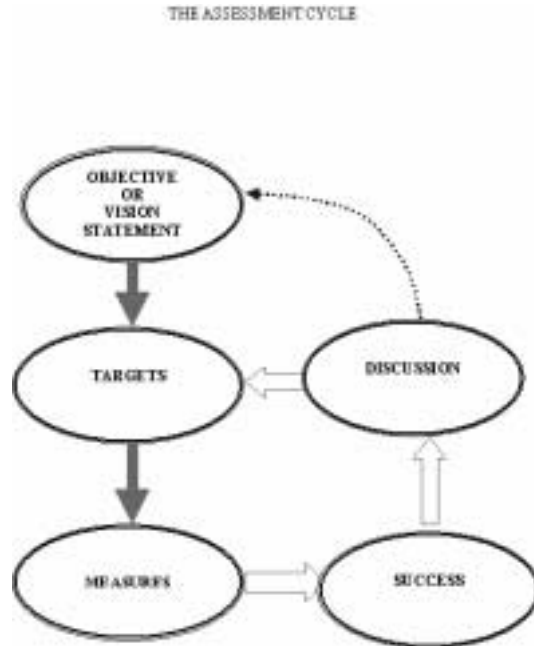
Box 1. The UK Small Business Service Strategic Objectives 2000/1

- Be a strong voice for small business at the heart of government – ensuring that government is aware of the needs of business.
- Strive for a regulatory framework which minimises the burdens on business.
- Develop and maintain a world class business support service to enhance the competitiveness and profitability of small businesses.
- Champion the importance of entrepreneurship across society, particularly in under-represented and disadvantaged groups.

Figure 1: The Assessment Cycle: Evaluation at the End of the Line



Figure 2: The Assessment Cycle: Evaluation as Part of The Policy Process



Box 2. UK Small Business Service Outcomes and Targets 2000/1

- Deliver support to 5000 firms by March 2002.
- Customer satisfaction at least 80% for all services.
- Unprompted and prompted awareness of Business Link branded services amongst small businesses to rise to 20% and 80% respectively by March 2002.
- 5 000 enquiries per day to the Business Link Web site.
- 2 000 calls per day to the national call centre.
- All advisors delivering Business Link services to meet new standards by April 2002.
- Establish Business Link University by September 2001.
- Develop new funding streams to help plug the financial gap for start up and small growing firms.
- Launch initiative to help small businesses understand and prepare for external investment by October 2001.
- Establish and make first investments in new Regional Venture Capital Funds by March 2002.
- Launch new Business Incubation Fund by September 2001 and announce first loans by December 2001.
- Achieve target of one million UK businesses trading on-line by March 2002.
- 1000 TCS Programmes by end of 2001.
- 15 000 STEP Placements by September 2001.
- Action Plan to implement the recommendations from the Council for Excellence in Management and Leadership by October 2001.

Returning now to Figure 1, this assumes that, once the policy has been in operation for a period of time, it will be subject to a “review” or an “evaluation”. This may be considered as the fourth stage and is often conducted when there is either a proposal to change the legislation, either to increase its scale, to reduce it or to abolish it. Or, as the APEC (2002) paper suggests, even if evaluation does not influence current projects, it provides learning for future projects.

The observation that the evaluation is the fourth, and in Figure 1, the last stage in the linear process, leads to the expression that evaluation is at “the end of the line”. However, it is then difficult to square this with the definition of evaluation above that refers to a “*process*”. Where evaluation takes place to simply provide guidance on the scale of budget allocated to a programme it is difficult to view this a “*process*”.

For this process to work effectively two changes are needed. However both involve a fifth stage. This is shown as an additional box called Discussion in Figure 2. The first type of Discussion recognises that when an evaluation is conducted, the most likely outcome is that the programme will have performed better in some respects than others. It is also the case that programmes frequently perform well in aspects that either were not anticipated by those developing them, or who viewed these aspects as of low importance when the programme was being framed. [See Box 3] The conduct of the evaluation provided the opportunity for the objectives of the programme to be discussed. Those responsible for the programme had to decide whether the programme should focus more heavily upon those areas where it was clearly being effective, and place less emphasis upon those where impact is less. Alternatively they could assert the continued importance of the original objectives and re-fashion the programme accordingly.

Box 3. Evaluation of the Shell Technology Enterprise Programme (STEP)

STEP is an eight-week programme in which second year undergraduate students during their summer holidays undertake a supervised and selected project in an SME. Prizes are awarded for excellent projects.

When first formulated, the objectives of STEP were framed in terms of changing the attitudes of graduates to working in small firms and starting their own firm. It also expected that, having employed a graduate for the first time, the small firm would be more likely to employ a graduate in future.

The evaluation found none of these initial objectives to be clearly achieved. Indeed participation in STEP made the student more likely to leave college and work in a large firm. However STEP was extremely successful in enhancing the likelihood of graduates going quickly into jobs, particularly those with poorer academic qualifications and from areas of high unemployment.

In the light of these findings STEP has expanded in scale, but less emphasis is now placed upon its "attitude changing" role.

Source: Fraser, Storey and Westhead (2003).

The example above shows that implementing evaluation as a *process* can be achieved, by feeding the results of evaluation back into the debate, once the evaluation is complete. A second, and perhaps even more powerful contribution to the *process*, is to incorporate the expertise of those who are to conduct the evaluation into the Objectives and Vision statement when legislation is being framed. This is shown as the dotted line in Figure 2. The role of the evaluator would be to confirm that the legislation was framed in such a way as to enable a satisfactory evaluation to be conducted. The presence of the evaluator would ensure the objectives and targets of the legislation were sufficiently clear to enable the evaluator to undertake their work, at a future time. The evaluator would also ensure that a process of data collection is established to enable the evaluation to be conducted. Finally, a budget for the evaluation, reflecting the scale and complexity of the programme, has to be set aside, together with a framework statement on how the evaluation is to be conducted.

Incorporating the evaluator into the earliest stages of legislative change in this way addresses three key problems. The first, as will be discussed in more detail later on, is that the objectives of programmes are frequently expressed in terms that are not sufficiently specific to enable evaluators to assess the effects of the programme. Incorporating the evaluator at the stage when legislation is being formulated requires legislators to address the two important questions:

- “How will we know if this legislation/ programme is successful?”
- “How will we know if this legislation/ programme has failed?”

The second problem faced by evaluators is the absence of suitable data. Sometimes there is not even data on the firms that benefit from the programme, but this is comparatively rare. Much more frequently the evaluator is hampered by the absence of data on firms that did not participate in the programme and so could be used as a “control group” to estimate policy impact. Finally, as will be shown later, the evaluation would also benefit from data on firms that applied to participate in the programme but were rejected, or those that were unaware of the programme but were eligible to apply, or those that were aware but chose not to apply. By including the evaluator at “the front end”, data collection at, or even before, the programme is implemented can begin, so enabling better quality evaluation to be undertaken.

The third problem is that, because consideration is not given to evaluation until, or sometimes long after, the programme begins, those managing the programme feel it is being imposed upon them. A debate then begins about the nature, scale and budget for the evaluation. If those responsible for delivering such programmes find an evaluation unexpectedly imposed upon them, it is understandable that their

response can be one of irritation. The evaluation is inevitably time-consuming, particularly for “front line” staff, for whom the opportunity cost of providing services to firms, may be thought to be very high. Where the commitment to the evaluation is funded from the programmes own budget there is likely to be an even more strident response. Addressing all these issues when the legislation is being framed, so that all parties are aware that the evaluation will take place, the grounds on which it will be conducted and how it will be funded, is hugely beneficial to all parties. For all these reasons it is critical for evaluation to be considered as part of a *process*, but one in which all parties are aware that evaluation is incorporated at all stages.

A second key phrase in the definition of evaluation by Papaconstantinou and Polt is “*as systematically and objectively as possible*”. Given that evaluation traditionally takes place “at the end of the line” there are likely to be strong entrenched interests in place once a programme has been in existence for a number of years. These entrenched interests include the direct beneficiaries of the programme, such as the businesses receiving funds, but they will also include those who are responsible for initiating and administering these programmes. All else held equal, it is to be expected that all these groups will choose the programme to continue or expand. The task of the evaluator, however, is to “*systematically and objectively*” assess the merits of the programme. In this task the evaluator is likely to conflict with those committed to the programme *per se*. Only through the use of objective techniques can the evaluator demonstrate independence to those delivering programmes, with many of these techniques being discussed later in the paper.

Nevertheless, whilst objectivity is important, the evaluator’s task is best conducted with the support of those delivering the programme. Whilst the evaluator is expected to be knowledgeable about evaluation, those delivering the programmes have valuable insider knowledge. They will often be able to provide simple but unexpected explanations for the “associations” identified by the evaluators. Without their knowledge and support there is the risk of ill-informed inferences being drawn. Hence there is a fine line between the need for the evaluator to exhibit objectivity and independence, yet also to gain the support and inside knowledge of those delivering the programmes. This support is much more likely to be forthcoming if, when the programme is devised, it is made clear to those delivering programmes that an evaluation will take place.

