

**Methodologies used in the Evaluation of
the Effectiveness of European Structural
Funds: A Comparative Assessment**

Final Report to the Scottish Executive

*Professor John Bachtler, Laura Polverari and Sandra Taylor
~ European Policies Research Centre ~
Professor Brian Ashcroft and Professor Kim Swales
~ Fraser of Allander Institute ~*

European Policies Research Centre and Fraser of Allander Institute
University of Strathclyde
40 George Street
Glasgow G1 1QE

Tel: 0141 548 3339
Fax: 0141 548 4898
E-mail: j.f.bachtler@strath.ac.uk

September 2000

Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment

Final Report

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
1. Introduction	9
2. Objectives and structure.....	11
2.1 Objectives	11
2.2 Methodology.....	11
3. Structural Fund evaluation in Scotland: review and critique.....	13
3.1 Context	13
3.2 Management	15
3.3 Programme-level organisation of evaluation	20
3.4 Methodological process	25
3.5 Targets, indicators and data	32
3.6 Methods and techniques	34
3.7 Overall assessment of Scottish evaluation practice	36
4. International Review of Structural Fund evaluation: context	43
4.1 Diverse attitudes towards evaluation	43
4.2 The influence of EC evaluation obligations.....	46
4.3 Evolution of Structural Fund evaluation in the EU.....	48
4.4 Assessment	54
5. International review and case studies: management and organisation of evaluation	55
5.1 Evaluation management.....	55
5.2 Integrated programmes of evaluation	62
5.3 Promoting partner involvement	71
5.4 Strengthening the capacity of the evaluation community	74
5.5 Improving the quality of data	85
5.6 Following up evaluation studies: validation, dissemination and exploitation.....	89
5.7 Assessment	93
6. International review and case studies: methods of evaluation.....	101
6.1 Evaluation issues	101
6.2 European comparative overview.....	107

6.3	Assessment	129
7.	Conclusions and recommendations	132
	Annex 1: Research specifications.....	141
	Annex 2: Evaluation documentation analysed for the Scottish review	146
	Annex 3: List of interviewees – Scottish review	149
	Annex 4: List of interviewees – international review.....	151
	Annex 5: Glossary	155

INDEX OF TABLES

Table 3.1: European Structural Fund Allocations and Commitments for Scotland at 31 December 1999 (£ million)	14
Table 3.2: Scottish Structural Fund Evaluation Studies – interim evaluations- June 2000	26
Table 3.3: Scottish Structural Fund Evaluation Studies – ex post evaluations- June 2000	29
Table 4.1: Evaluation traditions and experience in EU Member States.....	44
Table 4.2: Commitment to Structural Fund evaluation in EU Member States	52
Table 5.1: Relative merits of single and multi-programme evaluations	63
Table 5.2: Use of thematic evaluations among EU Member States.....	67
Table 5.3: Partnership involvement in evaluation.....	72
Table 5.4: Relative merits of using external and internal evaluators	74
Table 5.5: Capacity of the evaluation community	76
Table 5.6: Monitoring systems in EU Member States	86

INDEX OF CASE STUDIES

1. Institutional Arrangements for Managing Evaluation - Ireland.....	57
2. Institutional Arrangements for Managing Evaluation - Italy.....	58
3. A 'Nested' Approach to Monitoring and Evaluation - West Midlands.....	60
4. Meso-Level Programme Evaluation - Yorkshire and Humberside.....	64
5. Use of Interim and Thematic Evaluations in a Diagnostic Process - North-Rhine Westphalia.....	68
6. Evaluators as Process Consultants - Tuscany.....	70
7. Partnership Involvement in Evaluation - Finland.....	73
8. Cultivating a Relationship with an Evaluation Team - North-Rhine Westphalia.....	78
9. Documentary Sources on Evaluation Practice - The MEANS Programme and EC Resources.....	79
10. Checkpoint EVA - Austria.....	81
11. Lesson-learning from Evaluation - France.....	82
12. Lesson-learning from Meta-Evaluation - Sweden.....	91
13. Analysis of Assessability - Spain.....	109
14. Use of Control Groups to Assess the Impact of ESF Funding - UK.....	113
15. Use of Interviews and Evaluator's Judgement to Estimate Additionality and Displacement - Gibraltar.....	115
16. Measuring Net Effects of Technology and Innovation Support in the Evaluation of RTDI-related Actions - Austria.....	116
17. Identification of Net Employment Effects for a Whole Programme - Netherlands.....	118
18. Dynamic Input-Output Analysis - Beutel Model.....	120
19. Econometric Technique - The PARADISE Model.....	122
20. The Hermin Model.....	125
21. The CSF <i>ex ante</i> evaluation for the Mezzogiorno region of Italy.....	128

ABBREVIATIONS

CSF	Community Support Framework
DATAR	Délégation à l'Aménagement du Territoire et l'Action Régionale (Agency for Spatial Planning and Regional Development)
DETR	Department of the Environment, Transport and the Regions
DfEE	Department for Education and Employment
DoE	Department of the Environment
DTI	Department of Trade and Industry
EC/CEC	European Commission
ERDF	European Regional Development Fund
ESF	European Social Fund
EU	European Union
ICT	Information and Communication Technologies
MEANS	Research Programme on 'Methods for Evaluating Structural Policies'
MC	Monitoring Committee
ÖROK	Österreichisches Raumordnungskonferenz (Austrian Spatial Planning Conference)
PMC	Programme Management Committee
PME	Programme Management Executive
RTDI	Research, technological development and innovation
SE	Scottish Executive
SEP	Strathclyde European Partnership
SO	Scottish Office
SPD	Single Programming Document
SQW	Segal Quince Wicksteed

EXECUTIVE SUMMARY

Introduction

This document presents the final report of a study on the methodologies used in the evaluation of Structural Fund programmes across Member States. The research has been undertaken by the European Policies Research Centre and the Fraser of Allander Institute, both of the University of Strathclyde, on behalf of the Scottish Executive.

The study had five key objectives:

- (i) to analyse the Scottish evaluation framework in the context of the new round of Structural Fund programming;
- (ii) to examine how the evaluation of Structural Fund co-financed programmes is organised and conducted elsewhere in the EU;
- (iii) to review the spectrum of evaluation methods used across the EU for assessing the effectiveness of Structural Fund expenditure;
- (iv) to analyse the strengths and weaknesses of evaluation methods used in other parts of the EU; and
- (v) to provide clear recommendations for improvements to the monitoring and evaluation framework in Scotland.

It was commissioned on 4 April 2000 and completed on the 28 July 2000. The tight timetable is intended to allow results to be rapidly fed into the evolving monitoring and evaluation framework of the new generation of Structural Fund programmes in Scotland. This study has been designed to contribute to the overall objective of the new Scottish Executive - confirmed in the recent review of the PMEs - to measure the impact of the new round of Structural Funds so that they leave a real lasting legacy for the future. It will inform the development of an Evaluation Strategy for the Scottish Structural Fund programmes.

Within the short timespan, the study has combined an extensive review of evaluation guidelines, Structural Funds evaluations and policy literature from the UK and throughout the EU with survey fieldwork, including face-to-face interviews with Scottish evaluation practitioners and policy-makers and telephone interviews with a range of evaluation policy experts in the EU Member States.

In addition to setting out the report structure and objectives, the report consisted of five principal sections: a review of existing evaluation practice in Scotland; a review of the background to evaluation in other Member States; an assessment of the organisational structures for conducting evaluations in other Member States; a critique of evaluation methodologies from throughout the EU; and lastly, a series of policy recommendations for the Scottish situation. The executive summary covers each of these areas in turn.

Structural Fund evaluation in Scotland: review and critique

Far from merely satisfying the letter of EC guidelines - and in contrast to some Member States - the conduct and management of evaluation of the Structural

Funds has been taken seriously in Scotland. Within a common UK framework, a systematic approach to programme evaluation has been in operation for over five years. The investment in monitoring systems begun in the mid-1990s has already made a significant difference to comparative assessment of programme progress among the Scottish programmes, and overall, evaluation has been used to assess the progress of programme implementation and make corrective actions in the course of existing programmes and the preparation of new ones. While there are areas where improvements in the management of evaluation will continue to increase the value of evaluations, Scotland - and the UK more widely - has been recognised as being an exemplar in the organisation of Structural Fund evaluation.

Nevertheless, with respect to understanding the actual effects of Structural Fund expenditure, hitherto, there has been a lack of effective assessments of the impact of Structural Fund programmes. Scottish evaluations have been frequently characterised by an absence of methodological rigour in determining quantitative impacts and weaknesses in addressing the qualitative effects of programmes. Few thematic evaluations have been undertaken, and Structural Fund evaluation in Scotland has not been effectively integrated with the evaluation of domestic policy measures. Evaluation has also been hindered by deficiencies in the generation of adequate monitoring data. Moreover, while there is a strong evaluation capacity in Scotland, it continues to be based on a limited pool of potential evaluators and a restricted range of evaluation skills, particularly in relation to assessing ESF impacts.

International review of Structural Fund evaluation: context

Evaluation experience varies greatly across the EU. In some Member States, evaluation is deeply ingrained within the administrative culture, in others, it is only rooted within particular departments or policy areas and operates on an *ad hoc* basis, and in a few cases, it is viewed as having little policy relevance. The UK is generally regarded as the best example of a well-developed evaluation culture, with policymakers placing great emphasis on setting objectives, establishing monitoring systems and managing evaluations in regional policy over a 20-year period. Similar commitments to domestic policy evaluation can be seen in Germany, the Netherlands and Sweden.

Across the EU as a whole though, the evaluation needs of the Structural Funds have galvanised national evaluation practice, because of both the co-financing of programmes and the sheer volume and complexity of the evaluations demanded by EU guidelines. Structural Fund requirements have acted as a catalyst on differing national evaluation traditions, leading to significant variation in how Structural Fund evaluations have been organised (as well as in compliance at programme level). It has also created a number of useful lessons for Scottish practice, both in terms of the organisation of evaluations and the methodologies they have employed.

International review and case studies: management and organisation of evaluation

In many respects, Scotland has a strong foundation in the organisation of Structural Fund evaluations and has been viewed as a model for other countries. Nonetheless, several Member States offer useful approaches which can address the remaining weaknesses in the Scottish system. The development of a single framework for Structural Fund evaluation has been successfully undertaken in Ireland and Italy, where all-encompassing integrated frameworks for evaluating the range of programmes have been introduced on a cross-governmental basis. While no EU country has adequately combined Structural Fund and domestic policy evaluations, some Member States have established notable systems for measuring both Structural Fund and national policy activities within the same monitoring system (eg. France). Other areas have advanced the integration of Structural Fund evaluation by conducting a series of meso-level evaluations of programme elements - rather than a single global evaluation of a whole programme - allowing for more sophistication in the use of methodologies and thematic studies to gain insights into complex areas such as RTDI or community economic development. Further, the organisation of Structural Fund evaluation has been enhanced in some Member States by novel institutional arrangements promoting permanent and more constructive engagement of partners, as in Finland.

While all Member States have experienced problems in developing a capacity for Structural Fund evaluation, in several Member States, effort has been made to institutionalise evaluation experience (as in Austria, where a centralised resource centre for evaluation has been created in 'Checkpoint Eva'). Instructive documentary sources on evaluation are available for dissemination through the EU's MEANS programme, while a few Member States have developed their own approaches to quickly learning the lessons of evaluations, notably through the use of meta-evaluations to draw out good practice, as seen, for example, in Sweden. Lastly, while adequate monitoring systems has eluded all Member States, the range of experimentation across the Community provides a source of innovation for Scottish practice.

International review and case studies: methods of evaluation

No Member State has successfully addressed the problem of how to determine the actual impacts of the Structural Funds on an area's economy. Indeed, some Member States are not concerned with the issue, and view evaluation strictly in terms of satisfying EU guidelines. In this context Structural Fund evaluation in Scotland has been taken seriously and not as a mere fulfilment to imposed requirements. However, as far as gross impacts are concerned, the identification and measurement of different types of gross effects is one area where Scottish practice can learn from EU experience. On the other hand, as far as net additional effects are concerned, Structural Fund evaluations in Scotland and in the rest of the UK are more likely to identify key programme impacts – particularly additionality, displacement and multiplier effects - than in other EU countries. Nevertheless, over the past decade the EU has witnessed significant (if generally piecemeal) improvements in the potential

measurability of programme outputs and the overall growth of an evaluation culture in different Member States. Examples of useful methodological practice can be found in other countries in the two main approaches to economic policy evaluation: bottom-up and top-down.

Different Member States have strengths in different aspects of a bottom-up approach. Tackling the problem of the ‘assessability’ of programme impacts has been incorporated into the prior appraisal of programmes with success in Spain. Good examples of how to identify, measure and aggregate different types of gross effects can be found in Germany, where evaluation programmes give impacts at a detailed sectoral and company-size level. There are also some German and UK evaluation studies that identify the time path of employment change.

No Member State has successfully developed a comprehensive approach to measuring net impacts. However, a range of alternative methods provides new perspectives on the problem. In the UK there is extensive use of benchmarks to quantify additionality, displacement and multiplier effects. If used appropriately, these can be a very cost-effective but benchmarks have a very circumscribed role in quantifying additionality. Further, more work could be done in determining multiplier effects, especially in Scotland which is particularly well served with national and some regional Input-Output tables.

The use of control groups, which has tended not to feature in Scottish evaluations, has been successfully operated in Ireland and in the UK for assessing the impact of ESF funding. The use of survey techniques, which has been used in Scottish evaluations in an attempt to measure additionality, could be improved by following examples of good practice adopted in some French, German, Dutch and British evaluations. An Austrian evaluation also represents a good example of the application of the survey analysis for assessing the net impact of the RTDI projects implemented in the Objective 2 regions. However, this all having been said, the identification of the overall net impact of Structure Fund policies through the use of bottom-up approaches is generally done less well in the rest of Europe than in Scotland and the rest of the UK.

Good examples of a top-down approach in other countries also exist, notably sophisticated models such as PARADISE and HERMIN. Top-down and bottom-up methods are in some respects complementary and at the least are alternative takes on the same problem. Given the difficulty in quantifying policy impacts there is scope for using both methods. Within a Scottish context, it is likely that top-down methods are more appropriate for Scottish-wide, rather than Partnership-specific, evaluations of the Structure Funds. This is mainly due to data availability and the ability to spread the fixed costs of developing a model across a number of economic evaluation exercises (for both Structural Funds and domestic policies).

Policy recommendations

By combining a critique of the current Scottish approach with a review of international good practice, the report has produced a series of policy recommendations. These are presented under the following headings: approach; partnership; capacity; methods; and use of results.¹

Approach

1. A common framework for the integration of Structural Fund and domestic policies should be developed, including the establishment of a dedicated Evaluation Unit in the Scottish Executive and the development of a 'nested' approach to monitoring and evaluation.
2. Integration of Structural Fund evaluations should be improved by making more explicit the link between different phases of evaluation, adapting a more combined programme and thematic approach to evaluation and using more cross-programme comparative research.
3. Process issues could be addressed more roundly by involving evaluators on an on-going basis and establishing a continuing dialogue and trust between evaluators and Programme managers.

Partnership

4. At the programme level, the PME should continue to use Evaluation Steering Groups which already involve partners, but these could meet on an on-going basis, in parallel with the implementation of the programmes.
5. At the multi-programme level, and over the longer term, a more permanent dialogue on evaluation issues could be developed, perhaps through a permanent sub-group of the proposed Scottish Structural Funds Forum or the Scottish Co-ordination Team.

Capacity

6. The Scottish Executive and programme managers should invest in developing capacities for evaluation and should investigate the possibility of developing a publicly accessible resource library on evaluation, and a point of information (web-site) for practitioners and partners on evaluation.
7. The Scottish Executive and programme managers should promote exchange of experience and lesson-drawing, before and after evaluation exercises.
8. To ensure broader knowledge on the available evaluators, an open register of evaluators should be developed.
9. To encourage consultants and evaluators to undertake Structural Fund evaluations, a plan of activity on future evaluations should be publicised in advance by the PMEs and the Scottish Executive on a regular basis.

Methods

¹ The conclusions and recommendations of the report represent the point of view of the EPRC/FAI research team and are not the official position of the Scottish Executive. The usual disclaimers apply.

10. The Scottish Executive should take steps to re-organise the monitoring framework and set of indicators into a coherent and integrated framework for Structural Funds and, ideally, also non-co-financed domestic policies.
11. The work undertaken by the PMEs to increase the monitoring skills of implementers could be consolidated by dedicating increased resources to monitoring visits and other developmental activities.
12. The commitment of more resources is required if more accurate and programme-specific measures of the effectiveness of Structure Fund expenditure is to be achieved.
13. In the bottom-up approach, an accurate estimate of additionality and displacement requires direct contact with the recipients of Structural Funds assistance.
14. Where multiplier values are used to identify indirect effects, great care should be taken in the choice of multiplier. This is because multipliers vary across sectors, areas and the size and type of assisted activity.
15. Consideration should be given to the more intensive use of existing Input-Output tables for Scotland and sub-regions within Scotland for determining multiplier impacts.
16. More attention should be paid to the estimation of disaggregated impacts, embracing industrial sectors, labour market impacts such as gender and environmental impacts that have implications for sustainability.
17. Industrial surveys typically associated with the bottom-up approach should be linked to an evaluation model with a more active supply side of the Hermin type. Such a model is available for Scotland – the AMOS computable general equilibrium model - and this model has been used for policy evaluation.
18. Top-down and bottom-up approaches should be treated as complementary and not competitive. Top-down and bottom-up methods are not mutually exclusive. However, top-down methods are more practical at the economy-wide (Scottish) level. The modelling of Structural Fund impacts at the national level might be more appropriately dealt with in a wider programme of economic modelling.
19. There should be more formal and systematic integration of Scottish Enterprise and Highlands & Island Enterprise evaluations into Structural Fund evaluations at the Programme Area level.
20. The encouragement of the use of statistical and econometric methods to identify policy impacts both at the Programme Area and Scottish levels. Approaches similar to, but ideally more sophisticated than, the PARADISE model.

Use of results

21. An Evaluation Report series should be set up by the Scottish Executive, containing all Structural Fund evaluation studies as well as relevant reports on methodological issues.

22. At the programme level, each evaluation should contain a 'communication plan' to ensure appropriate dissemination to different target audiences.
23. The Scottish Executive could undertake or commission, after each round of evaluation, meta- evaluations, to identify the strengths and weaknesses of the evaluation reports, the degree to which they can be and/or have been exploited and new themes for future evaluation exercises.
24. The outcome of the suggested meta-evaluations could be discussed by civil servants, programme managers, partners, evaluators and other interested parties to encourage learning and lesson-drawing among the wider community.

1. INTRODUCTION

This document comprises the Final Report of a study on the development of Structural Fund evaluation in Scotland and possible lessons from other parts of the EU. The study has been undertaken on behalf of the Scottish Executive by the European Policies Research Centre and Fraser of Allander Institute, both of the University of Strathclyde.

The main aim of the study is to contribute to the development of the Scottish Executive's framework for the evaluation of the Structural Funds by drawing on experience elsewhere in the EU. The study was commissioned on 4 April 2000 and completed on 28 July 2000, a tight timetable intended to allow results to be fed into the evolving monitoring and evaluation framework of the new generation of Structural Fund programmes in Scotland.

The following report consists of six main parts. Following this introduction, Section 2 restates the objectives of the report in full and summarises the activities undertaken. Section 3 provides a review and critique of Structural Fund evaluation practice in Scotland, ending with a set of preliminary conclusions on the challenges facing the Scottish Executive, Programme Management Executives and partnerships in developing further the evaluation framework in Scotland. Section 4 begins the review of international practice with a general discussion of the contrasting 'cultures' of regional policy evaluation in different EU countries and the importance accorded to Structural Fund evaluation. Section 5 provides an extensive international review of the management and organisation of Structural Fund evaluation and presents relevant case studies. This is complemented by the international review and case studies of evaluation methods in Section 6. Finally, Section 7 presents recommendations and options for future Scottish Structural Fund evaluation. Annexes appended to the report provide: the research specifications; a list of evaluation documentation analysed for the Scottish review; details of those interviewed during the research process; and, a glossary of key terminology relating to Structural Fund evaluation.