The third key phrase in the Papaconstantinou and Polt definition is “*the relevance, efficiency and effect of an activity in terms of its objectives*”. The implicit assumption in this statement is that the policy has clear objectives that these are stated in sufficiently clear terms for them to be used by the evaluator. In practice this is often not the case. Objectives are often specified in generic terms, such as a desire to make an economy more “enterprising”, as shown earlier in paragraph 19, but this is of limited value to the evaluator. Taking the simple example of an objective of making an economy “more enterprising”, the evaluator could consider a possibly bewildering range of measures that could be used to assess policy impact. These might include, the number of new firms started, those started by young or disadvantaged people, those started by people who have not been in business before, the increase in the total number of firms, whether these firms grow, whether they internationalise and of course whether they survive. All could be considered measures of whether a country is “enterprising”, but it is likely that a country will perform better on some measures than others. Where there are a wide range of possible measures, in some of which the policy performs well and in others it performs less well, there is an inevitable tendency for those favouring continuation of the programme to select the measures or targets on which the programme has performed well. In contrast, those favouring programme closure or contracting will select their chosen measures. Richard Harrison and Claire Leitch (1996) delightfully encapsulate this in an essay, “Whatever

you hit call the target: An alternative approach to small business policy". It emphasises the importance of being clear on the targets that most closely reflect the objectives of the policy³.

The reality is, however, that the measures by which the effectiveness of the programme is assessed are often not specified. The evaluator therefore has to "second guess" the original objectives of the legislator when framing the programme. Frequently the evaluator has to, many years after the legislation was implemented, identify a number of measures that reflect the original objective of the legislation. This then leads inevitably to a range of performance measures being chosen, probably in conjunction with those responsible for implementing the programme.

In principle such discussion is of real value and does reflect the role of evaluation as a "*process*". In practice, unfortunately, it leads to those responsible for the programme identifying those measures upon which they have performed relatively well as being those upon which the programme should be evaluated. Frequently this leads to a somewhat unseemly debate from which neither the evaluators nor the programme managers emerge with credit.

The central conclusion from this review of Evaluation context is the integral role it has to play in the policy *process*. Evaluation cannot be left "at the end of the line". Instead, it has to be a key element of initial policy formulation. Once the policy is operational, all organisations and individuals responsible for delivery have to be aware that evaluation is to take place. Once the evaluation has been undertaken it should be used as the basis for dialogue with policy makers, with the objective of delivering better policy. The outcome of the evaluation can then become an input into a philosophical debate on the appropriate ways for governments and SMEs to interact.

The evaluation framework

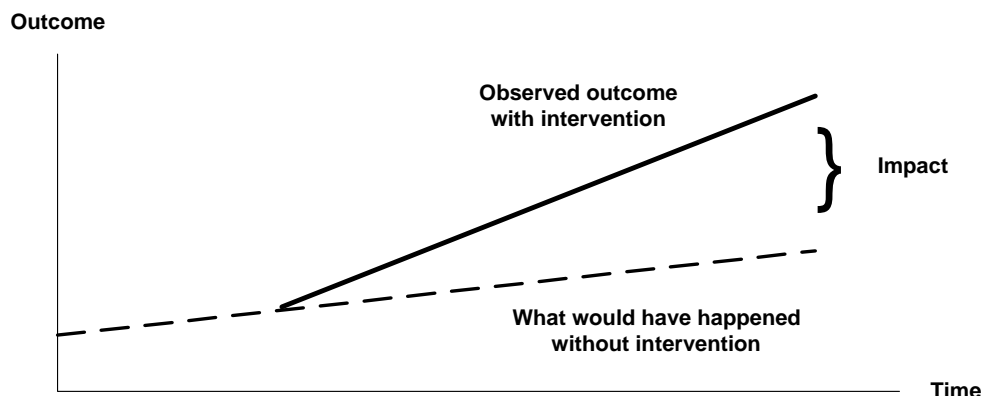
The SME Working Party has identified seven headings under which policies can be assessed. These are: Rationale, Additionality, Appropriateness, Superiority, Systemic Efficiency, Own Efficiency and Adaptive Efficiency.

Key to this assessment is the issue of Additionality. This is defined as the true impact of the scheme/programme. Whilst it is not always easy to quantify, it is likely to be reflected in a measure such as additional output, employment, sales or export activity that can be attributed to the existence of the programme. In other words activity that would not have taken place without the programme but is attributable to the firm participating in the programme. This is shown in Figure 3 below taken from Oldsman (2002).

It shows, for any given outcome, that policy impact can be considered as the difference between the observed outcome with the intervention and what would have happened without the intervention. The Figure shows these two outcomes diverging after the time when the policy is implemented.

³ This is, of course, not to imply that targets cannot be changed. Indeed a frequent desirable outcome from an evaluation is that it can lead to a change in targets. See the example in Box 3. What, however, is undesirable is for a number of targets to be selected and for policy makers to claim programmes to be successful whichever target they happen to hit at a particular time. It emphasises the importance of being clear on the criteria for unsuccessful policy.

Figure 3. The Impact of an Intervention



Source: Oldsman (2002)

Whilst this is a simple concept, identification of the programme impact “*as systematically and objectively as possible*” can be challenging. This is because programme impact cannot be easily observed for a number of reasons. The first is that it is not always clear what changes might have occurred in the firms as a result of participation; in other words the Outcome measures are unclear. Some programmes might be expected to lead to a greater likelihood of firm survival, others to growth in sales, profits or employment, others to the greater likelihood of innovating or selling into overseas markets. Other programmes might be expected to enhance all these characteristics and in other cases it is unclear what firm characteristics are expected to show improvement. Evaluation therefore requires a decision on appropriate Outcome measures.

A second problem is that participation in the programme will precede improvement. Using the example of Figure 3, the point at which the lines diverge will not be immediately after the programme is delivered. Some programmes will have their impact possibly years before others. As an example, a programme in which SMEs are subsidised to participate in an international Trade Fair so as to encourage them to internationalise⁴, might be expected to have a quick impact upon sales. In contrast, a programme to fund Research and Development in SMEs would be expected to have little impact for several years.

A third problem, is the myriad of influences upon the performance of an SME, other than that of programme participation. These include the skill of the owner, the sector and location of the business, macro-economic conditions and the role of chance. In principle, only when account is fully taken of these “exogenous” factors can the impact of the programme be estimated. The next section shows the various ways in which these issues have been addressed.

The Six Steps Approach

Many approaches to the challenge of impact assessment have been adopted, but it is possible to characterise them within the “Six Steps to Heaven” framework [Storey (2000)]. This is shown in Table 1 below.

⁴ "Access to International Markets", a presentation by the Delegation of Greece, OECD Working Party on Small and Medium-Sized Enterprises, Athens, 26-28 April 1998.

Table 1. Six Steps to Heaven: Methods for Assessing the Impact of SME Policy

Monitoring		
STEP I	*	Take up of schemes
STEP II	*	Recipients Opinions
STEP III	*	Recipients' views of the difference made by the Assistance
Evaluation		
STEP IV	*	Comparison of the Performance of 'Assisted' with 'Typical' firms
STEP V	*	Comparison with 'Match' firms
STEP VI	*	Taking account of selection bias

The "Six Steps" procedure recognises that there is a gradation of sophistication in the assessment procedures. The six steps show a gradation, with Step I being the least sophisticated procedure and Step VI being the most sophisticated. Here sophistication is intended to reflect the confidence the policy maker has in being able to attribute changes in businesses supported under the programme – often called the "treatment group" – to participation in the programme. Alternatively expressed, it may be considered as the confidence the policy maker has that all other influences are held constant.

Table 1 shows a distinction is made between Monitoring and Evaluation. Steps I to III are not considered to be Evaluation, but instead are referred to as Monitoring. The difference between Monitoring and Evaluation is that the former relies exclusively upon the views of the recipients of the policy. Evaluation however seeks, by some means, to contrast these views or actions with those of non-recipients in order to present the "counter factual" shown as the dotted line in Figure 3. The difference between actual changes and the "counter factual" is viewed as the impact of the policy – or its 'additionality'.

The remainder of this section will very briefly sketch the Steps in order to justify the use of the most sophisticated approach to evaluation. Table 2 shows that, whilst the characteristics of the firms that participate in programmes is valuable information, it does not help the policy maker in assessing whether the programme has enhanced the performance of the participant firms.

Table 2. STEP I: Take Up of Schemes

<p><u>QUESTIONS</u></p> <ul style="list-style-type: none"> * How many firms participated? * What sectors were they in? * What locations were they in? * How big were these firms? * How much money was spent? <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Tells you almost nothing about policy effectiveness. * Tells you almost nothing about satisfying objectives.

One simple, and very popular way to address this issue is to ask the firms that participated in the programme as shown in Table 3⁵. Whilst this can provide helpful insights into programme delivery, it has important limitations for assessing overall programme impact. Some respondents, if their business has done well since participating in the programme, may view this as reflecting their own personal achievement and so be very reluctant to attribute any change in their business to participation in the programme. In contrast, others may view it as in their interests for the programme to continue and so exaggerate the benefits to them. A third group may simply not know what the impact was, given all the other changes in the business. Unfortunately those surveying opinions of firms do not know which of these three responses is dominant. Finally, these surveys are much more likely to include surviving, rather than non-surviving SMEs. All else equal, the survivors are likely to be more positive, leading to an upward bias in programme estimation.

All these factors mean that simply asking for the views of recipients runs the risk of conflicting bias that it is difficult to address. Nevertheless this has not prevented many Step II “evaluations” being undertaken, presumably on the grounds of satisfying a requirement to conduct evaluation, and seeking to achieve this at lowest possible cost. The problems with this approach are, however, so considerable that they should be discouraged other than for the provision of “case-study” insights.

Table 3. STEP II: Recipients' Opinions

<p><u>QUESTIONS</u></p> <ul style="list-style-type: none"> * Programme participants: Did they like it? * Firms: Were there problems in applying the programme? <ul style="list-style-type: none"> : Were procedures too slow? : Cumbersome? <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Even if they like it, it does not tell you if it is effective. * All it can do is offer insights into policy delivery. <ul style="list-style-type: none"> - but that is not the key question.
--

Step III is a slightly more sophisticated approach. It still seeks to obtain answers from programme beneficiaries, but explicitly addresses the issue of the counter-factual. It asks programme recipients what would have happened to their firm if they had not participated. In this sense the respondent has to address the counter-factual.

⁵ As an example, see "A Thematic Evaluation of the Impacts of the European Commission Structural Fund on SMEs", Paper presented at OECD, November/December 1998.

Table 4. STEP III: Recipients' views of the difference made by the Assistance

<p><u>QUESTIONS</u></p> <ul style="list-style-type: none"> * Did firms think it provided 'additionality'? * Would firms have done it anyway? * Does it cause 'displacement'? <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Provide answers they think you want to hear. * No way of checking. * Only snapshot of surviving firms.

However this requires considerable mental gymnastics on the part of the respondent. Some may be able to do this complex calculation but other will not. Unfortunately the researcher has a highly imperfect idea of which respondents are in which category. These problems multiply when the respondent was not with the firm when the programme began.

For all these reasons Step III is also highly imperfect and emphasises the need to have an explicit counterfactual. This characterises Steps IV to VI, which are defined as Evaluation.

The simplest Evaluation is where the performance of "treatment" firms – those participating in the programme – are compared with another group of firms that have not participated. The logic of this comparison is that the difference between the "treatment" and the "non-treatment" firms is participating in the programme, and so this is reflected in observed differences in performance. The non-treatment firms in Step IV are usually other firms in the economy or locality.

Table 5. STEP IV: Comparison of the Performances of 'Assisted' with 'Typical' firms

<p><u>APPROACH</u></p> <ul style="list-style-type: none"> * Employment/Sales growth of assisted firms compared with 'typical' firms. * Survival of assisted firms compared with 'typical' firms. <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Assisted firms are <u>not</u> typical.
--

But, of course, programme participants may differ systematically from other firms in the economy. They might differ in size, in location, in sector, in age or other observable ways, but all of which may systematically influence their performance. These factors, since they can in principle be identified and measured, are referred to as "observables". Performance differences between the two groups may occur therefore, not only because of participation in the programme, but also because these "observables" are not held constant in Step IV. Failure to take account of these observables would mean, in terms of Figure 3 that the dotted line was incorrectly drawn. Some of the "impact" of the Programme would actually be a reflection of the characteristics of the firms, meaning the true programme impact is either over- or underestimated.

Step V seeks to address the issue of “observables”. It normally selects key observables such as ownership, size, sector and geography and ensures, through a matching process, that the treatment and the non-treatment firms do not differ from each other in these respects. Once the matching is complete then differences in the performance of the two groups is attributed to participation in the programme.

Table 6. STEP V: Comparison with ‘match’ firms

<p><u>APPROACH</u></p> <ul style="list-style-type: none"> * Compare assisted with ‘match firms on the basis of: <ul style="list-style-type: none"> – age; – sector; – ownership; – geography. * Compare performance of both groups over same time period. <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Perfect matching on four criteria can be very difficult. * Sample selection bias: <ul style="list-style-type: none"> – More ‘motivated’ firms apply; – Attribute differential performance to scheme and not to motivation.
--

Whilst the Step V approach reflects a real attempt to address the issue of the counter-factual, it cannot be considered as “best practice”. The results are still somewhat ambiguous because observed differences in performance may also reflect the presence of “unobservables”. The most obvious of these may be dynamism or motivation of the SME owner, with this leading to various forms of selection bias in the estimation techniques.

Two examples of selection bias are self-selection and committee selection. Self selection bias may occur where, for example, programmes seek to provide support for rapidly growing businesses. Here those businesses that are seeking growth may be more likely to apply to become participants in the programme than those with only modest or with no growth ambitions. The problem arises where this is only partly being reflected in “observable” factors, since the latter under-estimate how the firm would have got on even if the programme had not existed. The “counter-factual” dotted line, in terms of Figure 3, is specified too low and the programme impact is consequently over-estimated.

Alternatively expressed, applicant firms, because they have more motivated owners, would have been expected to perform better than the “matches”, even if there had been no programme. It is therefore unreasonable to attribute all differences in performance to programme participation, but the precise proportion attributable to the programme is unclear.

A second example is committee selection. This refers to programmes where only a proportion of applicants is successful. Here a committee or similar group makes a judgement, with the “better” firms/applications obtaining the funding. If the committee is effective it will eliminate those firms likely to perform poorly, so that the remainder are clearly non-randomly drawn from applicants. As before, even if the programme did not exist, the selected firms would be expected to have outperformed the other firms. Hence observed differences in performance between programme participants and matched firms cannot be attributed solely to programme participation.⁶

Table 7. STEP VI Compare assisted with ‘match’ firms taking account of sample selection

<p><u>APPROACH</u></p> <ul style="list-style-type: none"> * Use of Statistical Techniques: Heckman 2 Step: Estimator for testing and adjustment. * Use of Random Panels. <p><u>PROBLEMS</u></p> <ul style="list-style-type: none"> * Policy makers (and some academics) feel uneasy about statistical ‘adjustment’. * Use of random panels could mean public money is given to firms/people who we know will not benefit.

There are now a number of statistical techniques that address the issue of selection bias. It is a curiosity that, although these techniques are now well established in the evaluation of labour market and welfare programmes, they are much less frequently used in estimating the impact of business support. A helpful review of these techniques is provided by Smith (2002) and shown in Box 4 below. As Smith shows it is now technically possible to address the issue of selection and so deliver Step VI assessments.

⁶ It is of course possible that the Committee may not be good at selecting the good from the bad firms. In this instance the Committee may serve no useful function. Alternatively the Committee might play a different role in ensuring that a quota of certain types of firms receive funding. These might be firms from certain geographical areas or whose owners are disadvantaged in some respect. If that role is clear than the Step V procedure would address this. However, if it is not explicitly the role of the Committee then this could lead to the programme effects being under-estimated using Step V.

Box 4. Can we use statistical methods to make non-random comparison groups truly comparable?

There are at least five statistical techniques, not necessarily mutually exclusive, that can be used to detect the true effect of a program on some outcome variable, when the program users differ in other ways than program use from the nonusers. First, we can simply statistically control for observed variables that affect the economic outcome and might be correlated with program use by including these observed variables in the estimation equation that is used to predict the outcome variable. This approach is most effective in reducing the bias in estimation of program impacts when we have data on as many variables as possible that affect the economic outcome of interest and are correlated with program use. This approach cannot correct for biases that might be caused by unobserved variables that are correlated with both economic outcomes and program use. This approach also assumes that we know the functional form by which the observed variables affect economic outcomes.

A second approach that is a variant of the first goes under the label of “difference-in-differences” estimation, or “difference-in-differences-in-differences” estimation (DD and DDD for short) (Meyer 1995). Under a DD approach, we compare the difference before and after the program of the differences between users and non-users of the program or policy. Under a DDD approach, if we have reason to think that some types of users are likely to be more affected by the policy than another, we can compare the difference between the likely high impact and low impact groups in the user and non-user group before and after the policy. A DD approach is equivalent to assuming that one can do a good job of controlling for other factors affecting economic outcomes by allowing for effects of the time period, whether the entity is in a user or non-user group or a high-impact or low-impact group. The limitation of this approach is that there may be many other variables, both observed and unobserved, that also affect economic outcomes and are correlated with program use. The second approach can be combined with the first approach by adding some of these other observed variables to the estimation equation.