The report has been written by a study team consisting of Professor John Bachtler, Laura Polverari and Sandra Taylor (European Policies Research Centre) and Professor Brian Ashcroft and Professor Kim Swales (Fraser of Allander Institute). Specialist research input has also been provided by several other EPRC staff: Philip Raines, Mary Louise Rooney, Rona Fitzgerald, François Josserand and Rona Michie.

A preliminary version of Sections 2 - 4 was presented verbally at two consultation meetings, involving Scottish Executive officials and programme managers respectively, which took place in Edinburgh on 22 May and in an Interim Report submitted to the Scottish Executive on 13 June 2000 and discussed by the Steering Group on 19 June 2000. The study team is grateful for the feedback from participants at the meetings, as well as the time and assistance provided by programme managers, officials and other members of the evaluation community interviewed as part of this research. Further feedback in response to this report would also be welcome.

2. OBJECTIVES AND STRUCTURE

2.1 Objectives

The overall aim of the research is to contribute to the overall objective of the new Scottish Executive - confirmed in the recent review of the PMEs - to measure the impact of the new round of Structural Funds so that they leave a real lasting legacy for the future. The study will inform the development of an Evaluation Strategy for the Scottish Structural Fund programmes.

Specifically, the project has five key objectives:

- (a) *to analyse the Scottish evaluation framework in the context of the new round of Structural Funds programming*, providing a preliminary assessment of current approaches and the implications of new Structural Fund guidelines for existing practice;
- (b) *to examine how the evaluation of Structural Fund co-financed programmes is organised and conducted elsewhere in the EU*, giving consideration to institutional arrangements for managing and conducting evaluations, the resources used for evaluations, and the relationship between evaluation and programme design, programme management and project delivery over the short, medium and long term;
- (c) *to review the spectrum of evaluation methods used across the EU for assessing the effectiveness of Structural Fund expenditure*, providing a taxonomy of methods and techniques, and examining commonalities and contrasts with the Scottish approach;
- (d) *to analyse the strengths and weaknesses of evaluation methods used in other parts of the EU*, identifying examples of ‘good practice’; and
- (e) *to provide clear recommendations for improvements to the monitoring and evaluation framework in Scotland*, taking into account the difficulties associated with the evaluation of the overall effectiveness of the Structural Funds in the context of a devolved region.

2.2 Methodology

The research for this study was undertaken over a three-month period using a mix of desk research and field work. The research involved six main elements.

First, the desk research involved reviewing the evaluation guidance produced by the European Commission, the UK Department of Trade & Industry (DTI), the UK Department of the Environment, Transport & the Regions (DETR, formerly the Department of the Environment) and the Scottish Office/Executive. Other policy documentation (eg. the MEANS reports, EC conference papers, etc) and relevant academic literature were also considered.

Second, the study team examined an extensive set of Scottish Structural Fund programme documentation and evaluation studies, starting from the 1989-93 programmes through to the latest interim evaluations of the 1997-99 programmes. Some *ex ante* appraisals from the 2000-06 programmes have

also been examined. A full list of the documentation used is appended to this report (Annex 2).

Third, the desk research was complemented by a programme of face-to-face interviews with programme managers of all the Scottish Structural Fund programmes, the Scottish Executive, the Northern Ireland Department for Finance, Department of the Environment, Transport and the Regions, Department for Education and Employment and Department of Trade and Industry, and the main evaluators undertaking Structural Fund evaluation studies in Scotland (a list of interviewees is contained in Annex 3). The preliminary results of these interviews were 'tested' in two meetings with Scottish Executive officials and representatives of the Programme Management Executives, the results of which have been incorporated into this report.

Fourth, the team undertook a programme of over 30 telephone and face-to-face interviews with government officials and evaluation experts in all of the Member States (a list of interviewees is contained in Annex 4) to assess evaluation practice, to explore case studies and to consider their practical application in Scotland.

Fifth, the study team collated and analysed background information on various approaches to analysing regional policy, and the Structural Funds in particular, in other EU countries.

Finally, the research team undertook a series of internal discussion meetings to review the results of the analysis and identify conclusions and recommendations.

3. STRUCTURAL FUND EVALUATION IN SCOTLAND: REVIEW AND CRITIQUE

Reporting on the first part of the research, this section reviews the approach to evaluating Structural Fund programmes in Scotland. It begins with some general reflections on the context for Structural Fund evaluations – the number and scale of programmes and some of the ‘cultural’ influences on the evaluation process. The section then systematically describes the management and organisation of programme evaluation in Scotland, the methodological processes, targets, indicators and data, the methods and techniques. It charts the considerable progress in Scotland, underlining good practice to date, and highlights some of the outstanding issues which could usefully be addressed in the next programming period.

3.1 Context

Scotland has some of the most extensive experience of EU regional development programmes in the European Community. The first integrated development operations in Scotland pre-date the reform of the Structural Funds and begin with the early Community programmes focused on the West of Scotland, complemented by programmes for other parts of the country from 1989 onwards:

- Glasgow NPCI (National Programme of Community Interest), 1986-88
- Strathclyde Integrated Development Operation, 1988-92
- Western Scotland Operational Programme (OP), 1993
- Highlands & Islands NPCI, 1989-91
- Highlands & Islands OP, 1992-93
- Galloway Community Support Framework (CSF), 1989-93
- Eastern Scotland CSF 1989-93, comprising four OPs for Fife, Lothian, Tayside and Central Scotland

Following the reform of the Structural Funds in 1994, Scotland again benefited from a wide range of programmes under the spatial Objectives 1, 2 and 5b and the horizontal Objective 3 and (later) Objective 4:

- Highlands & Islands Objective 1 1994-99 Programme
- Eastern Scotland 1994-96 Objective 2 Programme
- Eastern Scotland 1997-99 Objective 2 Programme
- Western Scotland 1994-96 Objective 2 Programme
- Western Scotland 1997-99 Objective 2 Programme
- Dumfries and Galloway 1994-99 Objective 5b Programme
- Borders 1994-99 Objective 5b Programme
- Rural Stirling and Upland Tayside 1994-99 Objective 5b Programme
- North and West Grampian 1994-99 Objective 5b Programme
- UK 1994-99 Objective 3 Programme

➤ UK 1997-99 Objective 4 Programme

These programmes were complemented over the 1989-99 period by the EU Community Initiatives, such as Rechar, Retex, Resider, SME, Konver, Urban, Adapt, Employment, Leader, Pesca, Renaval and Interreg.

In the new programming period, a further five major programmes will be implemented:

- Highlands & Islands 2000-05 Special Transition Programme
- Eastern Scotland 2000-06 Objective 2 Programme
- Western Scotland 2000-06 Objective 2 Programme
- South of Scotland 2000-06 Objective 2 Programme
- Scottish 2000-06 Objective 3 Programme

The financial allocations associated with Structural Funds in Scotland since the mid-1980s are difficult to compare, but Table 3.1 provides an overview of the available figures.

Table 3.1: European Structural Fund Allocations and Commitments for Scotland at 31 December 1999 (£ million)

Programmes	1986-89 Commitments	1989-93 Commitments	1994-99 Commitments	2000-06 Allocations
Objective 1	0	0	227	194
Objective 2	109	580	587	521
Objective 3 & 4			250	310
Objective 5b		340	100	
Community Initiatives			52	68
Total	109	920	1,216	1,093

Source: Scottish Executive²

Over a period of almost 15 years, comprising several programming periods, there have been major changes in the content, management and implementation of Structural Fund programmes in Scotland. These have been responses to changes in the regulatory environment, the evolution of the institutional context for regional and local development in Scotland and learning processes associated with programme management.

First, and most notably, there has been a partial 'withdrawal' of the Scottish Executive from the active implementation of programmes through the outsourcing of practical administrative tasks to Programme Management Executives (PMEs). Second, there has been a major increase in partnership involvement in the design, management and delivery of programmes. Third, there has been a continuous process of innovation and refinement in the practices and methods of programme management, encompassing strategy

² The data available for the 1975-93 period are not broke-down by programme. The 2000-06 allocations for the Community Initiatives included in the table, furthermore, are not final.

development, programme steering, project generation, project appraisal and selection, and monitoring and evaluation.

With specific respect to evaluation in Scotland, the process of developing an evaluation framework has been largely ‘top down’, driven by the Scottish Office (and other UK Government Departments), initially with a view to establishing a common base for the spatial Structural Fund programmes in Scotland with common indicators and baselines. While the Scottish Office/Executive has had a key role in setting the framework and providing guidance, it has been the PME’s that have given practical effect to the guidance and the increasing commitment of resources to evaluation. Arguably, for much of the past decade, the exercise of evaluation in Scotland has been driven by the need to fulfil the regulatory requirements and meet EC expectations. However, this should not detract from the compliance effort: few Member States and regions in the EU were willing or able to respond to the relatively basic evaluation demands of the regulations. In Scotland, the evaluation framework delivered evaluations that were conducted, submitted and approved on time, responding adequately to EC requirements.

Evaluation methods could also be described as pragmatic, focusing on the requirements of the Regulations and addressing practical process issues relating to the management and delivery of programmes. A robust assessment of efficiency and effectiveness was not undertaken, but this reflected constraints and feasibility issues rather than being a deliberate exclusion of these themes from the agenda. Many constraints, in fact, were outwith the control of the Scottish Office/Executive and PME’s: for example, the imposed timetable for the evaluations was often inappropriate and did not allow optimal exploitation of studies for programme needs.

Lastly, as elsewhere, a challenging dimension of Structural Fund evaluations in Scotland has been the need to respond to the interests of different stakeholders including the EC, the UK Government, the Scottish Executive and Parliament, the PME’s, the programme partners and the final recipients, many of whom have different immediate objectives from evaluation. While evaluations to date have responded to the various agendas of stakeholders, they have done so with differing degrees of success, with relatively limited emphasis placed, for example, on robust impact analysis, and often a stronger focus on evaluating operational performance and partnership processes.

3.2 Management

The starting point for assessing the evaluation of the Structural Funds in Scotland is the management and organisational framework. As in other areas of EU regional policy in the UK, national authorities have historically exercised a leading management role with respect to the evaluation of Structural Fund programmes. The formal influence of the Scottish Office varied in the past for different programmes and funds. In the last programming period, the evaluations of Objectives 1, 2 and 5b were managed by the Scottish Office, while the Objective 3 and 4 ESF programmes were a UK-level competence, the Scottish Office having had only a minor role in their evaluation. The interventions within the Objective 3 and 4 programmes were co-ordinated and evaluated mainly by the Department for Education and

Employment (DfEE) and its Evaluation Unit in Sheffield. Following devolution this situation has changed, and the Scottish Executive now has the lead role for all aspects of Structural Fund evaluation in Scotland.

From a historical perspective, the pre-1994 period was one in which Structural Fund monitoring and evaluation activity was limited and relatively weak in Scotland – as in many other parts of the EU, although with steps progressively being taken during this time to strengthen arrangements.³ With the launch of the 1994-99 programming period, efforts were made to improve responses in the UK and in Scotland in particular, both in line with the requirements of the new regulatory framework and the importance accorded to evaluation by the UK Government. The first step was the analysis of previous evaluation reports, annual implementation reports and programming documents. This led to the identification of several monitoring and evaluation weaknesses during the 1989-93 programming period that needed to be addressed in the new programming period.

First, programming documents contained few precise objectives, indicators, and impact and output targets. At the time the programmes were drafted, the main focus was on the identification and funding of eligible projects and on the maximum exploitation of the resources available. At this stage, value for money was probably secondary to the primary requirement of distributing the funds. On the operational side, this led to a series of problems and inconsistencies. At the appraisal stage, projects were sometimes allocated funding under the wrong priority, and considerations of the specific relevance of a project to the aims of a priority were often left aside, leading to the approval of less effective projects.

Second, there was a lack of monitoring guidance for partners and applicants, leading to variations in the monitoring data collected within and across programmes. The monitoring activity undertaken was relatively poor in many cases, and oriented primarily towards financial information. The quality of the monitoring data supplied was constrained by the ability of project managers to report accurately on their activities. Core terminology and concepts, even for some basic issues (eg. definitions of tourism and differentiation between jobs created and safeguarded), were used differently by different implementers and PMEs, with predictable consequences in terms of drawing reliable conclusions about programmes outcomes.

Such factors negatively affected the efficiency and significance of evaluations: there was no agreed understanding of what programmes were trying to achieve and, therefore, of the basic premises against which they were to be evaluated. Evaluators had to make considerable efforts to gather basic activity, output and impact data before they could consider undertaking deeper analysis into the impacts generated.

Improvements to the evaluation framework, initiated by the Scottish Office, started from an acknowledgement of these criticisms. A working group, comprising Scottish Office staff and representatives of all seven programme areas, was established to discuss and identify ways to plan a ‘rolling

³ CEC (1997) *Ex Post Evaluation of the 1989-93 Objective 2 Programmes - Synthesis Report*, Series: Evaluation and Documents, No 4, Commission of the European Communities, Brussels.

programme of evaluation', that would allow the early and continued incorporation of evaluations into the programme implementation calendar, reducing the need for 'eleventh hour' responses to predictable management demands.⁴

The first task was mainly definitional, aiming to construct a taxonomy that could be applied to all programmes. An 'evaluation plan' was agreed for each programme, indicating what would be done, when and how in relation to monitoring and evaluation. These plans indicated that it was necessary to produce baseline information against which progress could be tested, and that economic statistics would be collected at top-down (macro) level and project information at bottom-up (micro) level. Project information would be gathered in the first instance from application forms and subsequently from quarterly claim forms, for which a single format was to be elaborated across all seven programmes.

For each programme, measure-level activity and intermediate output indicators were extrapolated. A uniform definition of activities and related intermediate outputs applicable to all programmes was identified, so that it would be possible to collect the same information - with the same meaning - across Scotland. The same exercise was carried out for macro-economic impact indicators, of which a list was drafted and circulated to all Monitoring Committees. Specific problems were also addressed, such as the difference between jobs created and safeguarded, and the problem of double counting of impacts.

At the end of this exercise, which took about two years (1994-96), a common language had been established and an effective monitoring system implemented. The work undertaken made it possible to implement a monitoring system common to all programmes which could form the basis for a comparative assessment of programme performance across Scotland. Consistent core data were collected across all programmes which enabled the implementation of the planned rolling evaluation and, as a result, an overall global picture of the outputs, and in part the results, which had been achieved with Structural Fund support.

Over the same period, the Scottish Office had also been working with other UK Government Departments in the context of GAMESF, first producing guidance on prior appraisal and project appraisal and then on interim and *ex post* evaluations.⁵ The latter work was undertaken jointly by the Scottish Office and Department of Environment, which together produced (in 1996) a common framework for the evaluation of European Structural Fund interventions which took into account the Commission's guidance on monitoring and evaluation.⁶ This framework provided the basis for interim

⁴ Described in Williamson E (1996) *The Development of the concept of Rolling Evaluation of European Structural Funds in Scotland*, Paper Presented to the 1996 Berlin Means/DG Regional Policy Conference on Structural Fund Evaluation, 2-3 December 1996.

⁵ DoE/SO (1996) *Framework for the Evaluation of Area Based European Structural Funds*, Department of the Environment (Regeneration Directorate) and Scottish Office (European Funds Division).

⁶ CEC (1995) *Common Guide for Monitoring and Interim Evaluation*; Commission of the European Communities, Brussels. CEC (1996) Discussion Paper; *Monitoring and Evaluation: Application of European Policies* Research Centre

and *ex post* work carried out on ERDF programmes during the second half of the 1990s.

The framework contained a clear and simplified description of the aims, types, approaches and methodologies for the evaluation of Structural Fund programmes and suggested draft terms of reference that the PMEs could apply in commissioning their interim evaluations. Furthermore, the specification of aims and objectives for the evaluations provided an operational framework for evaluators who used them to develop appropriate methodologies for specific evaluation needs. The indicative terms of reference drafted by the Scottish Office listed a precise set of tasks for evaluators to fulfil. This clarity, supported by the monitoring system which had already been implemented, made more reliable and comparable evaluations more feasible.

Overall, the approach taken to evaluation and the achievements to date have been positive. Monitoring systems have been established (in part through bottom-up initiative eg. by SEP in Western Scotland) and adequate evaluations have been delivered according to the required schedule. This has not been the case everywhere: due to the efforts made, Scotland is perceived by the European Commission as one of the more innovative countries in the EU in terms of evaluation. However, there are also some weaknesses that should be highlighted.

- *Standardisation has not always led to optimal responses.* The standardisation effort carried out by the Scottish Office, eg through the issuing of guidance and the structuring of evaluation responses, has improved the rigour and consistency of evaluation activities overall, but its scope may have meant that some important issues received insufficient attention, eg. the evaluation of issues specific to rural areas. For example, as part of the 'rolling evaluation' approach, SO co-ordination efforts anticipated the development of programme-specific evaluation plans for all 1994-99 programmes - a positive outcome of standardisation. However, these appear to have focused on organisational issues rather than the technical and economic evaluation needs of different programmes, leading to other issues being neglected.
- *Scottish Office/Executive guidance has improved the overall quality of evaluation activity, but may also have encouraged a somewhat passive approach to evaluation design among PMEs.* The guidance issued to support programme-level evaluation efforts has been beneficial in that it has efficiently provided technical support to programme-level actors with limited time, resources and expertise to dedicate to evaluation. However, it may also have encouraged a passive over-reliance on the SE/SO, leading to lost opportunities for evaluations to respond to programme-specific priorities and issues. Specimen terms of reference have sometimes been taken up wholesale by PMEs (especially the smaller ones), with only limited modification. Among the consequences is that the feasibility of an evaluation study as set out in the terms of reference may not always be verified in advance (relative to budgets, timetables and data availability), potentially undermining the ability of evaluators to meet expectations. A

further outcome has been strong similarities between the evaluation proposals made by some evaluators across programmes, with the same standard evaluation model being sold to different programmes with limited tailoring. The practice may also have contributed to the predominance of a restricted group of evaluators, who have developed quick and effective methodologies responding to the requests from PMEs.

- *Difficult past integration of ESF and ERDF evaluation activities.* ESF remained mainly outside the full competence of the Scottish Office in the last round, national ESF programmes being dealt with at UK level (DfEE). This resulted in a lack of coherence and uniformity in the data collected for ESF and ERDF programmes, leading to difficulties in integrated assessment of the overall impact of both kinds of intervention. This distinction was also reflected in multi-fund programmes, where the application and project selection procedures and processes for the two funds were separate.
- *Evaluation has not yet attempted to identify the real impacts of Structural Fund expenditure.* Formal evaluation requirements have been satisfied, but the next step, of making a qualitative ‘jump’ to identify the real impact of the Structural Funds in Scotland, has not been fully addressed.
- *There is no framework for the overall evaluation of economic development activity in Scotland.* There has been a lack of political will to connect all economic development policies under a common monitoring and evaluation framework in Scotland. Efforts to improve the evaluation of Structural Fund interventions have drawn on parallel initiatives in Scotland, eg. Scottish Enterprise evaluation methodologies, but an overall framework has not been developed which would enable the ‘bigger picture’ to be assessed. Issues such as additionality and displacement are therefore difficult to address. The fragmentation also means that it is difficult to identify the overall net impacts of economic development efforts either at Scottish level or locally.⁷
- *Insufficient involvement of other stakeholders in setting the parameters for evaluation.* Although there has been ongoing joint working between the Scottish Office and programme-level actors, notably programme managers, some Scottish Office initiatives to develop evaluation practices have been perceived by some actors at programme level as having been ‘imposed’, with insufficient emphasis given to integrating the experience of programme managers and partners. The evaluation framework, although unanimously deemed as a positive contribution, could also have been more flexible in order to meet the more specific needs of the various stakeholders - identifying the choices available to programmes depending on their circumstances, or highlighting the areas where identification of programme-specific responses would be most beneficial.