A third approach is matching program users with non-users who are similar in observed characteristics. Recent research has revealed that this matching should focus on finding users and non-users who are as similar as possible in their estimated “propensity score,” which is an estimated probability given observed variables that a given entity will use the program (Smith and Todd 2001; Heckman, Ichimura, and Todd 1999). This approach works well if we have data on all the variables that do a good job of predicting program use and are also correlated with economic outcomes. This matching approach will not work well if there are many unobserved variables that predict program use and are correlated with economic outcomes. In addition, in many cases there may be no reasonably close matches for some users with non-users, and the estimates from a matching approach hence are only valid as average program effects for the types of program users for which we can find good matches among non-users.

A fourth approach is explicitly modeling selection into the program and how it is correlated with unobserved variables affecting economic outcomes (Murnane, Newstead, and Olsen 1985). This requires estimating three equations: one equation explaining economic outcomes for program users, a second equation explaining economic outcomes for non-users, and a third equation explaining whether a given entity is a program user. The estimation of the third equation allows a “selection bias correction” term to be added to each of the first two equations, which in theory corrects for the bias caused by unobserved variables that affect economic outcomes and are correlated with program use. This approach assumes that we have accurately specified the variables and functional form that should enter all three equations. In addition, this approach assumes a particular statistical distribution for the unobservable factors (the “error term”) that enter all three equations.

A fifth approach requires finding some “instrumental variable” that predicts program use and is uncorrelated with unobservable variables that affect economic outcomes (Angrist and Krueger 2001). Under this instrumental variable approach, we only examine the change in economic outcomes that can be attributed to shifts in program use that are statistically associated with shifts in the instrumental variable. The intuition is that the effects on economic outcomes of these instrument-induced shifts in program use show the true effects of the program because these shifts in program use will be uncorrelated with unobservable variables predicting economic outcomes, as these shifts are generated by an instrumental variable that is uncorrelated with unobservable variables predicting economic outcomes. The problem is finding such instruments. The instrumental variable must do a good job of explaining program use, otherwise the estimation approach throws away too much information. But the variable must have little (ideally, zero) correlation with unobservable variables affecting economic outcomes, and it is difficult to test assumptions about the correlation of a proposed instrument with unobservable variables. Good instruments are hard to find and may not be convincing to all readers.

These five approaches can be combined in different ways. For example, researchers can create matched data sets, include controls for various observed variables in the estimation, and use instrumental variables for the program variable.

Source: Smith (2002).

Some modifications and limitations to the Six Steps approach.

The Six Steps Framework has benefited from comments made by the Japanese delegation to OECD. Their views are reported verbatim in Box 5 and may be summarised by saying that the most sophisticated – Step VI – approaches may be more appropriate in some circumstances than in others. They argue that in some programmes the causal link between programme participation and firm performance is clearer than in others.

Box 5. Comment on Cross-Cutting Theme: Evaluation of SME Policies and Programmes by the Delegation of Japan to the OECD's Working Party on SMEs and Entrepreneurship

The “Six Steps to Heaven methods for assessing the impact of SME policy” is an excellent framework to assess the additionality of programmes.

However, considering the variety of objectives and schemes of SME policy programs, the paper should address the issue that the appropriate monitoring or evaluation steps may vary by programme. The appropriate steps should be determined programme by program considering the causal relation between output and outcome and evaluation cost as well as the budget scale of the programme.

For example, in evaluating programmes such as provision of information and training, it is difficult and expensive to measure the impact for Step IV evaluation because the causal relation between output (information, training) and outcome (sales, employment, export, etc.) is rather vague and affected by other external factors. For these programmes, Step II or Step III seem to be appropriate targets and further steps might be too much.

On the other hand, it is rather easy to measure the impact of programmes such as grants and loan guarantees, because such programs are directly related to investment activities. For these programs Step V or Step VI seem to be appropriate targets.

The following is an image of the appropriate target steps and should be discussed by experts.

Programme	Monitoring			Evaluation			Causal relation	Measurement
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	output – outcome	cost
Provision of information	<input type="radio"/>	<input type="radio"/>					weak	high
Seminar, Training	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				weak	low
Consultation, Mentoring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				weak	low
Grant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	strong	low
Loan Guarantee	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	strong	low
Equity programme	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	strong	low
Tax incentive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		strong	high

This raises the very important issue of whether the same level of sophistication should be used in all programmes, irrespective of their scale and complexity.

The argument for seeking Step VI evaluations in all cases is that if some programmes are evaluated in a less comprehensive manner than others this causes problems in comparing effectiveness across all programmes. Assume SME policy has only two programmes. Programme A provides Grants and Programme B provides Training. If policy is delivered efficiently the marginal effect per unit of currency will be identical for the two programmes. However, such a judgement would be impossible to reach if Programme A were assessed in one way and Programme B in another.

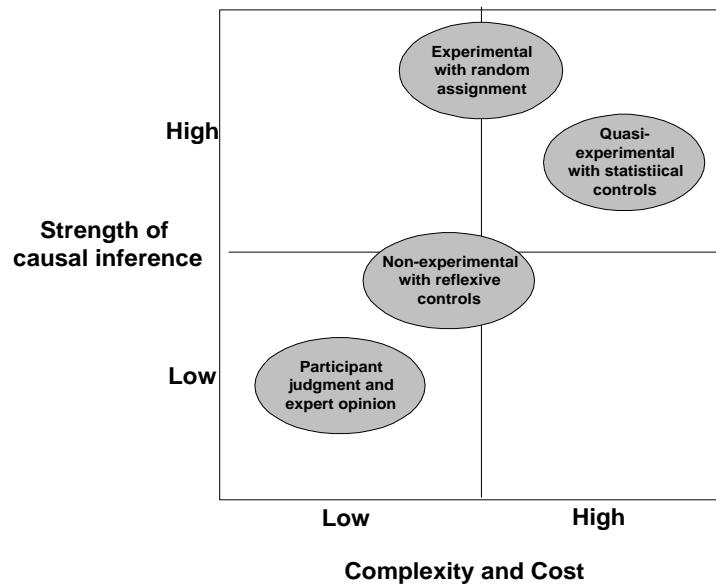
This is a very important point since the Six Steps approach shows that it is the more sophisticated processes that are the most likely to specify accurately the “counter-factual”. In most instances inaccurate specification of the counter-factual leads to programme impact being over-estimated. There is a clear pattern that programmes evaluated using the low numbered Steps approach are attributed to have a bigger impact than those using the higher numbered Steps approach. Unfortunately the studies using the more sophisticated approaches – those with higher Step numbers – are also much more expensive to conduct. Conducting an interview with a programme participant is easy and cheap and requires almost no analysis. This contrasts starkly with the more sophisticated approaches where evaluation needs to be built in from the start, non-programme participants as well as programme participants need to be monitored, and the statistical expertise to conduct the analysis is likely to be rare and expensive. There is therefore an almost ethical issue facing those charged with a responsibility for Evaluation. That is that the more sophisticated approaches are not only likely to generate lower programme effects, but that they are more expensive and probably more lengthy to conduct. This dilemma is discussed further in Section (vi).

The central problem therefore is that if some programmes were evaluated in a less challenging manner than others, it would lead to problems in allocating funds between programmes within the SME policy area. Equally, however, it is difficult to justify the same level of sophistication and hence evaluation budgets being applied to small as to large programmes. This is illustrated in Figure 3 below. It shows that impact accuracy increases with budget.

This highlights an important trade-off in evaluation. Ideally all programmes within SME policy should be assessed using the same methods in order to validly compare their marginal effects. In this way it may be possible to transfer resources from Programme A to Programme B and enhance the overall impact of SME policy.

The problem, however, is that it is difficult to justify the same level of analytical sophistication for small, as for large programmes. But if the smaller programmes are less rigorously assessed then they may appear to have a larger impact and so justify budget increases to the detriment of larger programmes/projects.

Figure 3. Trade-Offs in Evaluation Design



Finally, it is also important to emphasise that the “Six Steps” approach is designed to examine the impact of individual programmes and policies. Since it operates by comparing firms that participate in programmes or policies with those which do not, it cannot be used to examine policies that apply to *all* firms in the economy simultaneously. Hence tax incentives that are taken up by all SMEs are not appropriately examined using this approach. It is equally true that the totality of the impact of enterprise programmes could be underestimated using this approach where there was strong inter-dependency of programmes. Essentially the “Six Steps” approach estimates the marginal effects of individual programmes, but the sum of the marginal effects could be less than the total effects. It would be, however, unwise for these rather technical arguments to obscure the need for taxpayers to be satisfied that impacts of individual programmes were being assessed by the most technically sophisticated means.