⁷ In June 2000 the Scottish Executive published the document ‘The Way Forward: Framework for Economic Development in Scotland’ an operational document that identifies priority areas for action for the attainment of a sustainable and distributed economic growth. This overall policy framework may represent a first step also for the creation of an overall evaluation and monitoring framework, but it is yet to early to assess whether this will occur.

3.3 Programme-level organisation of evaluation

3.3.1 Experience of evaluation to date

The growth of an evaluation culture among Structural Fund actors has progressively permeated the activity of the Scottish Office/Executive, and also the PME's and subsequently the partners. Evaluation is perceived as a useful programming tool that has been used to focus strategies, adjust problematic interventions, test ideas and to develop a feeling of 'ownership' and joint steering among programme stakeholders. More generally, there is an acceptance of evaluation as an integral part of Structural Fund programming which seems to extend from programme managers to the partners. Independent evaluators are seen as being able to present a picture of programme performance to partnerships and to stimulate discussion and more open feedback on controversial issues.

In terms of the extent of activity, the regulatory requirements related to evaluation have been met, with studies undertaken according to the required schedule - even where this has not necessarily been the optimal timetable from the point of view of programme actors.

There are several examples where evaluation initiative has gone beyond the regulatory minimum. All programmes have undertaken or inspired at least one thematic evaluation (eg. the Dumfries and Galloway thematic studies on ICT, tourism and community economic development, inspired by the 1994-99 Objective 5b interim; the Highlands & Islands study on the environment; the Eastern Scotland study on tourism and sustainability; and the Western Scotland study on project selection undertaken within one of its interim evaluations). In all cases, these thematic analyses were motivated by specific objectives, for instance the need to re-launch or reshape specific measures that, although deemed strategic for the development of the programme area, were not bringing forward sufficient or appropriate projects. These evaluations were really programming and implementation tools and fed directly into the implementation process.

As far as the results of evaluations are concerned, evaluations have often confirmed suspicions about trends and processes as much as uncovering new issues. This is nonetheless a valuable function, especially from the point of view of PME's. The formal identification of issues by impartial evaluators focuses common attention on addressing them.

Where evaluation recommendations are seen as fair and accurate by programme participants, they have tended to be acted on, often very rapidly eg. the operational recommendations on project selection in the Dumfries & Galloway 1994-99 interim evaluation were quickly addressed to improve the performance of the second half of the programme. The overall process has been seen as worthwhile. Operational issues, in particular, have been addressed with success. Those contributing to evaluations (ie. as steering group members or evaluation interviewees) perceive the changes resulting from evaluation recommendations as evidence that their input to the studies has been useful. In turn, this reinforces the credibility of evaluation as a programming tool rather than a regulatory obligation.

The rigid timetable for evaluations, imposed by the Commission, on the one hand gave a certain urgency to evaluation requirements - and so ensured that studies were undertaken. On the other, it meant that the utility of studies was sometimes limited because they were done too soon. This has been the case for interim evaluations in particular, undertaken according to a strict timetable even after a delayed start to the programmes. Interim evaluations were sometimes undertaken when few projects had started and even fewer were able to demonstrate impacts, limiting what could be analysed. Conclusions could be drawn about operational issues at this early stage, but not about strategic relevance.

In fact, across all evaluations of Scottish Structural Fund programmes, while procedural and process aspects have been addressed effectively, and outputs have usually been summarised, there has been a general tendency to neglect the need to identify overall economic impacts. This is an issue discussed in greater depth in the methodological sections of this report.

3.3.2 *A developing culture of evaluation*

In terms of the culture of evaluation and the development of skills related to evaluation, there has been a significant increase of awareness and competence at programme and partner levels since 1994. Among programme management teams, the responsibility for evaluation has increasingly been taken seriously, associated with greater skills and commitment. In general, the PMEs tend to allocate the responsibility for evaluation to an individual member of the management team. This provides continuity and an historical memory, enabling expertise to be accumulated, making evaluation more 'visible' to partners and programme beneficiaries.

In addition to taking the evaluation responsibility seriously, several programmes have undertaken or participated in innovative work to reinforce evaluation frameworks, for example in regard to the horizontal dimensions including sustainable development, gender mainstreaming and RTDI. Part of this work has involved the definition of indicators and benchmarks to increase evaluability.⁸ In relation to environmental sustainability, for example, the establishment of core sustainability indicators has been undertaken in Eastern Scotland⁹ with impact assessment in support of target setting in Western Scotland. Also of note has been work in the Highlands and Islands during the mid 1990s into environmental aspects of Structural Fund programming, and especially their evaluation.¹⁰ For gender mainstreaming, the Toolkit elaborated with the support of the Scottish Executive, Equal Opportunities Commission and HERA 2001 included advice on the integration of gender mainstreaming evaluation into the programming cycle.¹¹ Western Scotland has also undertaken useful work in this area.

⁸ Clement K (2000) *Economic Development and Environmental Gain*, Earthscan, p118-121.

⁹ Eastern Scotland European Partnership (1998) *The Sustainable Development Project: Consultative Report*, ESEP Programme Executive, Dunfermline.

¹⁰ Fulton A and Siddy D (1996) *Strategic environmental assessment of the Highlands and Islands Objective 1 - Programme 1996*, Paper presented to the Berlin EC/Means evaluation conference, 2-3 December 1996.

¹¹ Fitzgerald R (1999) *Toolkit for Mainstreaming Equal Opportunities in the European Structural Funds*, Equal Opportunities Guidelines Steering Committee.

However, there are limits to the resources available at programme level and thus to the capacities which can be developed - especially since expectations and methods in this field are constantly evolving. Evaluation is one of many concerns, albeit an increasingly prominent one, and the PMEs are often too under-staffed to commit sufficient resources to this area. Dedicated resources are generally modest even in the largest programmes, and evaluation managers have multiple responsibilities.

The capacity of different programmes varies. The larger industrial/urban programmes have more staff, more resources and more regional organisations to draw on (eg. for statistics, parallel studies, analytical capability and expertise). Smaller PMEs inevitably rely more on Scottish Executive guidance which, in turn, has often responded more to industrial/urban issues than rural ones, because of their relative financial importance. In this framework, two additional sources of guidance have become valuable to the smaller programme management teams: evaluation consultants themselves; and the expertise of other PMEs, consulted on an *ad hoc* basis or in the course of regular Scottish programme manager meetings. In terms of developing evaluation capacity at programme level, this exchange of experience is a key resource.

Wider developmental information enabling the growth of more 'independent' expertise (such as training opportunities, and the materials produced by the EC-supported MEANS project) does not appear to be circulated widely. Even if it were, the limited time and resources available to PMEs for evaluation and the competing demands on staff time, mean that there is limited interest at PME level in independently having to invest in work on evaluation methodologies. Instead, as noted above, PMEs - especially the smallest ones - tend to trust Scottish Executive guidance and the advice of evaluators rather than taking a more pro-active, critical attitude to defining the evaluation needs of their own programmes.

3.3.3 *Partner participation in evaluation - growing awareness and understanding*

Structural Fund partners have become increasingly used to evaluation as an element of Structural Fund programming and are consequently more aware of the strategic relevance of evaluation. This awareness not only derives from a passive knowledge of a programme's evaluation activity, but more significantly from active participation in the evaluation process, either as members of Evaluation Steering Groups or as the subjects of survey interviews or questionnaires undertaken as part of evaluations.

Where partners understand the purpose and value of evaluation, they usually co-operate willingly and actively. In some cases, a considerable communication and development effort has been required by programme management teams and evaluators to achieve partner support. In some cases, partners have felt threatened by evaluations, especially where relationships with evaluators were difficult, where the purpose of an evaluation was poorly understood (eg. in the initial stages of some 2000-06 *ex ante* appraisals), or where it was not felt that the outcomes of evaluations accurately reflected the real situation (eg. the UK Objective 3 leavers' surveys, whose design gave an unfavourable picture of the performance of Scottish projects and alienated

some partners). Asking for too much information from project implementers, or changing what is required of them in terms of data supply *in itinere* has also damaged partner relations in some cases. Evaluators, moreover, are not always sufficiently sensitive of the need to work with partners and do not always involve them adequately or in an optimal manner.

3.3.4 *The evaluation community in Scotland - relatively skilled, but limited*

As far as the evaluation community is concerned, there have generally been good relationships with the available evaluators, both in terms of PMEs and partners. Over time, evaluators have developed an in-depth knowledge of programmes, areas and partnerships, and are therefore trusted.

Nevertheless, the population of evaluators working in Scotland is small, giving a limited number of actors dominance over the market. Some evaluators have proposed standard solutions to multiple programmes, rather than responding to the specific requirements of individual programmes. This means that while individual programme managers and partnerships might have objected to a multi-programme evaluation exercise being contracted, this is effectively what has been chosen in some cases.

The use of a single organisation to undertake multiple studies over the same time period has also in some cases led to an overload of work and caused discontent among PMEs, which felt they were not receiving the expected level of commitment and attention.

There is a possible lack of open mechanisms to extend the active evaluation community in Scotland. In an initial stage of selection, Evaluation Steering Groups agree a shortlist of potential evaluators who are then invited to respond to the terms of reference and put in bids for the project in question. This process assumes that the best potential bidders are already known to the Steering Group, but this is not necessarily the case, and arguably limits opportunities to widen the evaluation community. The approach, however, does help to avoid some problems which could be associated with more open mechanisms, notably the increased risk of moving outside known and proven suppliers and the greater effort which would be required to familiarise new service providers with programmes, partners and information systems.

Procedures to choose evaluators have been broadly similar in all programmes. A specific evaluation steering group is generally established to manage the evaluation, identifying the consultants and then steering the exercise. In some cases, procedures have been organised specifically to involve evaluators in defining the conduct of the evaluation exercise. For example, in the case of Eastern Scotland, pre-meetings have been held with interested evaluators to discuss the content of the study, timescale, methods, organisation etc, before the submission of the bids.

While experienced Structural Fund evaluators are well-known, some PMEs have experienced difficulties in finding specialist evaluators to address single issues. The Dumfries and Galloway PME, for instance, had some difficulties in finding an expert for a thematic study on telecommunications. ESF evaluation could also be a problematic area in the future: given the major role

that the DfEE has played until now, there could be a lack of experienced ESF evaluators in Scotland.

3.3.5 *The foundation for evaluation - monitoring*

There have been significant developments over time in the quality and coverage of monitoring in all programmes, but some of the basic initial data required for evaluations are still problematic. The quality of monitoring data was highlighted as unsatisfactory in early Structural Fund evaluations, and continues to be so according to the most recent evaluation reports (the interim reports of all 1994-99 programmes and the 1994-96 Objective 2 *ex post* reports).

Early programmes were often poor in terms of baselines and target definition, and information on outputs and impacts was not systematically collected from project implementers. In order to improve the 'evaluability' of programmes, active work has been underway for some time to improve baselines, to set core indicators and targets, to enhance the ability of project implementers to record outcomes accurately and meaningfully, and to increase the active use of the data collected.

Monitoring data are collected through the implementation reports which project sponsors attach to the expense claims submitted quarterly to the PMEs. A major problem has been the inadequate skills of project applicants in completing the relevant forms. In some programmes, efforts have been made to develop such capabilities through training activities, although not all have been successful. The Dumfries and Galloway PME, for instance, organised workshops for project applicants and partners, but these were discontinued because of the perceived excessive effort required to participate. Instead, PME meetings with specific sectors (eg. the voluntary sector) were organised.

Another important element of monitoring activities is monitoring visits, which are required by the Regulations to encompass at least five percent of the total eligible expenditure and a representative sample of projects. In general, all PMEs have tried to interpret the monitoring visits as a chance to guide implementers and help them respond to the information needs of the programmes, to identify implementation bottle-necks and to find possible solutions to these. Clawback is rare, and usually attempts are made to help project implementers overcome problems where good faith is not in doubt. The fact that applicants – in the view of PMEs - generally perceive monitoring visits as an opportunity for useful contact with programme managers is considered positive for PME-partner co-operation in monitoring and improving project performance.

Objectives 3 and 4 have been characterised by particular monitoring difficulties in the past. Since the monitoring forms have changed considerably over time, it has been impossible to draw out a clear picture of what has been achieved, at least in terms of impacts. This was reflected in the final evaluations, where the lack of standard data affected the relevance of the conclusions. The DfEE has developed a new monitoring framework that will be maintained for at least the next three years in England, and which should assist both evaluators and project implementers. The choice for Scotland is

whether to establish a similar framework - including potentially subcontracting aspects such as the leavers survey to the DfEE.

3.4 Methodological process

In considering the methodologies used to evaluate Scottish Structural Fund programmes, the following sections review first the terms of reference and the overall approach of the evaluations and then the specific targets, indicators and data, and methods and technique.

Table 3.2 and Table 3.3 provide respectively a description of the interim and final evaluations conducted for the Scottish Programme areas. Each table highlights for each study:

- whether terms of reference were included;
- the overall approach to evaluation;
- the conclusions drawn on the suitability of programme targets and benchmarks, and indicators and data; and
- the methodological treatment of additionality, secondary impacts, efficiency and effectiveness.

3.4.1 Terms of reference

The tables indicate that very few studies present the terms of reference governing the evaluation. Exceptions are the final report on the *Interim Assessment of the Highlands Islands Objective 1 Programme 1994-99* undertaken by the SQW consultancy, and the *Interim Assessment of the Dumfries & Galloway Objective 5b Programme 1994-99* undertaken by Roger Tym & Partners, which present the full terms of reference in an appendix to their reports. In addition, the *Interim Assessment of the Highlands Islands Objective 1 Programme 1994-99* offers a discussion in Chapter 1 on progress towards meeting the terms of reference.

In the *Interim Assessments of the Highlands Islands Objective 1 Programme 1994-99* and the *Dumfries & Galloway Objective 5b Programme 1994-99*, the terms of reference appear to reflect Scottish Office guidelines on methodology - which is unsurprising, since the Scottish Office Development Department approved the drafting of the terms. Nevertheless, interviews with programme managers revealed that, in the case of Highlands and Islands, the partners had a strong influence on the content of the terms of reference.

The SO evaluation guidelines were drawn up later (May 1996) than the two earliest final evaluation studies considered here - the *Final Evaluation of the Eastern Scotland CSFs 1989-93* and the *Final Evaluation of the Galloway CSF 1989-93*. Both studies mention, but do not include, the terms of reference in the report and both perhaps go further in their evaluation approach than required by the later guidelines - see below.

Table 3.2: Scottish Structural Fund Evaluation Studies – interim evaluations- June 2000

Report & Consultancy	Terms of Reference	Overall Approach	Targets & Benchmarks	Indicators & Data	Additionality & Deadweight	Secondary Impacts	Efficiency	Effectiveness
<i>Highlands & Islands Objective 1 Programme 1994-99</i> by <i>Segal Quince Wicksteed Interim Evaluation 1997 IAHI0</i>	Full terms included.	<i>Bottom-up</i> – sample survey of 101 projects; 19%; interviews with project officers of sponsoring body; use of files for applications and claims data. <i>Top-down</i> – overview of change in main aggregate indicators; comparison with Scotland, UK and EU on some indicators. No top-down assessment.	No specific recommendations.	Need for revised programme monitoring system.	Survey of projects asked whether project would have gone ahead wholly or partially without programme funding. Some attempt to account for additionality and deadweight. Use of judgement. Allowance for double counting. Use NAO research on percent net job creation after allowance for non-additionality, displacement & multipliers.	Not examined directly.	Not considered.	Judgement on progress towards targets and objectives.
<i>Eastern Scotland Objective 2 Programme 1997-1999</i> By <i>EKOS Interim Evaluation 1999</i>	Terms alluded to. Key requirements highlighted	<i>Bottom-up</i> – review of other Obj 2 SPDs for 1997-99 and previous SPD, review of ERDF and ESF data from PME: 40 interviews with Programme partners, staff within PE and officials of DG XVI using structured questionnaire; 10 policy workshops to review particular themes. Telephone survey plus monitoring data (via claim forms). <i>Top-down</i> – no socio-economic review since economic trends examined in review of 94-96 SPD.	Targets and indicators changed since the beginning of the Programme limiting Programme assessment.	No specific recommendations.	Impacts to be assessed in ex-post evaluation.	Not considered.	Not considered.	Judgement on progress towards targets and objectives.

Methodologies used in the Evaluation of the Effectiveness of European Structural Funds

Report & Consultancy	Terms of Reference	Overall Approach	Targets & Benchmarks	Indicators & Data	Additionality & Deadweight	Secondary Impacts	Efficiency	Effectiveness
<i>Western Scotland Objective 2 Programme 1997-99</i> by <i>EKOS</i> <i>Interim Evaluation 1999</i>	Terms alluded to. Key questions presented and answers Discussed.	<i>Bottom-up</i> - review of other Obj 2 SPDs for 1997-99 and previous SPD; review of ERDF and ESF data from PME; 30 interviews with Programme partners using structured questionnaire; postal survey (census) of 130 partners. <i>Top-down</i> - no socio-economic review since economic trends examined in review of the 94-96 SPD.	Targets based on 'sound rationale'. Earlier revision of some targets following independent review for DGXV1.	No specific recommendations.	Impacts to be assessed in <i>ex-post</i> evaluation.	Not considered.	Not considered.	Judgement on progress towards targets and objectives.
<i>Evaluation of Borders Objective 5b Programme 1994-99</i> by <i>Pieda</i> <i>Interim Evaluation 1997</i>	Mentioned but not included.	<i>Top-down</i> – overview of aggregate indicators before and after start of SPD. Comparisons with Scotland. Use of shift-share analysis of employment change. <i>Bottom-up</i> – sample of 30 projects for case study, interviews, grossing up.	No specific recommendations.	No specific recommendations.	Judgement.	Judgement.	Judgement.	Judgement.
<i>Dumfries & Galloway Objective 5b Programme 1994-99</i> by <i>Roger Tym & Partners</i> <i>Interim Evaluation 1997</i>	Full terms included.	<i>Bottom-up</i> - 47 project case studies 25% of total; analysis of application and claims forms; interviews with project officer of sponsoring body. <i>Top-down</i> – socio-economic overview only.	No specific Recommendations.	Confusion between activities and impacts. New target output measures proposed and appropriate indicators.	Judgement from case studies.	Not examined directly.	Not considered.	Judgement on progress towards targets and objectives.

<i>Report & Consultancy</i>	<i>Terms of Reference</i>	<i>Overall Approach</i>	<i>Targets & Benchmarks</i>	<i>Indicators & Data</i>	<i>Additionality & Deadweight</i>	<i>Secondary Impacts</i>	<i>Efficiency</i>	<i>Effectiveness</i>
<i>Evaluation of North and West Grampian Objective 5b Programme 1994-99</i> by <i>EKOS</i> <i>Interim Evaluation 1997</i>	Mentioned but not included.	<i>Top-down</i> – overview of limited indicators before and after start of SPD. Comparisons with Scotland and GB. Too early to gauge SPD impact - used as check on current relevance of SPD priorities & measures. <i>Bottom-up</i> - project files for application and claim form data for ERDF. ESF projects using SO info. 25 interviews with partners and PME. Use of semi-structured questionnaire. Project delays.	A number of targets and their numerical value less than ideal'. Reported project outputs and impacts often do not relate to SPD targets. Need to review and amend initial targets.	Programme area embraces only part of specific local authority or TTWA areas, so data for the Programme area is limited. Not all indicators were found to be reasonable. Too many indicators. Need to identify impact indicators for each Measure.	Focus on gross (physical) outputs and intermediate impacts (direct). Based on case studies, review of project files and judgement. Judgement offered on extent of deadweight in general.	Not specific to projects. General judgement offered.	Not considered.	Only suitability of targets and measures considered.
<i>Evaluation of Rural Stirling & Upland Tayside Objective 5b Programme 1994-99</i> by <i>EKOS</i> <i>Interim Evaluation 1997</i>	Mentioned but not included.	<i>Top-down</i> – overview of aggregate indicators before and after start of SPD. Comparisons with Scotland and GB. Too early to gauge SPD impact – used as check on current relevance of SPD priorities & measures. <i>Bottom-up</i> - project files for application and claim form data for ERDF. ESF projects using Scot Office info. 25 interviews with partners and PME. Use of semi-structured questionnaire. Project delays.	Several targets 'less than ideal' - 'only small proportion of project activity captured within SPD targets. No impact targets provided for a number of measures.	Output data from project applicants is of 'variable quality' Lack of clarity in definition of some indicators and problem of double counting were key issues to be resolved in monitoring arrangements.	Clarification of gross outputs ie. activities associated – and estimate of direct impact based on judgement from "experience of other projects and programme evaluation".	Not considered.	Not considered.	Only suitability of targets and measures considered.