Examples of Good Practice Evaluation

This section will provide examples of SME policy evaluations in OECD countries that achieve at least STEP V. These studies examine the performance of ‘assisted’ firms and compare their performance with otherwise similar firms that were not assisted. Depending upon the ‘controls’ used, this will provide a measure of additionality. For example, if the policy were to provide loans and grants to a group of SMEs, at a minimum it would be necessary to ensure that the assisted and non-assisted firms did not differ clearly in terms of size, sector, age, geography or prior growth.

However, as noted in the section above, statistical developments in recent years have enabled analysts to also address a potential problem that the Step V procedure ignores. The problem is that there may also be ‘unobservable’ factors, which influence the willingness of businesses to apply for grants, as well as “observable” factors such as age, sector ownership and geography. A second source of bias addressed by the Step VI procedures is committee-selection bias. This bias occurs because, in many programmes, there is insufficient funding for all applicants. In this case only some applicants are

successful, with a decision being made by funders on the basis of whether or not the business is likely to prosper. If those responsible for making this decision – the committee – are unable to distinguish the successful from the less successful, then this is a major source of bias. This is because the firms that obtain the support would have performed better than the firms rejected for support, even if the support had not existed.

The study in Box 6 is abstracted from the review by Oldsman (2002), and shows the impact of training and consulting advice provided by IRCs to SMEs in the United States. Using a matching procedure it finds a positive impact on the productivity of programme participation. What is less clear is whether this explicitly addresses the selection issue.

Box 6. Industrial Resource Center (IRC) program.⁷

The program was established in 1988 to help small and medium-sized manufacturers upgrade business practices and modernise their production capabilities in order to spur economic growth in Pennsylvania. The IRCs are designed to accomplish this mission through a comprehensive set of activities involving a combination of consulting and training services. Since its inception, the state government has committed roughly \$84 million to the programme.

A comprehensive evaluation of the program was conducted in 1999. The evaluation included an assessment of the impact of services on participating companies with respect to growth in productivity and output. To help control for potential selection bias, the analysis employs a two-stage procedure.⁸ The first step involves estimating the probability of companies becoming IRC clients as a function of characteristics of the firm. The second step involves estimating the impact of the IRC program on companies after controlling for factors that affect productivity and output growth as well as potential selection bias. The estimated model is based upon a modified Cobb-Douglas production function which includes plant specific factors.

The analysis is based on panel data for individual manufacturing plants – the Longitudinal Research Dataset (LRD) – maintained by the Center for Economic Studies at the US Bureau of the Census. The dataset provides detailed plant-level data on shipments, employment, factor costs, industry, and other legal and administrative identifiers. It is compiled from the Census of Manufactures carried out every five years and the Annual Survey of Manufactures. The LRD was used to obtain data for both clients and non-clients in Pennsylvania. Companies included in the IRC administrative database were linked to the LRD using a matching procedure developed by the Center for Economic Studies. The matching process identified 2 839 unique IRC client establishments in the census years based on its permanent plant number (PPN). The comparison group included a similar number of companies.

The analysis demonstrated that the program had a significant impact on IRC clients. For the pre-92 cohort, the difference in the growth rate in output and productivity directly attributed to IRC services is estimated at 1.8 percent and 3.6 percent per year over a ten-year period.

⁷ Nexus Associates, Inc. (1999).

⁸ See the following publications for additional information on procedures used in this analysis: Maddala (1994), Jarmin (1997), and Heckman (1974 and 1979).

Box 7. Small Business Innovation Research (SBIR) Program.⁹

The SBIR program was established by US Congress in 1982. The authorising legislation mandated that all federal agencies spending more than \$100 million annually on external research set aside a fixed percentage of these funds for awards to small businesses. Over time, the percentage has increased to 2.5 percent. Between 1983 and 1995, small firms received more than \$6 billion under the program.

An evaluation of the program was conducted in 1996, focusing on the impact of SBIR funding on sales and employment growth. The analysis involved comparing the performance of firms that received SBIR grants to similar companies that did not participate in the program, controlling for firm age, geographical location, prior venture financing, and overall venture capital activity in the region and industry. The sample consisted of 541 firms that received SBIR Phase II awards in the first three program cycles and 594 matching firms.

Data was compiled from program records and information contained in publicly available directories and databases.

The analysis demonstrated that SBIR awardees experienced greater growth in both sales and employment than similar firms, but these effects were confined to areas that attracted significant venture financing.

The very large US programme on SBIR, in which grants are awarded to SMEs to conduct Research and Development is reviewed in Box 7. This again is abstracted from Oldsman, but draws upon the original study by Lerner (1996). This study finds a positive effect of participation in SBIR, but uses a Step V approach. It is difficult to guess the impact of including the effect of Selection, since the programme has both Self- and Committee selection present. Although, as noted above, the effect of excluding selection is normally to enhance the value of the programme, this may not occur in this case since equitable geographical distribution of awards could be present within the remit of the committee.

A second concern with SBIR is the extent to which it provides real additionality to the R&D budgets of small firms, and the extent to which private capital is “crowded out”. The commitment of the United States to examining these issues is reflected in the National Research Council being asked to investigate these issues.¹⁰

Box 8 shows a UK study that attempts to estimate the impact of “soft” services provided by publicly funded Business Links. It shows that, using a Step VI approach, there is some tentative evidence of an impact upon the productivity of firms. However, it does not find a significant effect upon turnover or employment of the assistance provided.

⁹ Lerner, Josh (1996).

¹⁰ Charles Wessner, "US support for Innovation and Small Business: Measuring SBIR Outcomes", Paper presented at 19th session of OECD Working Party on Small and Medium Sized Enterprises and Entrepreneurship, 9-10 December, 2002.

Box 8. Measuring the Productivity Effects of Business Links¹¹

Business Links are the UK- based “one stop shop” for government supported (small) business advice and consultancy.

The purpose of the evaluation was to assess whether firms assisted by Business Links had performed better than non-assisted firms. Since a Business Link (BL) objective was to target firms employing between 10 and 200 workers with growth potential, a second objective was to assess the effectiveness of this targeting. The study tracks 137 firms assisted by BL and 196 comparison firms between 1996 and 2000. It identifies the factors which influence whether or not the firm applies for assistance from BL and, taking this into account, the impact which the BL advice has upon performance.

- It finds there is no evidence that in 1996 that Business Link were targeting faster growing firms.
- It finds tentative evidence that BL assistance in 1996 was having a positive effect on productivity growth.
- It finds a positive but statistically insignificant effect of BL assistance on turnover and employment was identified.
- The analysis has highlighted a number of other factors, which contribute to productivity, turnover and employment growth. The range of these factors – embracing market conditions, business strategy, the characteristics of the owner manager and the firm itself. This emphasises the complexity of the process of business growth and the consequent difficulties in both modelling and assisting the process.

In other instances good practice relates to seeking to evaluate the impact not of a single programme, but instead to a group. Box 9 takes the example of New Zealand and notes that different approaches are used to examine different parts of the policy package.

¹¹ Paper Presented by George Bramley , DTI, to OECD SME Working Party, 17th session, 2001.

Box 9. Evaluation of the *Growth and Innovation Framework* and enterprise development initiatives in New Zealand has three key components:

- *The Growth and Innovation Indicators Report* – this annual report brings together baseline information on a set of indicators intended to track progress in high-level growth and innovation outcomes.
- *The Annual Stocktake of Enterprise Development Programmes* – this annual report captures information on the efficiency and effectiveness (against programme objectives) of enterprise development programmes. It is a broad evaluation designed to capture new information and learning that has come to light during the year. It also provides, where possible, some comments on emerging conclusions about the impacts of programmes.
- *In-depth reviews of individual programmes* – these are comprehensive reviews of the policy foundations underpinning the programmes, the impact of the programmes, and the efficiency and effectiveness of programme delivery. Reviews completed in 2003 were the Regional Partnerships Programme review and the Investment Ready Scheme review.¹²

Box 10 shows the example of the Canadian SBLA evaluation. This seeks to provide public funding to underpin asset based lending from financial institutions. The challenge posed to the evaluator is to demonstrate that the lending would not have taken place without the SBLA.

This places a heavy reliance upon the views of the stakeholders such as the banks, government agents and loan recipients, all of whom cannot be regarded as disinterested parties. Nevertheless there is no clear “non-treatment” group with which to make comparisons. Hence the use of a variety of approaches is good practice where they point in broadly similar directions.

Box 10. Small Business Loan Act (SBLA) Canada¹³

SBLA began in Canada in 1961 to address the difficulties faced by small firms in asset-based debt financing. The programme seeks to use public money to leverage private funding from financial institutions. Government seeks to minimise “Deadweight” and “crowding out”.