Table 3.3: Scottish Structural Fund Evaluation Studies – ex post evaluations- June 2000

Report & Consultancy	Terms of Reference	Overall Approach	Targets & Benchmarks	Indicators & Data	Additionality & Deadweight	Secondary Impacts	Efficiency	Effectiveness
Evaluation of the Eastern Scotland CSFs 1989-93 by Hall Aitken Associates Final Evaluation 1995 IA	Mentioned but not included.	<p><i>Bottom-up</i> – developed programmed-base from SO's financial monitoring data plus data on output & impacts from each of 1032 ERDF supported projects. Plus analysis of ESF projects, 65 case study subsets plus 32 interviews with P/ship using semi-structured questionnaire.</p> <p><i>Top-down</i>: use of non-EU funded comparator areas in Nottingham & Derby plus GB average using S/S analysis. Scotland also used as comparator for some indicators. No direct link between 'b-u' & 't-d' approaches but compares CSF jobs outcome overall of 17,000 jobs or 4.5% of employment.</p>	No comments.	No comments.	Application of assumed factors for each. Variation according to priority. Grossing up from case studies.	Assumed displacement factor and standard multiplier applied variation according to priority. Grossing up from case studies.	Cost per Job.	Judgement.
Evaluation of the Galloway CSF 1989-93 Objective 5b by Piteda Final Evaluation 1995 IA	Mentioned but not included.	<p><i>Bottom-up</i> – representative sample of CSF projects: project records, interviews with project managers and sponsors. Grossing up.</p> <p><i>Top-down</i> – aggregate indicators, pre-CSF and trends for policy-off and some comparator areas plus national change. Use of S/S analysis of employment. Judgemental linking of b-u and t-d approaches.</p>	Some targets. Little meaning in development terms.	Lack of systematic monitoring for CSF as whole.	Judgement.	Drawn from secondary sources & judgement.	Cost per Job.	Judgement. Spend did not relate well to objectives.

Report & Consultancy	Terms of Reference	Overall Approach	Targets & Benchmarks	Indicators & Data	Additionality & Deadweight	Secondary Impacts	Efficiency	Effectiveness
<i>Eastern Scotland Objective 2 Programme 1994-96 by EKOS Final Evaluation 1999</i>	Terms alluded to only. Key requirements highlighted.	<i>Bottom-up</i> - review of project financial and monitoring data from PME; review of evaluations of projects funded by the 1994-96 Programme; telephone survey using structured questionnaire of sample beneficiaries. <i>Top-down</i> - socio-economic review; some assessment; no attempt to reconcile 'b-u' with 't-d'.	Lack of a clear hierarchy of targets at Programme, Priority and Measure levels.	Inconsistency between indicators and targets Poor quantification of some targets. Need for more data to assess progress towards meeting targets.	Estimates of additionality derived from 'benchmark data' for Obj 2 programmes produced by consultant; from specific evaluation data; and from review of earlier evaluations of Programme projects.	Application of 'off the shelf' displacement and multiplier factors.	Cost per job created or safeguarded.	Judgement on progress towards targets and objectives but limited by narrow range of indicators.
<i>Western Scotland Objective 2 Programme 1994-96 by EKOS Final Evaluation 1999</i>	Terms alluded to only. Key requirements highlighted.	<i>Bottom-up</i> - review of other Obj 2 SPDs on financial allocation and Priority & Measure structures; review of evaluations of projects funded by the 1994-96 Programme; project application data; no further survey work for ERDF; for ESF projects a survey of beneficiaries. <i>Top-down</i> - socio-economic review; some assessment; no attempt to reconcile 'b-u' with 't-d'.	Progress in developing targets from activities to output and impacts. Need for appropriate Targets for 'softer' activity such as networking between companies. Over-ambitious targets at project level.	Monitoring framework not sufficiently comprehensive to identify routes to impact of measures. Poor quantification of some targets.	Estimates of additionality derived from 'benchmark data' for Obj 2 programmes produced by consultant; from specific evaluation data; and from review of earlier evaluations of Programme projects.	Application of 'off the shelf' displacement and multiplier factors.	Cost per job created or safeguarded.	Judgement on progress towards targets and objectives.

3.4.1 Overall approach

Among the evaluation studies reviewed, both *interim* and *final evaluations* appear to adopt 'bottom-up' approaches to the evaluation.

In the *interim* evaluations, activities are identified and some attempt is made at establishing gross outputs, with some but not all studies seeking to calculate efficiency measures, principally cost per job. Some attempts are made at identifying net effects at the project level but not for the overall programme. Generally, the focus of these assessments appears to be more on assessing the appropriateness of priorities, targets, and measures than identifying outputs. This is understandable given the short time available for measures to have an impact by the interim phase and the possibility of introducing changes to some of the measures while the programme still has some time to run.

In the *final* or *ex post* evaluations, there is a greater focus on the identification of net impacts. But in the *Final Evaluations of the Western Scotland and Eastern Scotland Objective 2 Programmes 1994-96*, the official evaluator relied on pre-existing evaluations of ERDF funding that were available. In the *Final Evaluation for the Western Scotland Objective 2 SPD 1994-96*, for example, the evaluator had access to formal evaluation work that had to be undertaken by sponsors of 1994-96 projects submitting projects under the 1997-99 programme. In the *Final Evaluation of the Eastern Scotland Objective 2 SPD 1994-96*, a review was undertaken of evaluation studies carried out on projects submitted during the course of the Programme. A telephone survey of a sample of beneficiaries was also undertaken. In the case of the *Final Evaluation for the Western Scotland Objective 2 Programme 1994-96*, the official evaluator undertook no further survey work, although this did not apply to ESF projects where a survey of beneficiaries was undertaken. These two studies stand in contrast to the two earlier final evaluations, the *Final Evaluation of the Eastern Scotland CSFs 1989-93* and the *Final Evaluation of the Galloway CSF 1989-93*. In both these studies, a representative sample of projects for case study was identified, and survey interviews were conducted with programme partners in the former and project managers and sponsors in the latter.

In the *interim evaluations*, the bottom-up procedure appears to differ between those that undertake a sample interview survey of projects - often project officers from the sponsoring organisation (the *Interim Evaluations of the Borders and the Dumfries and Galloway Objective 5b Programmes for 1994-99*, and the *Interim Evaluation of the Highlands and Islands Objective 1 Programme 1994-99*) and those that survey and interview programme partners (the *Interim Evaluation of Rural Stirling and Upland Tayside Objective 5b 1994-99*, the *Interim Evaluation of North and West Grampian Objective 5b 1994-99*, and the *Interim Evaluation of the Western Scotland Objective 2 1997-99 Programme*). All interim evaluations used the data available in the project application and claims forms.

It is doubtful whether a survey of programme partners is truly bottom up, given the association of each programme partner with several projects. In these circumstances, identification of project performance would appear to rely almost wholly on the information contained in the application and claim

forms. Nevertheless, according to interviews with programme managers, one benefit of involving partners directly in the evaluation interview process is that it encourages partners to take evaluation more seriously.

In both the interim and final evaluations it must be questioned whether the attempt to identify outputs and impacts mainly in terms of jobs created is sufficient to offer guidance on the progress towards the development objectives of each SPD.

Some, but not all studies contain a *top-down* element insofar as surveys of the overall economic context or economic performance of the area, including the time period of the programme, are provided.

In the interim evaluations these area performance surveys, where they are provided, are primarily contextual with the main objective being to check the current relevance of the SPD priorities and measures. However, in the earlier *interim* evaluations (the *Interim Evaluation of the Borders Objective 5b Programme 1994-99*, the *Interim Evaluation of Rural Stirling and Upland Tayside Objective 5b 1994-99*, and the *Interim Evaluation of North and West Grampian Objective 5b 1994-99*) more effort appears to have been devoted to identifying variations in aggregate performance before and after programme inception and between the programme area and spatial comparators. This was particularly the case in the *Interim Evaluation of the Borders Objective 5b Programme 1994-99*, where a more formal attempt was made to control non-programme influences on aggregate employment performance (see section 3.6).

In the four final evaluations considered, all contained socio-economic reviews for the appropriate period and offered a top-down assessment of sorts (see section 3.6 below). In the later evaluations, the *Final Evaluations of the Western and Eastern Scotland 1994-96 Objective 2 Programmes*, no attempt was made to reconcile the limited top-down assessment with the findings from the bottom-up approach. This was not strictly the case with the two earlier final evaluations, (the *Final Evaluations of the Eastern Scotland CSFs 1989-93* and of the *Galloway CSF 1989-93*). In the *Final Evaluation of the Eastern Scotland CSFs 1989-93*, an estimate of overall programme impact on jobs was obtained *via* a top-down approach and this was compared with estimated outcomes from the bottom-up approach. The *Final Evaluation of the Galloway CSF 1989-93*, on the other hand, sought judgementally to link the outcomes of the bottom-up and top-down approaches.

3.5 Targets, indicators and data

3.5.1 Targets and benchmarks

In some programmes, especially those covering 1989-93 and 1994-96, targets appear not to have been adequately specified and quantified for evaluation purposes. Specific weaknesses highlighted in the reports covering these periods included:

- a tendency to specify in terms of activities rather than outputs;
- not all project activity is captured by the SPD targets;
- poor quantification of some targets eg. for technology and innovation;

- significant differences between programme and project targets;
- some targets with little meaning in development terms; and
- a lack of a clear hierarchy of targets at programme, priority and measure levels.

In later programmes, such as the 1997-99 Western Scotland Objective 2 Programme, following a specific indicator development exercise, the choice of targets appears to have been based on a firmer rationale, with better indicator choice and quantification. So, the position today is significantly better than at earlier stages of the Structural Fund programmes.

3.5.2 *Indicators and data*

For several of the early evaluations, the recommended core indicators for Scottish areas had not been incorporated in the programme monitoring system. These evaluations often complained about the inadequacy of data and the lack of a systematic monitoring framework

Later evaluations also found that the monitoring framework was insufficiently comprehensive to enable the process or routes for the impact of each measure to be tracked through appropriate indicators and data, eg. in the *Final Evaluation of the Western Scotland Objective 2 Programme 1994-96* there was concern that the impact of projects on the employment of local residents could not be established because data on the domicile of those gaining jobs was not collected. This problem appears to have continued after the 1994-96 period because the *Interim Evaluation of the Western Scotland Objective 2 Programme 1997-99* considered that there was a need to improve the reliability of monitoring data by requiring that more transparent monitoring arrangements be developed by project sponsors. Nevertheless, the position today is that considerable improvements have been made in the development of indicators and data appropriate to evaluation with the widespread adoption of a common set of indicators and improved monitoring procedures.

Assessments of project performance are very much dependent on the information provided in application and claims forms, which are held by the PMEs and made available to evaluators. However, several of the evaluations found that project sponsors were often confused between activities and outputs, or cause and effect, with double counting evident, eg. of jobs created, across several interlinked projects.

Lack of data, or inadequate data, is a problem for those programmes that embrace only a part of specific local authority or travel to work areas (TTWAs) eg. the *Interim Evaluation of North and West Grampian Objective 5b 1994-99*. It also appears to be a problem for the assessment of effects in certain horizontal themes such as the promotion of the environment and equal opportunities (see the *Interim Evaluation of the Western Scotland Objective 2 1997-99 Programme*, pages 109-110).

3.6 Methods and techniques

3.6.1 *Additionality and deadweight*¹²

Monitoring data on activities associated with project spending was a prime source of information for the estimate of impacts. However, as noted above, such data may be subject to double counting as, for example, when more than one project claims the same job, and/or over-optimistic reporting.

In principle, all projects should be additional and part of the purpose of the project selection process, where projects seeking European funding are subject to prior appraisal and scoring, is to identify and reject those projects that would have gone ahead anyway. There must be some doubt as to the ability of the PMEs and PMCs to 'weed out' all non-additional projects, principally because of asymmetric information problems. Moreover, the difficulty of validating *ex ante* job creation numbers and the precise amount of grant needed to tip the balance (absence of deadweight) would appear to be even greater.

In general, Scottish evaluators make little direct attempt to establish additionality at the project and/or measure level. This even applies at the Final Evaluation stage.

In one of the interim evaluations, the *Interim Evaluation of the Highlands and Islands Objective 1 Programme 1994-99*, estimates of additionality were obtained from a survey of projects that specifically asked whether the project would have gone ahead wholly or partially without programme funding. However, in most of the other interim studies, assumed additionality factors or coefficients were applied with little indication as to their source or relevance.

In the two earlier final evaluations, the *Final Evaluation of the Eastern Scotland Programme 1989-93* and the *Final Evaluation of the Galloway CSF 1989-93*, estimates of additionality reflected the consultant's judgement. In the two later final evaluations, of the *Western and Eastern Scotland 1994-96 Objective 2 Programmes*, estimates of additionality or deadweight were derived from 'benchmark data' for Objective 2 programmes produced by the consultant, or 'specific evaluation data', or a review of earlier evaluations of the programmes' projects. No indication is provided of how these deadweight factors were derived, nor of their relevance to the specific projects in the programme area.

The consultant's judgement, largely unsupported by theoretical argument or evidence, appears to be the main basis for these estimates. Alternatively, the evaluator relies principally on the views of project sponsors.

3.6.2 *Secondary impacts*

The wider economic impacts of projects and/or measures are generally not considered in interim evaluations, as per the Scottish guidelines, because it is considered that such effects take time to emerge. In the final evaluations, secondary effects were established by the application of 'off the shelf'

¹² The concept adopted in this paragraph of additionality is the economic one and relates to the net effect of an intervention. The reader is referred to the glossary for further explanation.

displacement and multiplier factors, which vary according to priority and/or measure. Little or no explanation is offered for the source of such factors or their applicability to the programme area. The consultant's judgement, largely unsupported by theoretical argument or evidence, appears to be the main basis for these estimates.

However, it is worth noting that in the two earlier final evaluations, the *Final Evaluation of the Eastern Scotland Programme 1989-93* and the *Final Evaluation of the Galloway CSF 1989-93*, the top-down analysis of employment performance involved the application of shift-share analysis in an attempt to control the influence of national (GB) share and industry structure factors on the employment performance of the programme area during the programme. While the deficiencies of shift-share analysis are well known, this is a procedure with the potential to isolate the effect of the programme measures taken together on area employment performance including direct and secondary effects. It is surprising that later final evaluations did not at least adopt a similar approach. In part, this is due to the 1996 Scottish Office guidelines on the subject of displacement, for example, which recommend that '*It may be considered appropriate to use displacement estimates produced elsewhere.....rather than evaluators attempting to produce their own estimates*'.

In the view of the study team, this is not helpful advice because it invites the evaluator to ignore the particular circumstances of the programme area and the measures adopted. The use of 'off-the-shelf' displacement estimates is particularly problematic where this does not provide a direct correlation with Structural Funds.

3.6.3 *Efficiency*

In the final evaluations, the main - only - efficiency indicator used is programme cost per job created or safeguarded by a measure. Cost per job estimates are offered even when the aim of a measure is not primarily job creation eg. image enhancement, or the promotion of applied research, technological development and innovation. No other objectives or opportunity costs are considered.

3.6.4 *Effectiveness*

To assess how far a programme's impacts have contributed towards achieving its objectives, evaluators chart progress towards programme and project targets.

Judgement on programme effectiveness, therefore, depends crucially on the:

- appropriateness of targets;
- appropriateness of target indicators;
- accuracy of information on associated activities and outputs; and
- appropriate measurement of net impacts.

In view of the difficulties that evaluators have had in obtaining accurate information and establishing net impacts, there must be some doubt about the evaluators' conclusions on programme and project effectiveness.

3.7 Overall assessment of Scottish evaluation practice

On the basis of the research undertaken on Scottish evaluation practice, it is possible to draw some conclusions about the strengths and weaknesses of Structural Fund evaluation in Scotland. These take account not just of past experience but also the more demanding evaluation challenges of the new programming period.

Scotland has taken an active, structured approach to evaluating Structural Fund programmes.

In contrast with some parts of the EU, the evaluation of the Structural Funds is taken seriously in Scotland. Together with other UK Government Departments, the Scottish Office has developed a common framework for evaluating programmes which has been applied systematically over the past five years. This has promoted shared understanding of the purpose and objectives of evaluation and a common language for discussing methods and techniques. Interim and *ex post* evaluations have been undertaken (largely) to specification and on time, meeting the requirements of the Regulations and the expectations of the Commission. In addition, thematic evaluations have been commissioned where needed by both the Scottish Office and specific PMEs. The quality of responses was recently confirmed when the recent *ex ante* appraisals prepared for the new generation of Scottish Structural Fund programmes (along with their counterparts from other parts of the UK) were commended by the Commission as some of the best *ex ante* evaluations in the EU.

Improvements to the quantification of programmes have significantly improved the consistency of data available.

The investment in monitoring systems initiated in the mid-1990s has made a significant difference to comparative assessment of programme progress among the Scottish programmes. The definition of core indicators, a common terminology, the systematic quantification of benchmarks, targets and outputs, and the enhanced collection of data has enabled much better evaluation studies to be undertaken and more reliable conclusions to be established, at least with respect to programme outputs.

Evaluation is increasingly valued as a useful and important programming tool, particularly for process issues.

Over the past decade, there has been a major change in attitude towards evaluation among those involved at different levels. In the early years of programming, the monitoring and evaluation requirements of the Structural Fund regulations were widely regarded as burdensome obligations which

needed to be met to satisfy the Commission services. Although this attitude persists in some quarters – at least with regard to the amount of work involved – there has been a growing recognition among officials, programme managers and partners and beneficiaries of the value of evaluation in delivering an effective programme. Evaluation has provided an opportunity to assess the progress of programme implementation and to gain feedback from programme partners, enabling corrective actions to be taken (in the case of interim evaluations) or contributing to the planning process in preparing new programmes (in the case of *ex post* evaluations). Programme managers have been using evaluation studies to address perceived shortcomings or verify the efficacy of procedures (eg. project appraisal and selection methods), while the combined interim/*ex post* studies undertaken in 1989-99 were explicitly timed to feed into the preparations for 2000-06. This is, however, an evolutionary process and there is scope for further involvement and motivation of programme partners in future evaluation exercises (see below).

There is a need for more serious evaluation of the impact of Structural Fund programmes.

By contrast with the evaluation of ‘process issues’, until now there has been no real assessment of the impact of Structural Fund programmes. Discussion of the effects of the Funds in Scotland tends to be hedged with uncertainty and qualification, as in many of the evaluation documents, or conducted in terms of outputs rather than impacts. The interview research brought out many observations on the difficulty of addressing impact in small-scale programmes, especially on a time-scale which does not seem to allow this kind of analysis. The problem is serious and the objections credible. However, in responding to the evaluation requirements of the regulations, the Commission’s expectations and the greater political exposure of Structural Fund programmes following devolution, there is a need for the impact of the Structural Funds in Scotland to be seriously addressed.

Methodological rigour in the design and execution of evaluation should be improved.

Some evaluation studies undertaken to date reflect a certain lack of rigour in the extrapolation of conclusions and recommendations. In many cases, consultants have used relatively weakly founded assumptions on which to base their findings. The use of ‘off the shelf’ multipliers, for example, whose relevance to the context in question is not verified and demonstrated, should be discouraged in future. In addition, a more rounded approach could be beneficial which aims to capture wider economic impacts than those on employment.

Further work is required to identify specific methodologies to address distinctive issues more roundly and to place greater emphasis on the quality of outcomes and impacts as well as their quantity.

Evaluating the contribution of Structural Fund programmes to specific horizontal issues, such as social inclusion, equal opportunities, environmental sustainability, innovation and competitiveness, will need the definition of dedicated methodologies and approaches. Past evaluations have mainly focused on identifying economic outcomes in terms of quantitative measures of outputs and, to a more limited extent, impacts. However, there has been limited attention given to capturing the contribution of programmes to, for example, raising technological capacity or improving social sustainability. One of the major challenges for the future will be to insert a qualitative dimension into evaluation, measuring the quality as well as the quantity of programme activity - especially in relation to more intangible outcomes.

In order to appraise the overall effectiveness of economic development policy in Scotland, it is imperative to place the evaluation of Structural Fund co-financed activity in a context of wider policy evaluation.

The impact on the territory of different policies, among which the Structural Funds often represent a minor part, can only be successfully identified by an integrated approach. Issues like displacement and additionality cannot be addressed by isolating a specific programme of interventions from all other factors and conditions. This implies a need to integrate the monitoring and evaluation frameworks used for both Structural Fund and non-Structural Fund supported activities.

Changes of responsibility for the European Social Fund may require a local 'gearing up' in terms of resources and expertise.

In future, ESF will be a full competence of the Scottish Executive, and this represents both an opportunity and a threat. Embracing all the Structural Fund interventions in Scotland within the same monitoring and evaluation framework, through homogenous baselines and indicators, will enhance the possibility of impact assessment and the chance to look at the various co-financed interventions as a whole.

On the other hand, there may be a lack of expertise in the field of ESF evaluation in Scotland because of the former delegation of the monitoring and evaluation functions for Objectives 3 and 4 to the DfEE in Sheffield. Over time, the DfEE has developed a high level of know-how, has elaborated methodologies and evaluation instruments (including the leavers' survey) and has taken an increasingly pro-active approach, through the work of a permanent and sufficiently well-staffed evaluation team. The future evaluation framework for the Objective 3 programme in Scotland is not yet

clear, and an extended co-operation with the DfEE is still under discussion. The reshaping of the monitoring and evaluation system in place, however, seems to be a current priority, given the discontent that previous evaluations (and in particular the leavers' surveys) have caused in Scotland. Whatever solution is adopted, the allocation of new responsibilities to the Scottish level will require a 'gearing up' and a clear re-think of what has to be achieved through monitoring and evaluation and how.

Evaluation would benefit from improving the capacities and skills of PME's and partners and widening the evaluation community.

Another challenge will be to widen the community of evaluators, these being deemed well-prepared and responsive, but too restricted in terms of numbers and range. It will also be important to develop evaluation skills and expertise among PME staff, allowing PMEs to be actively involved in the design of the overall Scottish evaluation framework and then in the design of specific evaluations, focusing consultants on the specific needs of each programme. This will also enhance the ability of PMEs to interest the partners more in evaluation. This last aspect is particularly relevant to the smallest partnerships, where capacity and expertise may be limited and where evaluation is not always perceived by everyone as a really effective management tool.

In some cases, the involvement of partners has been difficult in the past. Thus, it will be necessary in the future to manage evaluation sensitively: if recommendations are to be followed up, it is crucial that they are supported and 'owned' by main programme actors. There is a potential new role for the Scottish Executive in sensitising wider partnerships to evaluation. This could be a cost-effective way to introduce a common denominator. Each single PME could then build on this work with their own partners, exploiting the co-operative relationships developed over time, and building more specific skills and increased direct involvement.

Additional thought about the timing of future evaluations could be beneficial.

Appropriate timing is essential to the utility of evaluations. Each type of evaluation has a rationale and specific aims which are related to the stage of implementation which has been reached: *ex antes* support programme preparation; interim evaluations contribute to the identification of initial results, the assessment of the relevance of Structural Fund assistance and the operation of monitoring and implementation, and supply feed-back to the programmes; and *ex posts* evaluate the effectiveness, efficiency and impacts of assistance and identify the factors that have contributed to the success or failure of programmes. Timing is key to evaluation – but as long as it is imposed externally through the regulations and is not negotiable, the utility of

interim evaluations may be undermined. Ways need to be found to maximise the usefulness of evaluation even within this tight regulatory framework.

There is a need for the realism of evaluation expectations to be verified routinely at the outset of evaluation studies to avoid subsequent failure.

The passive re-proposal of indicative terms of reference can lead to insufficient attention being given to the feasibility of the study requested. It needs to be verified that the scope and objectives of a proposed study are proportional to the budget available. What can be achieved - and how - will in turn be influenced by the range and state of the starting data available to an evaluator and the time available for a study to be undertaken. A more active discussion process before evaluations are commissioned may help to ensure more realistic expectations. It is arguably preferable to limit the scope of studies to addressing the most important questions - especially where budgets are limited - rather than setting out objectives which are too ambitious and unlikely to be met.

Thematic evaluations have proved their worth in addressing specific programming issues. There is potential for continued use of such instruments, both on an ad hoc basis (as currently) and also as part of a more strategic programme of evaluation.

In addition to the obligatory evaluations, some thematic evaluations were undertaken during the last programming period, initiated variously by the Scottish Office/Executive and the programme level. These were deemed useful by those involved, responding to specific information deficits. In some cases, thematic evaluations were inspired by the findings of obligatory evaluations, notably at the interim stage, which highlighted specific problems that needed to be addressed in a targeted way. The value of such *ad hoc* thematic evaluations, commissioned quickly to resolve issues as they arise has been proven. Scope should be there for these to continue to be used where necessary.

A key principle of interim and *ex post* evaluation is that the framework for such studies should be established as early as possible (eg. defining the data to be collected for later analysis). It is possible that thematic evaluations could also benefit from such forward planning. To introduce a more strategic dimension to the multi-programme evaluations, the PME and SE could agree a limited number of thematic evaluations on issues of relevance to multiple programmes (eg. ICT and gender mainstreaming) early in the programming period. A strategic discussion about thematic evaluation could also provide a mechanism to focus efforts on those issues agreed by all core stakeholders to be of most interest.

A more explicit partnership approach to evaluation could be encouraged, explicitly recognising the priorities and objectives of different stakeholders.

The explicit acknowledgement of the various stakeholder needs and interests in evaluation could be increased. In particular, true impact evaluation has been relatively neglected in evaluations commissioned at the programme level, which have tended to focus on issues of most immediate interest to the horizontal partnership, (eg operational performance, programme balance and continuing policy relevance). True impact analysis is a stronger priority at Scottish Executive level, but this objective has until now been relatively neglected.

There is further room to strengthen the validation, dissemination and exploitation of evaluation findings.

Evaluations and their recommendations could be diffused more broadly, and not only among the restricted community of practitioners and programme managers. This need has been expressed clearly by the programme managers interviewed during the fieldwork undertaken. Evaluation, more than other aspects of policy-making, embodies a difficult trade-off between technocracy and democracy: a wide diffusion of the findings of evaluation will enhance the possibility of building a democratic consensus around programmes and of focusing new interventions more on the needs of beneficiaries and other stakeholders.

4. INTERNATIONAL REVIEW OF STRUCTURAL FUND EVALUATION: CONTEXT

Having explored the strengths and weaknesses of the approach to Structural Fund evaluation in Scotland, the following sections review how this compares with practices in other parts of the EU. This section begins the discussion by considering the different evaluation traditions and experiences in EU Member States and the influence of Structural Funds on national practice. It then traces the evolution of monitoring and evaluation across the EU over the past decade, highlighting the progressive improvements over successive programming periods and the variations in commitment to Structural Fund evaluation across Member States.

Chapters 5 and 6 then go on to discuss international Structural Fund evaluation practice in more practical detail, enabling Scottish approaches to be placed in context and using case studies to highlight distinctive and good practices in terms of (i) evaluation management and (ii) methodologies from which Scotland might benefit.

4.1 Diverse attitudes towards evaluation

There are major differences in the practice of evaluation across the European Union, for both Structural Fund and other evaluations. At one end of the spectrum are Member States with a long tradition of public policy evaluation. In these countries, evaluation is understood as an integral part of the policy making process, in which policy formulation and evaluation are linked in an iterative process of policy evolution. Regional policies are regularly examined in terms of their outcomes, ranging from the specific effects of individual regional aid schemes on individual companies, to the aggregated effects of individual policy measures, to the wider effects of a policy package on regional economic performance.

At the other end of the spectrum are Member States without a tradition of regional policy evaluation. In some cases, this may reflect the limitations of evaluation resources and skills in systemising policy design and revision, but it generally represents a broader scepticism towards the value of evaluation studies as tools of policy assessment. Where evaluations do take place in these countries, they tend to be one-off exercises, and are not a clearly-defined, regular stage in policy making.

In spite of the differing views on evaluation, attitudes have been changing towards policy evaluation across the EU as a whole. The experience of operating Structural Fund programmes and the role of Commission guidelines have significantly affected the way evaluation is considered in most Member States. There is evidence of more active evaluation in countries that hitherto conducted few evaluations, not just as a matter of necessity, but also as a more positive response to the benefits of this activity. This trend has also altered how evaluations are conducted and their results used in countries which are already active in this area.

Table 4.1: Evaluation traditions and experience in EU Member States

Member State	Evaluation tradition
Austria	Little tradition of evaluation, but several regional policy measures have been evaluated in recent years. Emphasis of studies appears to be broad, covering institutional and procedural issues, as well as quantitative evaluations of effectiveness.
Belgium	No regular evaluations of regional policy in either Flanders or Wallonia. Structural Fund experience is leading to some impact evaluation studies of individual measures being carried out.
Denmark	No national regional policy measures since 1992 – therefore no national regional policy evaluation.
Finland	Periodic evaluations of regional and industrial policy measures undertaken during the 1980s and 1990s, especially interview surveys of companies receiving government subsidies.
France	Public policy tradition of evaluating ‘public actions’ but no known evaluation of regional policy measures during the 1980s and 1990s.
Germany	Evaluation is theoretically part of the regional policy framework but the system only involves tracking the outputs of regional aid and analysing the performance of eligible areas. There has only been intermittent analysis of the effectiveness of regional policy measures.
Greece	No regular evaluations of regional policy. Some official and academic evaluations of regional policy measures were conducted in the 1980s. Superseded since by emphasis on Structural Fund evaluations.
Ireland	Little tradition of evaluation, although some studies have been conducted since the 1970s involving a mix of quantitative modelling and survey methods. Increasing emphasis has been placed on Structural Fund evaluation during the 1990s largely through the modelling work of the Economic & Social Research Institute.
Italy	No tradition of regional policy evaluation apart from Structural Funds. However, the creation of evaluation units in all national and regional administrations is changing the picture.
Luxembourg	Infrequent regional policy evaluation linked to policy redesign phases.
Netherlands	Regular regional policy evaluation, undertaken every 4-5 years in line with reviews of regional policy legislation. Methods involve evaluation of efficiency and effectiveness through surveys and modelling.
Portugal	No tradition of regional policy evaluation apart from Structural Funds.
Spain	No tradition of regional policy evaluation apart from Structural Funds. Some regional policy evaluation in the 1970s using methods similar to those used in the UK at that time, but practice ceased.
Sweden	Tradition of regular regional policy evaluation, mostly through surveys of firms receiving government subsidies.
UK	Continuing tradition of regular regional policy evaluation - very active from late 1970s to mid 1990s. Methods involve evaluation of efficiency and effectiveness of measures through mix of surveys and modelling.

To provide an indication of the overall attitude to evaluation in each Member State, Table 4.1 reviews the diversity of evaluation practice with respect to national regional aid schemes (financial incentives) and how this has been changing in recent years.

The above table illustrates how attitudes to evaluation vary across EU Member States. In some countries, evaluation may be deeply ingrained within the administrative culture, in others, it is only rooted within particular departments or policy areas, and in a few cases it is viewed almost universally as having little significance. The extent to which evaluation is embedded in a wider tradition of policy assessments has been the key factor determining the attitudes to regional policy evaluation. Where they are not part of a wider evaluation environment, regional policy evaluations tend to be one-off studies, dictated by *ad hoc* political or policy requirements rather than part of a regular timetable of assessment. By contrast, in countries where evaluation has been regarded as a critical tool of policy making, the process of commissioning, conducting and responding to evaluations is more systemised and frequent.

In this context, EU countries can be loosely divided into different groupings, characterised by their strength of commitment to evaluation. In some cases, the similarity of their attitudes to evaluation reflects common pressures and policy backgrounds (such as the need to reduce budgets and target spending); in other cases, it represents the same responses to quite different policy environments. The three groups are: (i) those countries viewing evaluation as an institutionalised part of policy making; (ii) those where evaluation is an occasional exercise; and (iii) those viewing evaluation as limited and largely irrelevant to mainstream policy practice.

The most positive attitudes towards evaluation tend to be found in north-western European countries. For example, in Germany, the Netherlands, Sweden and the UK, evaluation has been regarded as an important part of the policy culture, not just in regional policy departments, but across different levels of government. Similar attitudes – though more recently developed – can be found in Austria and Ireland. All share a relatively systematic approach to evaluation. Evaluation as a whole is a widely accepted part of the policy making process.

The UK is generally regarded as the best example of a well-developed evaluation culture. Policymakers recognise the importance of objective-setting, monitoring and evaluation and have invested heavily in evaluations of the efficiency and effectiveness of regional policy over a 20-year period.] In Germany also, the regional policy Framework Plan (*Rahmenplan*) explicitly acknowledges the need for a regular assessment of the ability of current regional measures to address policy priorities. The information produced by various ‘controls’ is meant to determine how far policy measures need to be revised, or indeed, whether they should be continued at all. Three types of ‘evaluation’ are used to address different aspects of policy, notably *Vollzugskontrolle* (governing individual recipient firm effects), *Wirkungskontrolle* (measuring the effectiveness of certain regional policy instruments) and *Zielerreichungskontrolle* (assessing overall regional economic performance and the need for public intervention). Dutch regional

policymakers also have a tradition of evaluating policy every 4-5 years in line with reviews of regional policy legislation.

In other countries, a commitment to evaluation is a more recent phenomenon, attributable in several cases to Structural Funds regulation compliance.¹³ For example, Austria and Ireland have both undertaken some policy evaluation in the past, but have developed more extensive and active approaches as a result of the need to comply with Commission regulations. This appears to have had 'knock-on' effects in terms of their attitudes to the evaluation of national regional policy measures. The same is becoming true for Italy.

In a second group of countries, evaluation is only used as part of extraordinary or intermittent policy reviews. This used to be the case in Ireland, before the more recent shift to regular programme evaluations. Countries such as Finland and Luxembourg link evaluation to infrequent overhauls of regional policy, often as part of wider consultation exercises. For example, the last major set of evaluations in Finland took place in 1985 as part of the then reappraisal of regional policy instruments; more have been commissioned recently as part of a similar re-assessment of regional policy measures. One-off evaluations may be inspired by specific changes in policy environment. Cutbacks in expenditure have increased the need for evaluations in some countries (eg. Luxembourg); in others, evaluation is viewed mainly in terms of fraud prevention, largely as part of the *monitoring* of regional policy measures to ensure applicant compliance with measure guidelines.

For a third group of countries, there appears to be little national evaluation of regional policy measures outside of the Structural Fund policy area. This applies to France¹⁴ and southern EU Member States which tended to have an extensive tradition of formal policy evaluation.

Policy makers in these countries exhibit considerable scepticism about the value of policy evaluations for measuring the effects of assistance on individual company performance or the wider economic impacts of measures. Officials cite methodological problems as the main reason for eschewing evaluation, particularly difficulties in isolating policy effects and determining the policy counterfactual. Hence, there is a tendency to rely on more informal means of defining and assessing policy needs, where the individual judgement of policy makers and their experience of the operation of different incentive schemes are critical.

4.2 The influence of EC evaluation obligations

The most important factor influencing trends in regional policy evaluation has been the European Structural Funds regulations. Guidelines for Structural Funds expenditure require evaluations of strategies and measures at different stages and place a high priority on improving the quality of evaluations. Although slow to respond across the EU as a whole, evaluations are becoming more commonplace in regional development policy because of the Structural

¹³ MEANS (1998) *Le développement de l'évaluation des interventions structurelles dans les Etats Membres*, C3E, Lyon.

¹⁴ Although, for example, requirements for the evaluation of the State-Region Planning Contracts have been increased for the 2000-06 period.

Fund regulations. While it is not always apparent that Member States are embracing evaluation with the same spirit – some view it as a technical requirement which must be met to satisfy the European Commission – it is clear that it has become a major force in changing attitudes to evaluation across the EU.

The change is particularly apparent in certain Member States. In countries where evaluation has not been deeply ingrained in regional development policy making in the past – such as Austria, France, Italy, Portugal and Spain – evaluations are being undertaken more frequently in order to comply with Structural Funds regulations. Indeed, in some of these countries, the only regional policy evaluations being conducted are those dictated by the Structural Funds. However, in countries like Austria and France, this has resulted in more pervasive changes in policy making and evaluation as a whole is being viewed more positively.

Changes are also evident in countries where evaluation has a long tradition. For example, Swedish evaluations are now being introduced at *ex ante* level for non-Structural Fund expenditure programmes in line with evaluation practice under EU programmes. In Ireland and Italy, it has prompted changes in the way evaluation is organised within government, particularly with the creation of new units to specialise in Structural Funds evaluations.

There are two clear reasons why the Structural Funds have had this impact on national evaluation. The first is the direct result of co-financing. Structural Funds monies often support existing national regional policy measures, which has led to their evaluation because of the presence of EU support. It is not surprising then that the results of such evaluations have fed through into the national policy making processes, although some countries – notably Germany – remain suspicious of an approach that blurs the role of different sources of expenditure.

Second, the example of Structural Funds evaluation has been influencing countries because of the sheer volume and complexity of evaluations that need to be conducted. In supplying the Structural Fund needs for *ex ante*, interim and *ex post* evaluations, many Member States have developed a pool of necessary skills in the conduct of evaluations and the effective use of evaluation results which might not have existed otherwise. In some cases, because of the administrative changes made to adapt to EU requirements, it has been relatively simple to incorporate regular evaluation as part of policy making because of Structural Funds regulations. Hence, just as other principles in the design and delivery of Structural Funds programmes have shaped national regional development thinking, so too are diverse national policy evaluation practices being influenced by the universal requirements of EU regional policy.

4.3 Evolution of Structural Fund evaluation in the EU¹⁵

Before the 1988 reform of the Structural Funds, it was generally recognised that the monitoring and control of European Community regional expenditure was inadequate. This was partly due to the primacy of the Member States in the field of regional development. ERDF was used to co-finance regional development projects, largely subsumed within national regional policy budgets. Evaluation research was poorly co-ordinated within the European Commission, with different Commission services determining their own priorities and methodologies. There were also conflicts with Member States that had different expectations and traditions of evaluation and contrasting methodological approaches. Monitoring and evaluation techniques varied greatly throughout the European Community, with little detailed guidance for the Member States.¹⁶

The 1988 reform led to the appraisal and assessment of European Union regional policy being progressively enhanced in scope, scale and rigour.¹⁷ The new programme-based approach was associated with a comprehensive monitoring and evaluation system. The regulations required a Monitoring Committee to be set up for each and stipulated that structural operations were to be assessed *ex ante* and *ex post* to highlight their impact on objectives and their effects on specific structural problems.