Given that SBLA had a long history, albeit with a number of policy changes over the years, a review was established in 1996. Demonstrating that small businesses would have not received a loan without the SBLA is difficult, using any single methodology. A wide range of methods was therefore used including case studies, econometric methods and stakeholder consultations.

The evaluation concluded that up to 86% of SBLA lending was incremental. In most cases this was because the loan would not have been made at all or, if they had, that the terms would have been harsher for the SME or the procedure would have been slower.

¹² Davis, Nick, “Best practice policies for SMEs: Country presentations on SME programme evaluations, New Zealand: evaluation of Investment Ready Scheme”, DSTI/IND/PME (2003), 11 May 2003. “Regional Partnership Programme Review”, September 2003.

¹³ “Evaluation of Small Business Loans Act (SBLA)”, paper presented at Working Party on Small and Medium Sized Enterprises, OECD, Helsinki, 20-21 May 1999.

Finally, Box 11, drawing on the work of Lars Barger-Sjogren (2003), shows an example from Sweden of a Step VI procedure being used to evaluate a major programme of “on the job” workforce training.

Box 11. On the Job Training in Sweden

This sophisticated study defined competitiveness to be productivity, gross profit margin and capital structure. It examined more than 2000 firms including both treatment and non-treatment firms. It also addressed the issue of selection bias.

The standardised measure of productivity used showed that, prior to the programme in 1995 the treatment group were lagging behind the control group by almost 30%. At the end of the programme in 2000, however, the gap has been significantly reduced.

Lessons

This section interprets the above findings, seeking to relate them to the concerns of SME policy makers.

- Funding for public programmes to assist SMEs comes from the taxpayer. Those responsible for developing and delivering such programmes therefore have a responsibility to ensure that value for money is generated. The sums of money are not trivial, and the total public budgets on SMEs are rarely quantified, primarily because there are so many such public programmes. An exception is a recent study, summarised in Appendix I, which reported that the total annual UK Government Expenditure on providing services to Small Business for the most recent year available was approximately £8 billion. To place this in context, the UK public expenditure on Police was £7 billion. This emphasises the importance of being aware of the total budget and of ensuring value for money is derived from programmes. Evaluation is expected to play a key role in informing politicians and taxpayers.
- The example above demonstrates not only that the sums of public money are substantial, but also that responsibility for that money is widely distributed amongst different departments within government. This constitutes a major policy challenge because whilst Departments may have “budget lines” for small businesses, their overall Departmental priorities may lie elsewhere. This means that whilst the total public expenditure on small firms may be substantial there is a strong risk that it lacks coherence and focus. This is reflected in the recommendation of the Finnish delegate Annukka Lehtonen drawing upon the outcome of “Workshop on Micro Policies for Fostering Entrepreneurship”, held in Helsinki in April 2002, together with the government entrepreneurship project. The quotation below reflects the key concept of “policy coherence”

Box 12. Policy Coherence¹⁴

“The key to successful enterprise policy is true, horizontal cooperation within which various ministers and ministries work together. It means appointed officials of various ministries are responsible for ensuring that legislation is drafted to promote entrepreneurship and that the impact of various policy measures on entrepreneurship is taken into account.

True, horizontal cooperation means that, besides ordinary administrative cooperation, responsible officials are combined from various ministries, who have the Government’s mandate to act as a “policy entrepreneur”. The task of the responsible officials is to act proactively, answering for the preparation of measures promoting the business activities of their own ministries. This means that the preparatory responsibility of the various ministries is not changed, but the responsible officials are to verify that the work is actually taking place. The aim of this is to knock down the barriers between the administrative branches, which are traditionally closely watched at the ministries. For the decision-making we also need a working group of sectoral ministers or ministers representing the government parties”.

- Given the goal of policy coherence, this report has re-emphasised previous OECD recommendations that Evaluation has to be an integral part of the policy process. It is not an optional "add-on". Instead, considerations of evaluation need to be incorporated in all stages of policy development. This means that when legislation is being framed that objectives should be set so that evaluation can be conducted. It may also mean that an evaluation methodology and budget is formulated, and even that data collection may begin.
- The report also makes it clear that the last few years have seen major statistical advances that mean that it is now possible to undertake much more accurate assessments of programme impact than was the case only a few years ago. It is no longer valid to argue that assessment procedures are very imperfect. Quite simply, the technology is there and is used extensively in many other aspects of public programmes, such as labour market programmes. It is curious that these techniques have been slow to develop in SME policy assessment, but there is now experience of them being developed for several types of SME programmes.

Nevertheless there remain important issues for discussion and debate. These are outlined below.

Whilst the statistical techniques available are able to much more accurately assess policy than has been the case in the past, there is some nervousness amongst politicians and public servants about their widespread adoption. This is partly because of the mathematical and quantitative skills needed for them to be understood are not available to all; it is partly because of widely held mistrust of attempts to quantify policy impact on the grounds that this leads to the "softer" aspects of policy delivery being ignored. For some groups, therefore, no matter what improvements are made in the accuracy of quantitative techniques, there will always be a real reluctance to their widespread adoption.

Other groups may have less of a "root and branch" objection, but continue to have misgivings about the implementation of more sophisticated evaluation. These issues are addressed below.

¹⁴ Extract from a presentation by the Finnish Delegation to the OECD Workshop on Entrepreneurship, Budapest, Hungary, 8-10 September 2003.

(i) *The link between the technical quality of evaluation techniques and the estimated programme*

There is clear evidence that the more sophisticated evaluations, in general, identify lower programme impact. It is also clear that the more sophisticated evaluations are also more expensive, primarily because more items are held constant rather than attributed to the programme. This could easily lead to a conflict of interest since, if those responsible for the programme are also responsible for setting the budget for the evaluation, there may be pressure for favoured programmes to be evaluated using less sophisticated procedures. A priori, it would seem that this type of discretion is undesirable, and could suggest a separation of responsibilities between those Evaluators and Programme Managers.

(ii) *The frequency and scale of the evaluation, related to the scale of the programme*

The experience in OECD countries is that some programmes are evaluated, sometimes more than once, and others are never evaluated. As noted above, it is also the case that some programmes are evaluated using considerably more sophisticated techniques than others.

In part this understandably reflects governments choosing to link evaluation budgets to programme scale. The problem, however, is that this makes comparisons difficult between the results of small and large scale evaluations. Yet, if SME policy is to satisfy the OECD criteria of *Superiority* and *Systemic Efficiency*, and to the concept of *coherence* then the marginal efficiency of all policies should be identical, but this information will not be available if policies are evaluated using different methods.

The pragmatic response may be to, as a minimum, agree that all programmes should be evaluated, so that none slip through the net. It also seems appropriate to link the scale of the evaluation budget to the scale of the programme. Finally, to facilitate the broadest comparisons, it might be suggested that no programme be evaluated using less than Step IV, and that the larger the programme the greater the expected use of Step V and Step VI approaches.

(iii) *Evaluation by internal or by external organisations?*

A judgement has to be reached about whether programmes should be evaluated by the departments responsible for delivering them, by other organs of government or by external organisations. There are arguments favouring evaluators from all of these groups.

The Departments responsible for delivering the programme clearly have unique inside knowledge and understanding, which will not be available to any outsiders. They may also be more likely to engage the support of those delivering the programme. On the other hand there is the risk that their "independence" may be impaired through this closeness, and risk being subject to "capture" or political influence.

An alternative is for such evaluations to be conducted by a specialist part of government, independent of the programme delivery department. The advantage here is that they will have greater specialist evaluation skills, and be seen to be more independent than the department delivering the programme. Their detailed programme specialism will however be less and so they will rely more heavily upon the co-operation of the programme delivers.

A third option is to engage specialist "outsiders" such as consultants or academics. Both these groups have the advantage of being less clearly subject to "capture", and being likely to be specialists within their subject¹⁵. The central disadvantage of the employment of outsiders is that if Evaluation is

¹⁵ A case could be made that consultancy practices are not wholly independent, since their business depends upon repeat contracts from governments. It is appropriate here to repeat the point made by Kuusisto et al. (1999), that,

viewed as one element in the *process* of policy improvement then this has to lead, as in Figure 2, to Discussion. That Discussion 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.

A judgement therefore is needed which reviews these conflicting arguments and reaches the important decision on who is to conduct evaluations.

(iv) *Implementing the findings of the evaluation*

Since evaluation is not “the end of the line”, experience of where evaluation has led to policy modifications is critical. These modifications may be either changes in the objectives of the programme or to changes in the way the programme are implemented or delivered. For example programmes may be extended to include businesses in sectors that were previously excluded; changes in the terms and conditions of the programme or changes in the desired outcomes.

The process by which any programme changes are implemented is likely to vary depending on the scale and profile of the programme, and according to political processes. Nevertheless the key purpose of conducting evaluations is for them to be an influence upon policy. In some countries, particularly for high profile programmes, it may be valuable for information stemming from the evaluation to enter the public domain. This is more likely to occur where “outsiders” have conducted the evaluation.