Over the 1989-93 programming period, structures for monitoring and evaluation began to be introduced by the Member States and the European Commission, but only slowly. Data was poor or non-existent, targets and indicators were largely absent, and monitoring systems (for physical indicators at least) were rudimentary and unsystematic, with major variations between regions.¹⁸ In some regions, monitoring was limited to purely financial indicators; more effective monitoring was sometimes seen as being, at best, of secondary importance.

Subsequently, Member States were required to provide more information on the current regional development situation, the impact and effectiveness of Community-financed operations undertaken in the 1989-93 programming period, and the anticipated objectives (quantified where possible) and impact of future operations. In order to upgrade the data available to evaluators, 'baselines' of the regional development situation were to be identified at the outset of the planning period against which the effect of subsequent operations could be assessed. The Commission established a co-ordinated approach to regional policy evaluation with a specialist evaluation unit to have lead management responsibility for evaluation of Structural Fund operations in

¹⁵ This review draws on text from Bachtler J and Taylor S (1999) *Objective 2: Experience, Lessons and Policy Implications*, Report to the Commission of the European Communities (DG XVI), EPRC, University of Strathclyde.

¹⁶ Bachtler J and Michie R (1995) A New Era in EU Regional Policy Evaluation? The Appraisal of the Structural Funds, *Regional Studies*, 29 (8), 745-751. Eskelinen H, Kokkonen M and Virkkala S (1997) Appraisal of the Finnish Objective 2 Programme: Reflections on the EU Approach to Regional Policy, *Regional Studies*, 31 (2), 167-172.

¹⁷ Bachtler J and Michie R (1993) The Restructuring of Regional Policy in the European Community, *Regional Studies*, 27 (8), 719-725.

¹⁸ Bachtler J and Michie R (1995) *op.cit.* Eskelinen H *et al* (1997) *op.cit.*

Objective 1 and 2 regions. The MEANS programme, *Evaluation Methods for Actions of a Structural Nature*, was launched with the task of developing a 'European evaluation culture' by improving the quality and utility of evaluations. In addition, exchange of experience among EC countries was organised by the Commission through several fora, including an evaluation working group comprising representatives of the Member States and periodic evaluation conferences¹⁹.

As a result of these efforts, there were signs of improvement by the second programming period, which started in 1994. Many programmes were judged as having satisfactory quality indicators, with the best programmes having indicators at three levels (programme, priority, measure). However, there were still important deficiencies: incomplete indicator lists; lack of quantification at global objective, priority and measure levels; missing targets for monitoring; and unclear assessment indicators.

A comprehensive review of 1994-96 SPDs for Objective 2 regions and those approved in 1995 for the new Member States²⁰ found that, while most made reference to envisaged positive employment effects, just under half had absent, unquantified or obviously incomplete employment indicators. Where targets were set, programmes usually failed to specify the type of impacts (eg. direct permanent jobs only or also indirect jobs and temporary employment generated in implementing measures). They also tended to omit timescales, preventing future evaluators from comparing performance meaningfully with the targets. There was a lack of precision and consistency in the terminology used to refer to the main types of possible outcome.

Over the 1994-96 period, a considerable amount of further development work was undertaken. This was associated with: clearer, quantified targets; monitoring and evaluation indicators at all levels of the programme, distinguishing between outputs and impacts; and more efficient monitoring and evaluation of organisational arrangements eg. management information systems. Commission services worked closely with programme managers and secretariats to improve the range and quality of indicators as well as monitoring systems. The MEANS programme issued a range of monitoring advice, leaflets and programme manager handbooks as well as providing support for evaluators. A series of high-profile evaluation conferences was organised by the Commission (in Brussels, Berlin and Seville) to promote the monitoring and evaluation 'message' and to encourage exchange of experience among regions.

In several countries, national government authorities were increasingly establishing an overall evaluation framework or common guidelines for evaluation. In the United Kingdom, the Scottish Office and the Department of the Environment formed a sub-group to produce a guidance note, *Framework for the Evaluation of Area-Based European Structural Funds Intervention*, for all government departments. A similar approach was adopted in France where DATAR produced a *Vade Mecum for the Structural Funds* backed up by

¹⁹ Bachtler J and Michie R (1995) *op.cit.*

²⁰ Bachtler J, Taylor S with Kearney C (1996) *Extended Synthesis of Agreed Single Programming Documents in Objective 2 Areas 1994-96*, Report to the Commission of the European Communities (DG XVI), EPRC, University of Strathclyde.

specific guidance notes on specific types of evaluation (agreed with the Commission) and held a series of seminars with programme managers. DATAR also encouraged regional managers and partnerships to build links with the long-standing government regional evaluation units which provide substantial expertise and resources for evaluating government policies.

The provision of central support and resources was taken a stage further in Austria. The *Österreichische Raumordnungskonferenz (ÖROK)* and Federal Chancellery invested heavily in promoting evaluation awareness and interest since evaluation did not have an established culture or tradition in Austria. All Structural Fund evaluations were co-ordinated by ÖROK and involved a range of federal and regional interest groups; exchange of experience meetings between the evaluators for Objective 2 and other objectives were organised, and a central resource, 'Checkpoint EVA', was created by the Federal Chancellery to promote new thinking in the field of evaluation.

At the mid point of the 1994-99 period, research by the European Commission on the Objective 1 and 6 programmes showed that monitoring and evaluation practices were improving²¹. There was evidence, for example, that some countries were fixing clearer quantified targets and indicators for monitoring and evaluation eg. Austria, Portugal, Italy and the United Kingdom. It was also noted that evaluation was increasingly being seen as a management tool, capable of assisting policy makers and programme managers with insights into the effectiveness of measures and programmes. A further trend during this period was significant investment in the modelling of Structural Fund impacts, especially in the cohesion countries. Among the most prominent examples have been the Beutel dynamic input-output analysis model, the HERMES and HERMIN models and the econometric PARADISE model, all described in more detail in Section 6. Nevertheless, it was equally clear that progress was not universal and that there was still a widespread need: to develop new sets of indicators to enable better measurement of programme results and impacts; to improve integration between financial and physical indicators; and to ensure more quantification of baseline and target indicators and provide benchmark data to allow for better comparison within and between programmes.

The evidence for Objective 2 regions was similar. The reprogramming of Objective 2 in 1997 showed that in the more advanced cases (the longest-running programmes), there were ambitions of creating fully integrated physical and financial reporting systems, and comprehensive monitoring and evaluation frameworks. Also, regions increasingly accepted the value of monitoring and evaluation exercises. A review of the 1997-99 Objective 2 programmes carried out by the Commission found that quantification of effects for the 1997-99 programmes had improved significantly, 55 of the new SPDs giving detailed measure-level information.²²

Several increasingly common good practice features were evident. First, more programmes were specifying what was included in employment targets,

²¹ CEC (1998) *Mid-Term Review of Structural Interventions Objectives 1 and 6 (1994-99)*, DGXVI, Commission of the European Communities, Brussels.

²² CEC (1997) *The New Regional Programmes 1997-99 under Objective 2 of the Community's Structural Policies – Focusing on Job Creation*, COM (97) 524 final, Commission of the European Communities, Brussels.

detailing some of the assumptions on which these targets were based, and indicating the timescales over which they should be realised. Second, the type and quality of jobs created was specified with increasing care. There was greater differentiation between ‘temporary’ employment generated by implementing projects (eg. in construction, often referred to in terms of ‘person years of work generated’) and ‘permanent’ posts (eg. created in the longer term by the use of infrastructure). Also, there was more specification of the nature of the jobs created (full-time, part-time, seasonal, etc). Third, there was increasing precision in the use of employment-related terminology and concepts, indicating greater familiarisation with the field and increased reflection. Definitions were used which more accurately reflected the true nature of employment impacts. No longer were people entering employment as a result of enhanced skills described as ‘jobs created’ (except where they entered self-employment).

In spite of improvements, these reviews make clear that experience across programmes is still mixed. Several of the above observations relating to the position in 1994 still apply eg. inadequate indicator information, lack of appropriate and practical targets, and lack of consistent and regular data collection systems. Employment targets still often only encompass direct effects and are expressed in gross terms. The negative effects of deadweight and displacement and positive multiplier effects are not yet being taken sufficiently into account. In addition, the assumptions on which targets are based are still not consistently set out. These features are symptomatic of the virtual absence, in many cases, of robust methodological approaches to quantifying programme outputs and impacts.

Examining the present situation concerning commitment to Structural Fund evaluation among EU Member States, a subjective assessment is provided in Table 4.2. This indicates the perceived level of commitment in each country, whether a national evaluation strategy or framework is in place, and whether the evaluation of the Structural Funds is primarily externally driven - to meet the regulatory requirements - or whether it is also strongly driven by the domestic agenda and objectives. According to this classification, Member States can be divided into three groups:

- strong commitment to evaluation (Denmark, Ireland, Netherlands, UK)
- evolving commitment to evaluation (Austria, Belgium, Finland, France, Greece, Italy, Luxembourg, Portugal, Sweden)
- variable commitment to evaluation (Germany, Spain) ie. major differences between regions and limited drive from the national level

First, there are interesting differences between countries in the degree to which there is a co-ordinated approach to Structural Fund evaluation. In Member States such as Sweden, France, Austria, Ireland, the UK and the Netherlands, attempts have been made to establish a co-ordinated approach to evaluation – in several countries with an explicit framework or strategy – with national guidance on terminology, evaluation processes or methods. In the Dutch case, the country has a decentralised approach to programming, and the national evaluation framework was established to ensure that a ‘common language’ was in place to enable comparisons to be made between evaluation studies.

Table 4.2: Commitment to Structural Fund evaluation in EU Member States

Member State	Commitment	Strategy	Comment
Austria	Evolving	Yes	Evaluation generally restricted to regulatory requirements. However, it is gradually being seen less as an obligation and valued as a management tool and integrated in the programming process.
Belgium	Evolving	No	Evaluation goes beyond regulatory requirements. Extensive investment in interactive monitoring system for programme management and modelling for impact assessment.
Denmark	Strong	No	Evaluation goes beyond the regulatory requirements. Studies have been conducted independently to improve the effectiveness of the programmes.
Finland	Evolving	No	Evaluation generally restricted to regulatory requirements, but active efforts to improve quality and relevance.
France	Variable	Partial	Evaluation organised to meet regulatory requirements, but examples of regions using studies also to address programming issues.
Germany	Variable	No	Evaluation organised to meet regulatory requirements, but examples of regions using studies also to address programming issues.
Greece	Evolving	No	Evaluation generally restricted to regulatory requirements.
Ireland	Strong	Yes	Evaluation goes beyond regulatory requirements. Institutionalised management of evaluation, and extensive investment in modelling for impact assessment.
Italy	Evolving	No	Evaluation generally restricted to regulatory requirements.
Luxembourg	Evolving	No	Meet regulatory requirements.
Netherlands	Strong	Yes	Evaluation generally restricted to regulatory requirements. Effort to establish national frameworks.
Portugal	Evolving	No	Evaluation generally restricted to regulatory requirements.
Spain	Variable	No	Evaluation generally restricted to regulatory requirements.
Sweden	Evolving	No	Evaluation generally restricted to regulatory requirements, but pro-active responses.
United Kingdom	Strong	Yes	Evaluation organised to meet regulatory requirements, but regions frequently use studies to address programming issues.

Despite co-ordination, not all regions necessarily comply fully with the guidance (for example in France) leading to regional variations in the approach to evaluation and the quality of studies.

By contrast, in other Member States (eg. Finland, Germany, Spain), evaluation is not co-ordinated nationally. In the case of Finland, this reflects the lack of communication and co-ordination between national government departments on evaluation matters and the split in competence for Structural Fund administration at regional level (between State district offices and regional authorities). In Germany, as a federal country, fulfilment of the evaluation obligation is the responsibility of the *Länder* and the federal government only has an oversight function; the national ‘hands off’ attitude perhaps also reflects a high level of scepticism about the value and purpose of Structural Fund evaluation, at least outside the Objective 1 regions. There is considerable regional variation between the German states, with major investment in evaluation in the east German *Länder* and less commitment in the West German states – with some notable exceptions (eg. North-Rhine Westphalia, Bremen)

Second, several of the Member States in the above table are described as having an ‘evolving’ or ‘variable’ commitment to Structural Fund evaluation. In some cases (Austria, Finland, Sweden) this reflects the fact that Structural Fund programming is still relatively new; during the first programming period following the accession of these countries (1995-99), there was a process of institutional adaptation to the novel management and administrative requirements of the Funds. As these have become familiar, evaluation has slowly come to be seen less as an obligation and more as a programming tool, reflected (in Austria, in particular) in investment in monitoring and evaluation systems.

In other countries (Belgium, France, Greece, Italy, Portugal, Spain), the evolutionary commitment to Structural Fund evaluation is associated with the complex inter-relationship between national and EU evaluation traditions/expectations. In these countries, evaluation has had (or will have) to be embedded in the national public policy system for Structural Fund evaluation to be taken seriously. As noted earlier and in the next section, changes to the evaluation framework in Ireland have contributed to more integrated evaluation practice, notably the requirement that all departments have meaningful objectives, three-year spending reviews for all departments, performance management and publicity for evaluation studies. Partly driven by the Structural Fund experience, this national approach has also served to increase the resources and commitment for Structural Fund evaluation.

A similar process is taking place in Italy where recent legislation (1999) has begun a process of institutionalising evaluation units within national and regional administrations. The hierarchical approach is beginning to deliver a consistent framework for Structural Fund evaluation, although there is still wide regional variation in compliance. The French commitment to evaluation dates back further, and was significantly strengthened in a 1992 law underlining a requirement for the regional evaluation of all public investments. A ‘Scientific Evaluation Council’ has also been established there to improve evaluation practice. While this has helped to raise the profile of evaluation

and focus efforts, some regions have been reluctant to respond – an experience mirrored in the approach to Structural Fund evaluation. Lastly, in Spain and Portugal, historically there has been no strong culture or tradition of evaluation and the Structural Funds have been a catalyst for the re-emergence of regional policy evaluation - but this time of a new type, involving the assessment of complex programmes involving multiple actors. For many in the national and regional administrations, the evaluations undertaken because of the Structural Fund regulatory requirements were their first experience of evaluation.

4.4 Assessment

Evaluation experience has historically varied greatly across the EU. In some Member States, evaluation has been deeply ingrained within the administrative culture, so much so that it has been systematically integrated into the policymaking process. Within this category, the UK has generally been regarded as the best example of a well-developed evaluation culture, especially with respect to regional policy: over the last two decades, great emphasis has consistently been placed on how objectives and monitoring systems are set and how evaluations should be managed. Similar commitments to domestic policy evaluation can be seen in Germany, the Netherlands, Sweden and more recently, Austria and Ireland. Other countries – notably Finland and Luxembourg - have operated a more *ad hoc* approach to evaluation, where it has been undertaken on an occasional basis by particular departments or for individual policy areas. Lastly, countries such as France and the southern Member States have very limited evaluation traditions and have tended to view evaluation as having little policy relevance.

This diverse experience has been transformed by the evaluation requirements of the Structural Funds. As a result of the co-financing of programmes and the scale and complexity of the evaluations demanded by EU guidelines, Member States have had to develop an evaluation capacity, not just for the Structural Funds, but for domestic policy areas as well. The ‘Trojan Horse’ effect has resulted in the introduction of the first systematic policy evaluations in France, Portugal and Spain and the development of sophisticated institutional approaches to evaluation in Austria, Ireland and Italy (from which Scotland can draw good practice, as will be shown in later sections).

Nevertheless, Member States continue to display a wide range of policy approaches to evaluation (as well as commitment to its performance, particularly at regional level). Over the last decade, there has been a steady improvement of evaluation practice across the EU, but not a harmonisation of organisational or methodological approaches to the design, delivery and use of Structural Fund evaluations. No one Member State stands out as an exemplar of all aspects of evaluation – though the UK combines good practice in more areas than most – but a degree of experimentation has resulted in the proliferation of good (and bad) practice.

5. INTERNATIONAL REVIEW AND CASE STUDIES: MANAGEMENT AND ORGANISATION OF EVALUATION

The earlier critique of Structural Fund evaluation in Scotland (Section 3) noted that there is a relatively robust framework in place for evaluating Structural Fund programmes which largely delivers what is required, to specification and on time. Improvements to the quantification of programmes have significantly improved the consistency of data available, while Scottish Office initiatives to encourage a more structured and ongoing response to the evaluation requirement have improved the quality and consistency of evaluations. In turn, evaluation has come to be increasingly valued as a useful and important programming tool, particularly for process issues.

Nonetheless, it was also suggested in Section 3 that the management of Structural Fund evaluation in Scotland could benefit from further consideration of several issues, among them:

- organisation of the evaluation of Structural Fund co-financed activity within a wider framework of economic development policy evaluation;
- improving the capacities and skills of PMEs and partners (including a 'gearing up' to respond to new responsibility for European Social Fund evaluation), and widening the evaluation community;
- additional thought about the organisation of future evaluations, their timing, targeting and inter-relationship with each other, including continued and potentially more strategic use of thematic evaluations;
- a more explicit partnership approach to evaluation, recognising and integrating the priorities and objectives of different stakeholders; and
- strengthening arrangements for the validation, dissemination and exploitation of evaluation findings.

Following on from Section 4 which provided a wider context by examining how the status of evaluation has evolved over the last decade across the EU 15, this section reviews international experience with the management and organisation of Structural Fund evaluations in more detail, also presenting case studies of international practices which could help to address the monitoring and evaluation issues facing the Scottish Structural Fund community at the start of the 2000-06 programming period.²³

5.1 Evaluation management

The first key question is how evaluation can be optimally managed. Although there is a robust framework in place in Scotland, Structural Fund evaluation takes place largely in isolation. The resourcing of Structural Fund evaluation in the Scottish Executive is limited, and there appears to be little explicit or formalised institutional co-ordination or exchange with the evaluation of

²³ As noted in Case Study 9, a detailed, step-by-step guide to the conduct of evaluation is beyond the scope of this study, and would duplicate the work of a variety of detailed projects including MEANS initiatives and, for example, publications such as *Improving Evaluation Practices: Best Practice Guidelines for Evaluation and Background Paper*, published under the OECD PUMA Programme, Paris, 1999.

domestic programmes. This is part of a wider lack of connection between the assessment of domestic and EU interventions in Scotland. Structural Fund programmes are delivered only in parts of Scotland and represent one component of a wider economic development picture, but there is - as yet - no framework to measure the economic development impacts of Structural Fund programmes as part of a wider economic development context.

This is a common problem. Across the EU, it has been recognised that it is difficult to address effects of Structural Fund interventions in isolation, partly because of the methodological problems involved (see section 6) and partly because of institutional constraints: Structural Fund evaluations are just one of many demands on technical assistance resources, and are generally not sufficiently resourced to address all programme dimensions in detail.²⁴ Arguably, there are advantages in working ‘holistically’ to identify the impact of overall economic development activity, or of all activity in a given sphere (eg. business development) rather than trying to ring-fence the diverse activities of a Structural Fund programme.

The foundation for such an approach is to have an adequately resourced institutional arrangement for evaluation which: (a) encompasses the main departments and agencies involved in economic development, at national, regional and local level; (b) combines monitoring and evaluation resources and expertise across EU and domestic programmes; and (c) has an integrated programme of evaluation that seeks to make a real contribution to understanding the efficiency and effectiveness of policy interventions. One of the best examples of an all-encompassing institutional arrangement for EU policy evaluation of this kind is in **Ireland** – see **Case Study 1**. A further, more recent example is in **Italy** – see **Case Study 2**.

Of course, there is greater potential to seek coherence with parallel frameworks and activities when these other activities are subject to a similar evaluation regime, which is by no means always the case. Where there is a strong national context for evaluation, as in Ireland and France, this may be more likely. In Ireland, changes to the national context for evaluation which took place in 1997 have strengthened the emphasis on integrated evaluation practice, among them the requirement that: (i) all departments' strategies have meaningful objectives; (ii) three-year spending reviews take place for all departments; (iii) performance management be taken up; and (iv) citizens have access to evaluation documents.²⁵ In France, the formal emphasis placed on evaluation of the *Contrats de Plan État-Région* has increased, leading to a greater coherence between national requirements and those imposed by the EC for Structural Fund programmes. Similar potential exists in England, given the

²⁴ There have been notable exceptions, eg the Yorkshire and Humberside and Nordrhein Westfalen evaluation choices described in case studies 4 and 5.

Comments in this report about the resources dedicated to evaluation are based on qualitative information derived from written commentary and from the interview programme. It has not been possible to identify accurately either the direct or indirect resources dedicated to Structural Fund evaluation in different Member States. In terms of personnel inputs, in particular, this is because many Member States do not have permanent personnel dedicated to Structural Fund evaluation, this function usually being combined with a wider range of Structural Fund co-ordination tasks or with domestic evaluation responsibilities.

²⁵ Means Internal Bulletin, February 1998, C3E, Lyon, p4.

inter-linkage and similarities between EU programmes and the Single Regeneration Budget.

CASE STUDY 1: INSTITUTIONAL ARRANGEMENTS FOR MANAGING EVALUATION - IRELAND

Ireland is distinctive among EU Member States, in that it has institutionalised the evaluation of structural interventions in the Irish economy.²⁶ To date, four Evaluation Units are operating.

- The **ESF Programme Evaluation Unit** was created in 1992 by the Department of Enterprise, Trade and Employment in partnership with the European Commission. It undertakes: evaluations of individual vocational education, training and employment programmes; thematic evaluations across programmes and agencies; and survey analysis. It reports jointly to the Department of Enterprise and Employment, where it is located, and to the EC.
- The **Analysis and Evaluation Unit** for the Agriculture, Rural Development and Forestry Operational Programme operates within the Department for Agriculture in partnership with the Commission. It undertakes evaluations of the agricultural OP under the aegis of its Monitoring Committee.
- The **Industry Evaluation Unit (IEU)** operates within the Department of Enterprise, Trade and Employment. It is a quasi-autonomous Unit jointly funded by the Department of Enterprise, Trade and Employment and the European Commission. It undertakes evaluations of the measures and sub-programmes contained within the OP and reports to the Industry OP Monitoring Committee. The Unit also manages the Retex Community Initiative.
- The **CSF Evaluation Unit** was established in 1996 by the Department of Finance in partnership with the European Commission. It is co-financed by the Structural Funds under the CSF Technical Assistance Operational Programme and reports to the Monitoring Committee for the Operational Programme. The Unit's central function is to advise and assist the national authorities and the European Commission in the evaluation of Structural Fund programmes, but its work also includes the co-ordination and promotion of best practice in Structural Fund evaluation work.

These Units '*provide a high professional standard and a valuable resource to Monitoring Committees on an on-going basis*'.²⁷ Given the extent of EU co-financing in Ireland, the Evaluation Units have to date effectively evaluated EU and national programmes together.

²⁶ CEC, *Better Management through Evaluation. Mid-term Review of Structural Fund Programmes: Objectives 1 and 6 (1994-99)*, p. 15.

²⁷ Honohan P (Ed.) (1997), *EU Structural Funds in Ireland. A Mid-Term Evaluation of the CSF 1994-99*, ESRI, Dublin, p. 194.

The mid-term evaluation of the 1994-99 CSF, highlighted some elements of the Units that could be improved, in particular the perceived excessive breadth of work programmes. A more focused approach, concentrating the limited resources available on achieving improvements in the implementation and impact of the CSF was judged to be more appropriate. The report suggested that the Units should:

- address operational issues rather than policy or thematic matters, which require varying skills and know-how, in order to improve implementation efficiencies and the impact of the programmes;
- conduct statistical and other quantitative analysis on a measure-by-measure basis, aimed at measuring performance and impacts and in order to establish cause and effect relationships; and
- establish and implement a clear and explicit process to assist the review and revision of targets and objectives on a regular basis.

Relevance to Scotland

The core rationale for the Irish structures of course lay in the significant Structural Fund resources allocated to that country, and the fact that being entirely eligible for Objective 1 until 1999, there was a strong inter-linking of Structural Fund and domestic development policies. It is unlikely that such a complex structure would be required in Scotland. However, while the Structural Funds are only a part of the budget allocated to economic and social development policies, it is possible that a more modest Evaluation Unit in the Scottish Executive, with the task of co-ordinating evaluation efforts across Structural Fund and other economic and social development policies could be beneficial, bringing increased consistency and a higher profile to evaluation and ensuring the generation of strategically useful information, enabling the overall impact of economic development efforts in Scotland to be better understood and policy to be better targeted. It would need to work collaboratively alongside other Scottish institutions including the Enterprise network, and could focus, for example, on methodological inputs, statistical support, coherence issues and synergies.

CASE STUDY 2: INSTITUTIONAL ARRANGEMENTS FOR MANAGING EVALUATION - ITALY

In Italy, recent legislation (Law 144/1999 - collegato alla finanziaria) requires all public administrations (national and regional) to establish Units for the Evaluation and Verification of Public Investments. These will operate in co-ordination with the parallel, pre-existing National Evaluation Unit within the Ministry of Treasury and Budget.

The National Unit has operated since 1980 and now has a staff of 30. Its activities relate to all public investments and include, amongst other things, the development and diffusion of methods for evaluating Structural Fund programmes and ensuring that these are applied in Structural Fund programming. The Unit also undertakes evaluations and had a leading role in

the *ex ante* evaluation of the Objective 1 CSF, developing an *ad hoc* supply-side macro-economic model and providing guidance to other relevant administrations. For the CSF, in particular, the Unit has a strategic and co-ordination role – in collaboration with the national competent administrations for ESF and EAGGF and with the working group ‘Evaluation and monitoring that will be constituted within the CSF MC. The aim of the Evaluation and Monitoring working group is to:²⁸

- launch the network of evaluation Units foreseen by law 144/99 by the end of 2001 (with reference to the administrations involved in the management and implementation of the CSF);
- in the context of the evaluation of the CSF, define and apply methodologies and techniques for the assessment of results and impacts;
- ensure homogeneity and consistency in the evaluations undertaken at different levels and assess the management and continuing relevance of monitoring indicators;
- undertake ‘second degree evaluation’, addressing the quality of the interim and thematic evaluations that will be undertaken for the CSF and the various OPs, also with the aim of providing exchange of good practice; and
- promote and provide methodological guidance to thematic evaluations, for the *in itinere* assessment of large projects, Integrated Territorial Projects and other relevant projects. These thematic studies are conceived as instruments to improve the implementation performance of programmes and to engender an evaluation culture.

Law 144/1999 strengthens the role of the evaluation Unit, complementing it with similar units in every administration. Furthermore, the law foresees a tight deadline and the provision of resources - about €4.1 million for 1999 and circa €5.2 million for 2000. The new Units operate to:

- provide technical support to the planning and evaluation of public investments, with particular attention to the issues of sustainability and quality of the environment;
- contribute to a new monitoring system for public investments (see below); and
- ensure the gradual extension of effective Structural Fund techniques, in relation to the programming, evaluation, monitoring and control phases, to all programmes and projects implemented at the territorial level.

The law also foresees the creation of a new Monitoring System for Public Investments, whose function will be to provide up-to-date information on the implementation of all development policies, with particular reference to Structural Fund programmes. The new Monitoring System (MIP – *Monitoraggio degli Investimenti Pubblici*) will be overseen at the CIPE (Inter-ministerial Committee for Economic Programming) and will relate to the

²⁸ Ministry of Treasury, Budget and Economic Planning (2000), *CSF for the Italian Objective 1 2000-2006*, Ministry of Treasury, Budget and Economic Planning – Department for Development and Cohesion Policies, Rome, Italy

procedural, physical and financial progress of the interventions. The MIP will also generate a database that will be maintained by CIPE.

Relevance to Scotland

As described in Case Study 1, the creation of an Evaluation Unit for both Structural Fund and domestic policies would contribute to building a common evaluation framework and to delivering the idea of evaluation as a continuous exercise.

The second requirement highlighted above is that there should be an integrated sharing of resources and expertise for monitoring and evaluation across domestic and EU programmes. This would allow exchange of expertise as well as a common approach to monitoring and evaluation. There are few examples of evaluations of small Structural Fund programmes which try to take account of, or interact with, the wider economic development evaluation context. Among them are initiatives at national level in France, and at regional level in the **West Midlands** (see **Case Study 3**) and Bremen.

The French example relates to a new computerised monitoring system (*Présage*) which is being introduced nation-wide in France. This has scope for regions to record not only Structural Fund programme activities but also the activities supported under the contractual expenditure agreements between the State and regional councils - the *Contrats de Plan État-Région* - even where these projects are not co-financed by the Structural Funds. If a region chooses to input both Structural Fund-supported projects and wider *Contrat de Plan* activity - and at this stage it is very much up to them to decide - then this will potentially provide them with a more powerful evaluation tool, enabling evaluations to take a wider scope and a broader overview to be easily compiled. Of course, it remains a regulatory requirement for the effects of the Structural Funds to be separated out from the bigger picture. To facilitate this, the project information which is input includes an indication of the sources of funding used and the sums involved. This enables projects funded under different programmes to be separated out rapidly and easily, and an estimate to be made of the relative contribution of different funds to the impacts achieved.

CASE STUDY 3: A 'NESTED' APPROACH TO MONITORING AND EVALUATION - WEST MIDLANDS

The Government Office for the West Midlands (GOWM) has a good track record for the monitoring and evaluation of Structural Fund programmes. The new regional development agency, Advantage West Midlands (AWM), has been keen to exploit the Structural Fund expertise of GOWM to develop a **region-wide, comprehensive monitoring and evaluation framework** encompassing all interventions.

This initiative is still in its early days, but the AWM has developed a basic 'AWM framework' for the region, within which is a more sophisticated framework for the Objective 2 programme (which, in reality, is driving the AWM framework). AWM are encouraging colleagues in the region to develop their own 'nested' systems within the AWM framework. In the first instance, the intention is to have commonality, from the top downwards in:

- definitions of outputs,
- the data used, and
- how the data is collected.

It is anticipated that this will make it easier to disentangle the outputs associated with different policy interventions. The ultimate aim is to get **real impact evaluation**, and to separate out the impacts of parallel socio-economic interventions such as the Structural Funds and Single Regeneration Budget. However, the proposal is still in its early days and a long way short of this goal. For example, as yet there is no input-output framework for the West Midlands. Nevertheless, over the next seven years, the framework should certainly deliver what is needed for the management and delivery of the Objective 2 programme and improve the quality of monitoring and evaluation.

The work on developing the AWM framework, which is being subsidised by technical assistance, has been driven by AWM and GOWM. There is a partnership consultation group and they have consulted closely with other partners on what should be covered by the system. Development has been undertaken by the consultancy firm SQW based on EC work.²⁹

There is as yet no compulsion for other organisations to fit into this framework; AWM is not able to exert sanctions on partner organisations. Instead, it is believed that by demonstrating the utility of the system, it will progressively be adopted actively. The emphasis has been on joint working, and the fact that the work has been undertaken with a partnership consultation group and involving extensive dialogue with other partners should help to ensure that it is actively taken up.

Relevance to Scotland

This example demonstrates how an agency operating at regional level has taken pragmatic steps to establish an integrated evaluation framework in a context of partnership in order to be able to establish an overall picture of main economic development impacts at regional level. This initiative has been pursued at least in part because it is seen as politically important, in particular for strategic policy design.

²⁹ Including: CEC (1997) *Guide to Cost-Benefit Analysis of Major Projects In the Context of EC Regional Policy*, DGXVI, Commission of the European Communities, Brussels; and Ernst & Young (1995) *Appraisal and Evaluation Methods for Assessing Net Value Added*, Department for Education and Employment, Sheffield.

The AWM approach could also be feasible in Scotland. As in the West Midlands, Structural Fund interventions represent just a small part of the local development framework, and there is clear political interest in capturing the bigger picture. Furthermore, in Scotland an Input-Output framework already exists, which would allow impacts to be addressed.

The Bremen case involves Structural Fund evaluators exploiting the results of other policy evaluation studies which overlap with Structural Fund programming activities. The Structural Fund Co-ordination Unit in Bremen has made it a priority to make an input to (and exploit the outputs of) the studies produced by other departments to support its own evaluation efforts (eg. with respect to the analysis of the environmental impacts of policy). Interview evidence indicates that this has been particularly useful in enabling complex areas to be addressed in more detail and more cost-effectively than the 'global' evaluation of a small Structural Fund programme could deliver on its own.

5.2 Integrated programmes of evaluation

A further issue cited above is the need to have an integrated programme of evaluation. In general, Scotland – like most other parts of the EU – has tended to respond to the EU regulatory requirements with a relatively *ad hoc* approach to evaluation - in spite of the concept of rolling evaluation. *Ex ante*, interim or *ex post* evaluation studies are commissioned as they are required in the programming cycle, and there is often insufficient inter-relationship between them, or systematic follow-up on issues raised by the programme evaluations. Where specific issues are investigated further, it has tended to be at the initiative of the programme managers, although the Scottish Office has also initiated some thematic studies. With some exceptions, the 'global' evaluation of whole programmes has been the main approach rather than using specific well-developed methods for discrete programme elements.

Different approaches have been used by EU countries/regions: some have restricted themselves to undertaking separate evaluations for each individual programme; others have initiated multi-programme studies for groups of programmes; and others have used a combination of programme-specific and thematic evaluations.

Austria and Denmark provide examples of a 'multi-programme' approach. In Austria, the four Objective 2 programmes were evaluated as part of a single, multi-programme study, providing an opportunity for a standard methodology to be employed across all four regions and for commonalities and contrasts in programme performance to be identified. A combined approach, as used in Denmark for Objective 2, is to undertake joint evaluations of common core elements, with additional thematic studies at programme level addressing programme-specific issues (eg. SME development and internationalisation in North Jutland, which were subjects chosen by the programme-specific steering group).

While the Scottish approach involves commissioning programme-specific evaluations, what has effectively been commissioned in some cases is multi-programme evaluations: the same evaluators are contracted to undertake studies for more than one Scottish programme at the same time, using methodologies which are almost identical. This practice is similar to multi-programme evaluation, but it fails to deliver the potential benefits of true multi-programme evaluation, or of combined multi-programme and single programme evaluation, as set out below in Table 5.1, because, formally and contractually, the various studies are separate. If programmes believed there could be benefits from the joint commissioning of evaluations, this could easily be undertaken. Less formally, programmes using the same evaluator could propose a session of exchange for comparison of their results.

Table 5.1: Relative merits of single and multi-programme evaluations

Multi-programme evaluation	Single programme evaluation
<ul style="list-style-type: none"> ➤ Methodological unity across different evaluations enabling comparison of results. ➤ Potential cost benefits - the consultants develop one set of questionnaires and other frameworks and apply them. ➤ Potential added value – consultants get to know several programmes in depth and can draw more accurate and useful comparisons based on their comprehensive knowledge. 	<ul style="list-style-type: none"> ➤ The methodology is a response to a specific programme and so may address more of the important issues more usefully. ➤ The timescale can be shorter than when the same evaluation team is undertaking multiple studies, and the evaluator may have more time to devote to developing the necessary relationships with programme actors

One difficulty in evaluation is to ensure that the studies carried out are able to capture the main relevant impacts and trends which arise from the wide range of activity which Structural Fund programmes typically support. In 'global' evaluations – ie. of a whole programme - evaluations are limited to reflecting a very limited range of impacts which are relevant to most interventions. Frequently, the focus is on employment outcomes, as with most Scottish evaluations. In addition, a problem in the 'global' evaluation of multi-fund programmes is that the dominant Fund receives most attention, with only cursory attention being given to the other(s), which may require different methodologies and significant resources to be successfully addressed (eg ESF evaluation in Objective 2 studies).

One potential solution is to conceive evaluation studies as a series of meso-level evaluations of programme elements rather than as a single global evaluation of a whole programme. The practice promises a significant improvement in evaluation quality, improving insights into the functioning and impacts of the programme, but requires considerably greater inputs, not only in terms of financial outlay, but they also have a potentially longer duration, greater complexity, a greater need for wider range of methodological responses, and a greater requirement for thematic specialists rather than

generalists to deliver insights in complex areas such as RTDI or community economic development.

Among the examples of programme evaluations undertaken as multiple meso-level studies are the evaluation of the Finnish Objective 3 and 4 programmes from 1994-99, and, from the UK, the major **interim evaluation of the Yorkshire and Humberside 1994-96 Objective 2 programme** (see **Case Study 4**).

CASE STUDY 4: MESO-LEVEL PROGRAMME EVALUATION YORKSHIRE AND HUMBERSIDE

Evaluation of the Yorkshire and Humberside Objective 2 programme for 1994-96 was distinctive for several reasons: not only did it combine quantitative impact measures with qualitative assessments of programme operation, but it did so using a series of inter-linked smaller 'meso-level' studies. The results of one set of studies were fed into a higher 'tier' of studies, ultimately fitting together into an evaluation hierarchy which covered geographical and thematic aspects of the programme.

The original tender for the evaluation specified three separate sets of interlocking evaluation activities: assessments of the three Sub-Regional Partnerships making up the programme area; a cross-regional appraisal of several thematic issues with a view to drawing lessons for the content and delivery of the 1997-99 Objective 2 programme for the area; and a 'quality assurance' service, which would not only provide monitoring of the evaluation process but also place the experience of the region's programme in a national context. Completed in October 1997, the final report elaborated on these different strands by using a tiered approach to evaluation. These can be described in two stages: the sub-regional analyses; and the Programme Wide Appraisal.

Stage 1: Sub-regional analyses

Under EKOS co-ordination and completed by Easter 1997, the sub-regional analyses constituted the major programme-wide data-gathering stage of the evaluation. Examining separately the programme partnerships of South Yorkshire, West Yorkshire and the Humber, the analyses focused on the operation of the locally-based partnerships, development of methodologies for better output measurement and identification of key, programme-wide issues for further investigation in the second stage of evaluation (while noting distinctive, sub-regional programme concerns). This was largely undertaken on the basis of interviews with the main participants in the partnerships and a review of their individual systems for gathering outputs from awarded projects.

Stage 2: Programme wide appraisal

The programme wide appraisal was awarded in early 1997 to a university-based consortium, led by the Centre for Training Policy Studies at the University of Sheffield and including staff and research units at Sheffield Hallam University and the universities of Bradford and Hull.

Using the sub-regional analyses, different teams undertook thematic assessments of different aspects of the programme, sometimes introducing another tier of analysis within the evaluation (as described below in relation to the employment studies). For the most part, these assessments were based on a combination of desk research (involving the sub-regional analyses and programme output and monitoring data), interviews with officials in the Programme Monitoring Committee, Government Office and other agencies, and telephone and face-to-face interviews with a sample of project participants.

The programme wide appraisal considered two sets of effects: efficiency in programme delivery; and project impacts.

(i) Programme delivery issues

Administrative issues in the programme were subject to several self-contained studies, designed to address the main issues of concern identified by the programme officials in connection with the forthcoming 1997-99 programme.

Partnership capacity. Largely using the sub-regional analyses and supplementary interviews, the task of this study was to evaluate the partnership organisations set up to implement the programme, identifying the factors affecting their performance and suggesting improvements for the next programme.

Quality and relevance of economic assessments. A desk-based review was undertaken of the quality and coverage of the data in the 1994-96 economic assessments produced for the three Objective 2 sub-regions, with particular emphasis on the extent to which these assessments could be used to aggregate impacts for the programme area as a whole.

Programme implementation. The detailed study focused on the effectiveness of the monitoring and implementation arrangements for the programme (including project selection methods and project management systems), reviewing progress in achieving programme-wide goals and assessing the coherence and level of integration between ERDF and ESF funding in the programme.