A number of options can be considered. One of these could be that all programmes have only a fixed ‘life’, and that continuation depends upon a satisfactory evaluation. This could be considered as a variation on ‘sunsetting’ legislation. It has the advantage of ensuring that all participants agree, when the programme is first developed, its objectives and how it will be assessed. It would also enable programmes to be reoriented to reflect changing economic circumstances.

(v) *Significance for middle and lower income countries.*

Whilst it is important for taxpayers in all countries to be satisfied that government programmes are effective, it is also important to recognise that resources for evaluation are likely to be very scarce in middle and lower income countries. In these situations a key role can be played by donor organisations who need to establish an audit trail for the funds, but who also can take a lead in highlighting the benefits of undertaking appropriate evaluation. One example of this link is the work of the Inter American Development Bank (IBD) and FUNDES International. They sought to develop instruments capable of designing, monitoring and evaluating SME policy instruments in Latin America, most notably Mexico, Chile and Columbia. A summary of the key findings of this project is found in Appendix 2.

A similar role could also be played by international organisations such as OECD in providing case studies of evaluations appropriate for middle and lower income countries. Scholarships and opportunities to study abroad might also be provided for early career government economists to become familiar with research developments.

on balance official external evaluations have been more favourable than those evaluations that were undertaken independently. However it is also appropriate to point out that consultants have a reputation for jobs being completed on time, with this not always being a characteristic of academics!

CONCLUSIONS

This paper has provided an extensive review of the methods used to evaluate elements of SME policy. It draws upon experience primarily from OECD member states but, where appropriate, has also made reference to non-member states. The key conclusions are now presented.

First, that evaluation cannot take place adequately until the objectives and targets of SME policy are clearly specified. The paper suggests using the COTE framework for policy formulation. In this, the letter “C” stands for both policy clarity and policy coherence. ‘Clarity’, in this context, means that the policy is fully understood by those delivering it and those expected to benefit from it. ‘Coherence’, however, is a slightly more sophisticated concept. It refers to the need to ensure that all parts of government seek to collectively work in the interests of SMEs. Specifically SME policies are unlikely to be ‘coherent’ where they are delivered by many different government departments without co-ordination or where some departments focus upon regulation of SMEs and others focus upon ‘support’, without an adequate dialogue between them.

The letter “O” stands for ‘Objectives’ of policy. To ensure clarity the objectives of policy must be clearly stated; in practice such objectives might be to increase the number of firms, or to increase the formation rate of firms, or to increase the value added of firms. It might also include seeking to raise business ownership amongst groups defined in terms of ethnicity, gender, disadvantage or location.

The letter “T” stands for ‘Targets’ which need to be specified in a manner that enables an assessment to take place of whether or not they are achieved. An example of converting an ‘objective’ into a ‘target’ would be the objective of ‘to increase the formation rate of firms’. Here the target might be ‘to raise new firm formation by 10% over the next three years’. Only then is it possible to be clear whether or not the target is met.

The letter “E” stands for Evaluation, which is the prime focus of this paper. However the first key conclusion of the paper is that evaluation cannot be undertaken unless targets exist. This is because evaluation can only take place in a framework in which the expected policy impacts are clearly specified.

A second key conclusion of the paper is that Evaluation needs to become more central to the policy-making process. It should not be undertaken solely as a historic accounting exercise to determine whether public money has been spent wisely, although that role is of value. Instead of being “at the end of the line” evaluation should be used to inform current policy, so that the objectives and targets may be modified in the light of evidence of policy effectiveness. Hence considerations of how policy is evaluated should therefore be incorporated into policy formulation when new ideas are being developed.

Ensuring that evaluation is given a higher priority means that all, rather than merely some, programmes are assessed. The paper is however less clear about the extent to which all programmes should be assessed with the same level of sophistication. The key argument in favour of a consistent level of sophistication is that this enables valid comparisons to be undertaken between one programme and another. The argument against is that it is inappropriate to have similar evaluation budgets for programmes of widely differing scales.

Nevertheless the paper makes it clear that it is now technically possible to provide accurate measures of programme impact. It shows, through its “Six Steps to Heaven” approach that, although there are examples of the most sophisticated approach being used, these are the exceptions rather than the rule. The cause for concern is that the less sophisticated approaches tend to provide misleading answers, most frequently over-estimating policy impact. The third key conclusion of the paper is that, where appropriate, the most sophisticated – Step 5 or Step 6 – approaches are to be used.

In undertaking these evaluations it is necessary to bear in mind some potential conflicts. The first is that ideally those undertaking the evaluations need to be independent of those responsible for the programmes, as in any audit role. But, 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. This may be more difficult to achieve if the evaluator is viewed as an “outsider”. On balance, the paper concludes that the independence of the evaluator is of the greater importance, but this is a close call.

Finally the paper concludes that OECD itself could play a greater role in disseminating “good practice” in SME policy evaluation techniques.

ANNEX 1

Table A.1. United Kingdom Support for Small Businesses, 2001

(in millions of Pounds Sterling)

Expenditure		5 342
<i>of which</i>		
Small Business Service Budget	349	
Other expenditure by Department of Trade and Industry and its agencies	275	
British Trade and Industry	44	
Department of Culture, Media and Sport and its agencies	332	
Department of Education and Skills and agencies	138	
Regional Development Agencies	274	
Department of Work and Pensions	71	
Local authorities	300	
Tax agencies (support services)	81	
Department of Environment, Food, and Rural Affairs (grants, advice in-kind support)	3 120	
Department of Transport, Local Government and Regions and agencies	107	
EC and EIB	246	
Favourable tax treatment (R&D Tax Credit; 10; Corporation Tax; 20p Small Company Corporation Tax Rate; Venture Capital Trusts; Enterprise Investment Scheme)		2 590
Total		7 932
(as a percent of GDP)		(0.8)

Source: "Cross Cutting Review of Government Services for Small Businesses", Department of Trade and Industry, September 2002, URN 02/1324.

ANNEX 2

GUIDE TO THE EVALUATION OF POLICIES TO SUPPORT SMALL BUSINESS⁺,*Maria Vega, Genente Internacional, Entorno Empresarial, Fundes International***Executive Summary**

The Inter-American Development Bank (IBD) and FUNDES International, in the framework of technical cooperation, embarked on the development of an instrument capable of helping those responsible for the design, evaluation and monitoring of support actions. Thus the methodology set out in this Guide will be useful to anyone undertaking a comprehensive evaluation of micro/SME support policy (based on a common set of indicators). It is also relevant to the evaluation of specific policies (by area) at national, regional and/or local level. Using quantitative and qualitative results indicators, it is possible to obtain a good idea of a policy in a given context, compared with the best possible practice.

The guide is divided into five main sections: (i) the objective, definitions and content of a small business policy; (ii) the methodology as such – criteria, scope and sources of information; (iii) the general policy and its impact on micro/SME; (iv) evaluation by area; and (v) evaluation of the support system as a whole.

For the purposes of this guide, we consider that the ultimate goal of a small business support policy is to strengthen the sector in terms of more businesses and better competition between them. Given this goal, small business policy consists of three components: (i) a group of strategic areas where the actions have a direct impact on micro/SME; (ii) the degree of coordination, interrelation and cooperation between them and the various institutions involved; and (iii) what we term general policies which, although not designed to affect the behaviour of micro/SME, do in practice influence their behaviour.

In this connection, given the obvious gaps in the area of evaluation, the methodology has a “bottom-up” structure. This means starting from a general analysis and evaluation of the components of the support policy, *i.e.* the programmes, the areas to which they relate, the people and organizations involved, their internal cohesion, the degree of coordination between them, and finally their overall functioning. It also briefly deals with other public sector actions (general measures and their impact on micro/SME). In terms of the areas concerned, it is assumed that the chief challenges are financing, technology and information (including business development centres), marketing and foreign trade, the legal, regulatory and competitive framework, human resources, education and training, and cooperation among micro/SME.

The Guide considers three evaluation criteria: **final** (which reflect the economic wellbeing of society as a whole), **intermediate** (rather more straightforward and angled at the impact on the performance of micro/SME) and **institutional** (which measure the quality of institutions involved in the provision of support measures). Thus the criteria range from the economy in general, the beneficiary of a given service and the institution providing the service. The following is an example of the criteria in the area of technology.

⁺ Both here and in the Guide itself, the term micro/SME or small business is used synonymously to define the group of medium-sized, small and micro enterprises with development potential.

TABLE A. 2. CRITERIA FOR THE ANALYSIS OF AREAS

Area Criteria	<i>TECHNOLOGY</i>
INTERMEDIATE CRITERIA	Did the service (programme or other type of support) reach a satisfactory number (compared with the target) of small businesses? For example, what proportion of small businesses examined visited a technological development centre?
	Did the service meet the needs for which it was designed? For example, what proportion of the businesses that visited the centre found the visit useful?
	Did the service or support have the desired positive impact on growth, productivity, employment, profits or other goals of the business (or group of businesses)? For example, is there any evidence that businesses which use the technology development centres increase their productivity more quickly than other comparable small businesses?
FINAL CRITERIA	Did the package of measures (or the overall support system) have the expected impact on the general growth of the product, productivity and employment in small businesses and incomes distribution in the country? For example, can it be shown statistically that the presence of business development centres makes a difference in terms of growth and job creation in the small business sector?
INSTITUTIONAL CRITERIA	The government's commitment to the design of the system of technological support to micro/SME and its effectiveness, which in practice means availability of economic resources, qualified human resources, clear objectives and targets and evaluation of the results.
	The support of the non-governmental sector (intermediate agencies and private sector) in the form of cooperation in the design, execution and/or monitoring of public sector support activities, through co-financing, direct provision of technical assistance services, etc.
	The quality of the institutions which participate in the provision of services, reflected in the professional qualities and experience of the members of, for example, national science and technology councils, the degree and form of decentralization of activities, etc.
	Wise choice of instruments: are the technologies suited to the specific needs of the micro/SME sector of the country or region? And is there a proper balance between them?

Depending on the areas to be covered and the decision criteria, appropriate information can be collected on the performance of policies and programmes, taking advantage of existing quantitative and qualitative information (*e.g.* such as that compiled by support institutions) and developing new sources as needed. The Guide includes a questionnaire to assist in obtaining information from both clients and support services, and experts on the subject. The following table is useful in planning the types of indicators that should be collected by area.

TABLE A. 3. INDICATORS TO BE COLLECTED BY AREA

Area Variables	<i>Technology</i>
Objective(s) of the area	<ul style="list-style-type: none"> • One or more • Concrete or vague (deliverable?) • Complementary or conflicting
Institution(s) responsible	<ul style="list-style-type: none"> • One or more • If more than one, do they coordinate with each other?
Description of the area	<ul style="list-style-type: none"> • Brief description of the policies, programmes and/or actions involved (are they redundant, complementary, etc?). The human and economic resources. Degree of coordination with private enterprise, local government, etc.
Programme dissemination mechanisms	<ul style="list-style-type: none"> • Print (newspapers, magazines) • Audiovisual (radio, TV, Internet) • Leaflets, flyers, documents of institutions, etc. • Business chambers, local guilds, universities, technical schools • Other
Operational mechanisms	<ul style="list-style-type: none"> • Public • Private • Mixed
Evaluation mechanisms	<ul style="list-style-type: none"> • Evaluation of areas / specific programmes • Variables involved (coverage, satisfaction, productivity gains, cost savings, etc.) • Monitoring after a given period of time
Cost by area	<ul style="list-style-type: none"> • Total cost of area actions (in local currency and US\$) • Costs by programme (in local currency and US\$) • Human resources and % of costs that they represent in area actions – can be broken down by programme • Are the programmes co-financed? (In what proportion and by whom?)
Degree of coordination with other areas	<ul style="list-style-type: none"> • Is there coordination with other areas as an integral support policy? • Coordination with the private sector / international organizations in: <ul style="list-style-type: none"> ○ Subcontracting ○ Consultancy ○ Joint initiatives ○ Other
Beneficiaries (eligibility criteria)	<p>In theory:</p> <ul style="list-style-type: none"> • All types of enterprise • Only SME • Micro • Other <p>In practice, who are the beneficiaries?</p>
Quality certification	<p>Is there any type of quality certification in the area programmes:</p> <ul style="list-style-type: none"> • At local level • A international level

This information, supplemented by a brief description of the micro/SME in the region and the state of general policies is the basis of the evaluation itself and the production of the report. The evaluation may be carried out at different levels of sophistication, depending on the needs of the users, the information and economic and human resources available to undertake it.

Finally, the Guide for the evaluation of support policy is not a static instrument. It must be constantly updated as the support system adjusts to new realities and business needs. It must therefore be a continuous process of adaptation and improvement both in its general structure and the specific areas of evaluation and the instruments for each. As the available information improves, both overall and for specific actions, together with the level of qualifications of the human resources, availability of economic resources, etc, the degree of refinement of the methodology will be constantly enhanced.

REFERENCES

- Angrist, Joshua D. and Krueger Alan B. (2001), “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments”, *Journal of Economic Perspectives*, 15(4), pp. 69-85.
- APEC (2002), “Evaluation Framework for Economic and Technical co-operation work across APEC Fora”, Paper submitted by Australia for the ESC Small Group on Evaluation.
- Bondonio, Daniele and John Engberg, (2000) “Enterprise Zones and Local Employment: Evidence from the States’ Programs”, *Regional Science and Urban Economics* 30: 519-549.
- Bager-Sjogren, L. (2003), “On the Job Training and Small Business competitiveness – EU Structural funds Impact in Sweden”, Swedish Institute of Growth Policy Studies (ITPS), presented at ICSB Conference, Belfast 16-18 June 2003.
- Bartik, T., (2002), “Evaluating the Impacts of Local Economic Development Policies on Local Economic Outcomes: What has been done and what is doable?”, paper presented at ‘Evaluating Local Economic and Employment Development Conference, OECD, Vienna, 20-21 November.
- Fraser, S., Storey, D.J. and Westhead, P. (2003), “Small firm work placements for Undergraduates: Does it pay off?”, University of Warwick, SME Centre.
- Greenbaum, Robert T. (1998), “An Evaluation of State Enterprise Zone Policies: Measuring the Impact on Business Decisions and Housing Market Outcomes”, *Doctoral dissertation, Carnegie Mellon University*, December.
- Harrison, R.T. and Leitch, C.M. (1994), “Whatever you hit call the target: an alternative approach to small business policy” in Danson, M. (ed.) *Small Firm Formation and Regional Economic Development*, Routledge Series in Small Business, London.
- Heckman, James J., Hidehiko Ichimura, and Petra Todd, (1997), “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *Review of Economic Studies*, 64(4), pp. 605-654.
- Holzer, H.J., Block R.N., Cheatham M. and Knott J.H., (1993), “Are Training Subsidies For Firms Effective? The Michigan Experience”, *Industrial and Labor Relations Review*, 46, (July), pp. 625-636.
- Jarmin, Ronald S. (1999) “Evaluating the Impact of Manufacturing Extension on Productivity Growth.” *Journal of Policy Analysis and Management*, 18(1), pp. 99-119.
- Kuusisto, J., Berney, R. and Blackburn, R (1999) “A Critical Evaluation of SME Support Policies in the United Kingdom and the Republic of Ireland – An in depth Delphi Study of Selected SME Support Policies and their Evaluation”, Ministry of Trade and Industry, Finland, Studies and Reports 6/1999.

- Lerner, J. (1999), "The Government as Venture Capitalist: The Long Run Impact of the SBIR Program", *Journal of Business*, Vol. 72, No. 3, pp. 285-318.
- Meyer Bruce D. (1995), "Natural and Quasi-Experiments in Economics", *Journal of Business and Economic Statistics*, 13(2), pp. 151-161.
- Murnane, R.J., S. Newstead, and R.J. Olsen, (1985), "Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias", *Journal of Business and Economic Statistics*, 3, pp. 23-35.
- Oldsman, E., (2002), "Evaluating Business Assistance Programme", paper presented at 'Evaluating Local Economic and Employment Development conference', OECD, Vienna, 20-21 November.
- OECD (1997), *Policy Evaluation in Innovation and Technology: Towards best Practices*, OECD, Paris.
- OECD (2000), *Small and Medium Enterprise Outlook*, OECD, Paris.
- OECD (2003), *Policy coherence: vital for Global development*, Policy Brief, July, OECD, Paris.
- Papaconstantinou, G. and Polt, W. (1997), "Policy Evaluation in Innovation and Technology: An Overview", in *Policy Evaluation in Innovation and Technology: Towards Best Practices*, OECD, Paris.
- Roper S. and Hewitt-Dundas, N. (2001), "Grant Assistance and Small Firm Development in Northern Ireland and the Republic of Ireland", *Scottish Journal of Political Economy*, Vol. 48 No. 1 pp. 99-117
- Smith, Jeffrey A. and Todd Petra E., (2001), "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods." *The American Economic Review*, Vol. 91(2), pp. 112-118.
- Storey, D.J. (2000), "Six Steps to Heaven: evaluating the impact of public policies to support small businesses in developed economies", in Landstrom H and Sexton D.L. (eds.) *Handbook of Entrepreneurship*, Blackwells, Oxford, pp. 176-194.
- Thomas, T. and Landry B. (2001), "Evaluating Policy Outcomes- Federal Economic Development Programs in Atlantic Canada", ACOA, Moncton, New Brunswick.