Programme impact

While noting that it was still too early for a comprehensive assessment of the programme's economic effects, a synthesis study was made of the actual and sustainable employment impacts of the programme. This was based on a 'sub-tier' of employment studies which evaluated job creation and safeguarding of the different sets of measures collected in the programme's four priorities. Each priority was the subject of a separate study:

- business support measures (especially for SMEs);
- new technologies and innovation support;
- infrastructure, property and tourism and cultural industries measures; and
- community economic development measures.

Each report provided individual net employment impact and cost-per-job measurements for each measure as well as more general discussion of the main delivery issues affecting the priority. Project impacts were based on analyses of monitoring/output data from the programme managers and interviews with project samples. In measuring project impacts, the four studies made use of a common, supplementary study which provided a standardised list of ‘discounts’ and ‘multipliers’ for assessing net impacts. This study – undertaken by PACEC - reviewed 250 evaluation studies in order to generate a list of ‘parameters’ for different type of measures. The parameters were percentage figures for either discounting gross job impacts for individual projects (in the case of displacement or substitution) or multiplying them (for income multiplier and supply chain linkages). Different parameters were identified for different types of project and matched to the particular measures in the Yorkshire and Humberside programme. Hence, the parameters for a type of community economic development measure in another evaluation study were applied to analogous measures here (listed individually for the benefit of the evaluators). Consideration was made of differences in project aims, project populations and business cycles.

Subsequently, a further final study was undertaken on the performance of the ESF in the programme. As with the other studies, the evaluation was based on a review of the sub-regional analyses, interviews with programme officials and visits to a sample of projects, although as the PACEC study did not include parameters for ESF projects, little formal investigation was made of net impacts. This study began about four months after the rest of the work, when it became clear that this dimension required separate treatment in order to deliver information of comparable quality to that generated on other measures. Since this study was not originally foreseen, it was undertaken by a single individual and with limited resources which constrained the methodologies chosen. The study was undertaken in two phases.

First phase:

- desk research based on the 1994-96 Yorkshire and Humberside Programme and on other relevant Objective 1 and 2 implementations;
- desk research on the three sub-regional assessments and on the programme wide appraisal undertaken;
- analysis of the records of the Government Office for Yorkshire and Humberside relating to ESF implementation in the 1994-96 programme;
- analysis, with reference to ESF issues, of the Sub-Regional Assessments and of the programme wide appraisal questionnaire responses; and
- formulation, on the basis of the above-mentioned resources, of key questions on the implementation of ESF measures.

Second phase:

- interviews and meetings with staff of the GOYH European Secretariat;
- visits to project sponsors.

The report highlighted that no further investigation was possible because of the resource and time constraints and included some recommendations on the need to survey companies and beneficiaries in future, in addition to project sponsors, to achieve a full consideration of impacts.

Relevance to Scotland

Organising an evaluation as multiple, meso-level studies mobilises specialist expertise and specific methodologies to allow a significantly greater depth of analysis of policy performance and relevance than traditional 'global' evaluations focused on employment impacts across a whole programme. Such approaches, however, are time consuming and costly, and should only be used for large programmes (where even a small percentage of the technical assistance budget represents significant sums) and where the information generated will be put to genuine use over the remainder of a programme lifecycle. It might be expected that once a major study had been undertaken, the insights from this would continue to guide future evaluation, so that such a detailed, involved exercise did not need to be repeated in the short term.

Given the high costs associated, a fully developed approach of this type might only be useful in Scotland for combined, comparative analyses of more than one programme. An alternative would be to limit the use of meso-level evaluations to specific thematic issues where there is a genuine need for additional insight because measures are innovative and/or because their primary impacts are not captured by the measure of job creation alone (eg RTDI and sustainable development).

A related practice is to undertake 'global' evaluations, supplemented subsequently by one or more thematic evaluations on issues which require closer analysis. It was mentioned above that, in addition to the basic diet of evaluations, thematic studies are encouraged where they improve programming by responding to specific questions or challenges. These have been used extensively by Commission services but also in most Member States, especially for investigating 'horizontal themes' and programme management issues (see Table 5.2).

Table 5.2: Use of thematic evaluations among EU Member States

Member State	Extent of thematic evaluation
Austria	Limited use, mainly on RTD and linked to national policy evaluation.
Belgium	None known.
Denmark	Frequent use for specialised Structural Fund issues eg. RTDI, SME development, internationalisation. Seen as more useful than 'global' approaches to evaluation.
Finland	Frequent use for issues such as equal opportunities, environment and programme management.
France	Frequent use at regional level, some use at national level.

Germany	Frequent use at regional level. None at national level.
Greece	None known.
Ireland	Frequent use, especially for RTD.
Italy	Limited use. Some studies to 'deepen' programme evaluation findings.
Luxembourg	None known
Netherlands	Little evidence of thematic evaluation.
Portugal	None known.
Spain	None known.
Sweden	Limited use.
UK	Frequent use at regional level across the UK. None known at UK/GB level. Examples at Scottish level.
EC	Frequent use for a range of issues, including equal opportunities, environment, RTD, partnership, programme management and monitoring issues. Not only commission cross-national studies (eg for RTDI and partnership) but also enable selected programmes to undertake their own thematic studies as part of an EC programme of research- eg into good practice in delivering sustainable development.

The **use of programme evaluations and thematic studies** in an integrated, diagnostic process can perhaps be best illustrated by the example of **North-Rhine Westphalia** (see **Case Study 5**).

CASE STUDY 5 : USE OF INTERIM AND THEMATIC EVALUATIONS IN A DIAGNOSTIC PROCESS NORTH-RHINE WESTPHALIA

The 1994-96 interim evaluation study of the North-Rhine Westphalia Objective 2 programme contained three core elements:

- at the heart of the study, an assessment of the effects of the NRW Objective 2 programme at the interim stage;
- the effects of the Objective 2 programme on the diversification of the industrial structure, focusing on SMEs; and
- international benchmarking against other Objective 2 programmes, looking for experiences which could be transferred to the North-Rhine Westphalia programme.

The evaluation, as well as identifying the initial results of the programme in quantitative and qualitative terms, also contained recommendations addressing the strategy used, the allocation of resources, support for SMEs, programme steering, the integration of ERDF and ESF strategies, and programme publicity.

From a management perspective, the evaluation also raised questions about certain aspects of the programme. For this reason, a series of evaluations was commissioned from the same consortium, for 'deeper' evaluation of some strategic elements. The studies commissioned included:

- thematic case studies to examine the impacts of various priorities over the 1994-96 and 1997-99 programming periods, covering: the decontamination and reclamation of old industrial wasteland; technology-based sectors (information and communication technologies, logistics and the ChemSite initiative); and education and training;
- the effects of the Objective 2 programme since 1989 on a large long-term project, which involved converting an ex steel works into a multi-media attraction; and
- a discussion paper used for the preparation of the 2000-06 programme.

The thematic case studies were commissioned to examine the results of various priorities that had been set in the Objective 2 programmes, and to establish how relevant the strategy was in addressing the regional problem. The final element of these studies was to establish how such strategies could contribute to future regional development aims. In addition, an interim evaluation study also commented on the complications involved in collecting project data as this had been done via the traditional implementation structures in the region (the implementing units) and suggested centralising this information within the co-ordinating unit, a request to which the co-ordination unit responded.

Relevance to Scotland

The use of several specialist evaluators is more expensive and contractually demanding than hiring a single evaluator. It is potentially only really necessary for larger programmes, where the resources used are greater, justifying a greater evaluation effort, and only really feasible there, since the funding available for evaluation from the technical assistance budget (as a percentage of the whole programme) is also greater. However, it does deliver benefits in terms of range and depth of analysis.

The Case Study also illustrates the value of thematic evaluations in providing a flexible means of addressing specific issues highlighted by evaluation studies or by programme partnerships. Such practice has been used to an extent in Scotland, and this Case Study provides a well-developed example confirming that this practice is valued elsewhere in the EU.

Lastly, there is scope for consolidating the evaluation of *process issues* which has already been a strong dimension of Structural Fund evaluation in Scotland. As with the PME review conducted recently in Scotland, process evaluation can provide deeper insights and understanding into how administrative functions actually operate and inter-relate. An international example is provided by the approach to evaluation in **Tuscany** – see **Case Study 6**.

CASE STUDY 6: EVALUATORS AS PROCESS CONSULTANTS TUSCANY

In the interim evaluations of the 1994-96 Objective 2 and the 1994-99 Objective 5b SPDs in Tuscany, the evaluators interviewed all the officers in the regional administration responsible for managing the individual measures, using this to trace the processes and progress in the implementation of measures and to collect information on the operational difficulties encountered. The trust established and the co-operation derived, allowed the evaluator to elaborate, beside an accurate analysis of the programme's impacts, a clear set of process-related suggestions and also to isolate specific aspects in 'thematic analyses' on issues identified together with the regional administration. The activity of the evaluator has been interpreted extensively as a support to programme management, and this allowed the following:

- the identification of procedural bottle-necks and structural weaknesses, with the help of those responsible for the single measures, and, successively, the development of potential solutions and options to overcome the problems;
- the provision of specific recommendations on a measure-by-measure basis, in order to improve programme performance (these were followed up in most cases);
- to establish a climate of trust which activated a 'virtuous circle' firmly embedding evaluation: contact with the evaluator → increased awareness of the potential value of evaluation as a support function → increased co-operation → sharing of evaluation findings and results → identification of new tasks for evaluation.

The evaluator addressed operational issues using multi-criteria analysis. In particular, the steps taken by the evaluator included:

- re-assessment of the *ex ante* evaluations (often lacking in indicators coherent with the objectives of the programmes, and their priorities and measures). This made explicit the aims of each measure, allowing the definition of performance indicators;
- development of individual summary assessments for each measure;
- interviews with measure-managers, in order to identify the major implementation problems and risks of loss of resources;
- elaboration of questionnaires for each measure (related to delivery, procedural and impact aspects);
- discussion of the questionnaires with the measure managers;
- elaboration of the results of the questionnaire-based interviews in the framework of the summary assessments previously prepared; and
- 'translation' of the qualitative opinions contained in the questionnaires into weighted scores, in order to be able to compare the different measures.

This last stage was carried out according to a scoring system. Each measure was scored according to various criteria, for example financial, procedural and physical implementation, and the probability of achieving the anticipated goals. The scores, ranging for example from 1 to 5 (ie. for the physical implementation, insufficient information = 1 point; serious delay = 2 points; minor delay = 3 points, etc) were then weighted according to the degree of complexity ascribed to each measure (for instance: the lowest weight being applied to the measures with a traditional approach and with previously identified beneficiaries, and the highest weight attributed to innovative and open approaches, ie. with non-identified recipients, measures). The sum of the weighted scores attributed to the various criteria for each measure, would determine a total weighted measure score. Both this total score and the measure weighted scores related to each criterion allow objective comparison of the performance of the heterogeneous measures and an overview of the overall programme.

The Tuscany evaluation reports include several analyses of this type, for example relating to the selection criteria used for project selection, and assessments of the efficacy and efficiency of implementation of the measures. Efficacy has been assessed with respect to the capacity of the various measures for meeting the forecasts formulated in the (re-assessed) *ex ante* analysis. Efficiency has been evaluated through scoring the dimensions of financial implementation, physical implementation and implementation uncertainty. The methodology utilised and the premises for the analysis (the weights and scores attributed) are clearly illustrated in the reports.³⁰

Relevance to Scotland

In many cases, the evaluations undertaken in Scotland have addressed process issues effectively. The above example provides an illustration of a more involved approach to process evaluation than is usually used in Scotland. It also highlights the idea that process evaluation could be improved by also involving evaluators in-between evaluations, with an approach to evaluation which is more integrated into the programming cycle.

5.3 Promoting partner involvement

Since Structural Fund evaluations are addressing programmes which are designed and delivered on a partnership basis, involving both a wide vertical and horizontal partnership, it is appropriate that evaluation studies are undertaken with the active participation of more actors than simply the programme manager. The effective participation of multiple actors in Structural Fund evaluation can bring several benefits. In particular, it raises the profile of evaluation among the partnership, encourages more realistic expectations about such studies, contributes to the accuracy of reports by

³⁰ The method utilised to attribute the weighted scores is also set out by Manuela Crescini, the leader of the evaluation team that undertook the studies, in the paper *The on-going evaluation of the SPDs, Objectives 2 and 5b of the region of Tuscany*, presented at the EC's Seville Conference on evaluation, March 1998.

ensuring timely feedback, and ensures wide acceptance of results. Each of these makes it more likely that an evaluation report will be exploited, whether to adjust policy or operational practices.

Table 5.3: Partnership involvement in evaluation

Member State	Co-ordination of evaluation exercises
Austria	National evaluation steering group
Belgium	Regional evaluation steering groups
Denmark	Programme level and, in some cases, multi-programme and national evaluation steering groups
Finland	Programme level evaluation steering groups
France	Programme level evaluation steering groups
Germany	Programme secretariats
Greece	National evaluation steering group
Ireland	National evaluation steering group
Italy	Programme secretariats
Luxembourg	Not known
Netherlands	Monitoring Committees
Portugal	National evaluation steering group
Spain	Programme secretariats
Sweden	Programme secretariats
UK	Programme level evaluation steering groups

In terms of the positive virtues of involving partnerships in evaluation, practice in Scotland is in line with that in most other Member States and regions (see Table 5.3), with the routine establishment of regional evaluation steering groups or committees to accompany evaluation studies from the definition of objectives and the design of the terms of reference to the finalisation of the report and, in some cases, the exploitation of its results. This approach is common in several other countries – Belgium, Denmark, Finland, France – but is not universal. The cohesion countries (and also Austria) tend to use a national evaluation steering group in which different government departments as well as some regional interests are represented, and in some countries (Italy, Germany, Spain, Sweden) evaluations are managed directly by programme secretariats.

Given that one of the issues of interest for Scotland is how to improve partner involvement in the evaluation process, there are some interesting practices in **Finland** concerning the use of **more permanent engagement of partners** as well as their more efficient deployment through evaluation planning (see **Case Study 7**).

CASE STUDY 7 : PARTNERSHIP INVOLVEMENT IN EVALUATION FINLAND

In Finland, it was noted earlier that there is relatively poor co-ordination of the evaluation process at national level, with problematic relations between government departments. Within individual departments, however, co-ordination is good in some cases. For ESF, for example, an effective structure is in place.

A strategic Steering Group for ESF Evaluation has been established to promote intensive, open and on-going dialogue between a core group of actors involved in the evaluation of ESF programmes, including relevant evaluators. It operates as a permanent evaluation steering group, which meets on a long-term basis, and is not confined to the periods when evaluations are actually underway. This enables a more strategic and anticipatory approach to evaluation, also permitting genuine expertise to be built up more easily over time among the participants. The function fulfilled by the group is highly valued by participants - so much so that more people now want to attend the meetings, which are seen as a good source of feedback and information about policy. The perceived risk now is that by enlarging the group - particularly to passive observers - it could lose its momentum and become less useful. The focus could shift from open, problem-solving discussions between experts (the current most valued function), to information provision for non-experts.

A further good practice from Finland in terms of partnership involvement is at programme level: some Finnish evaluation steering groups establish an explicit **evaluation plan** which structures the process of evaluation in advance, setting out timetables, and the roles and objectives of the various participants. This may help to give momentum and transparency to the process, promote partner engagement and ensure more strategic planning, helping to ensure that momentum is not lost before the study can be exploited.

Relevance to Scotland

Improving partner involvement is one of the issues on the agenda for the new round of Scottish Structural Fund programmes. The earlier key partners are involved in evaluation - including designing evaluation frameworks - the more likely they are to be appropriate and acted on willingly.

At Scottish level, evaluation groups similar to the Finnish Steering Group for ESF could help to develop integrated approaches to evaluation. Indeed, a group has already been proposed for high-level exchange. The Finnish experience indicates (i) that the role and objectives of the group need to be clearly defined, and (ii) that there may be a need to distinguish between strategic working groups and groups with a role more focused on information dissemination.

At programme level, Evaluation Steering Groups like those in Finland are already used in Scotland. The Finnish example, however, formalises the activities of the group with the agreement of a written Evaluation Plan. Such approaches can help to bring additional clarity and transparency to a complex process.

5.4 Strengthening the capacity of the evaluation community

5.4.1 Working with evaluators

The evaluation community is defined here as the population of professional evaluators available to respond to requests for evaluation studies. In the field of Structural Fund evaluation, the predominant approach has been to commission external evaluations from the private sector, ie. consultancy firms, and/or from experts in the non-profit sector, eg. universities. An alternative used in some other Member States and regions is the establishment of 'in-house' public sector evaluation units, with a specific role to undertake independent evaluation of government activities at various levels (eg in some French and Spanish regions, in Ireland and in Northern Ireland, and, within Great Britain, in the DfEE for previous ESF evaluations).

External evaluation of Structural Fund programmes is much more frequent than the use of public sector evaluation units, in part because of the EC's requirement for visibly independent evaluation. However, for non-regulatory evaluation, and for *ex ante* appraisals in particular, there are numerous examples of evaluation work being internalised within governmental evaluation units. The practice is also pursued for regulatory evaluations in some cases, but on condition that the independence of the nominated unit be demonstrated to the EC in advance. Each of the two approaches has certain merits and difficulties (see Table 5.4).

Table 5.4: Relative merits of using external and internal evaluators

Advantages of external evaluators	Advantages of internal evaluation units
Ability to select relevant experts for different types of evaluation. Clear impartiality.	Possible to ensure continuity in successive evaluations of a given programme - with an accumulation of data and knowledge.
Difficulties with external evaluators	Difficulties with internal evaluation units
Difficulty in guaranteeing the quality of the result - need to track the evaluation closely.	Need to demonstrate impartiality to the Commission and other programme partners - may be questioned. Expense of retaining personnel year-round when the demand for evaluation does not merit this.

The evaluation community in every EU Member State has grown in size and skill in response to the demand for Structural Fund programme evaluations.³¹ The current research suggests that three sets of factors play a role.

³¹ MEANS (1998) *Le Développement de l'Évaluation des Interventions Structurelles dans les États Membres*, C3E, Lyon, France.

- *The length of time a Member State has been in the EU.* For instance, the evaluation community in Sweden, which joined the EU in 1995, is still characterised as small, with a limited number of private, competent evaluators. There is a growing number of centres developing evaluation skills, but they often have little knowledge of regional policies or Structural Funds.
- *The size of the market for evaluation services.* This is related to the number of programmes operating in a given state. The evaluation community in Denmark is limited by the size of the Structural Funds market.
- *The overall status of evaluation.* While Spain has been a member of the EU since 1986, and has a large demand for Structural Fund evaluation, the status of evaluation is low, limiting the development of the evaluation community.

Table 5.5 provides an overview of the availability of evaluators in different EU Member States. In smaller countries such as Austria and the Nordic countries, the evaluation community is relatively small, and there is generally a need to use the same consultants for almost every evaluation or draw on foreign expertise (although this is not always possible for language reasons). This also applies to some cohesion countries (Greece, Portugal) where there are only a few, small consultancies specialising in Structural Fund work. In the case of Finland, it appears that disenchantment with Structural Fund evaluation requirements is actually leading to a declining pool of EU evaluators. At the other end of the scale are the larger countries – France, Germany and the UK – where there are sizeable and active evaluation communities and a wide range of EU evaluators to draw on from private consultancy firms and academic departments. The UK probably has among the most vigorous population of evaluators. However, within the UK, for cultural and geographical reasons, Scotland (in common with Northern Ireland) has a relatively clearly defined and distinct evaluation community, which is limited in size. Availability is, however, not synonymous with quality. The Irish evaluator pool, for example, whilst small is of high quality, while that in some other Member States is larger but more variable (eg France and Spain).

The way in which evaluations are commissioned may influence the openness of the evaluation community, and thus the ability of new providers to join the market. The predominant approach in Scotland is to make restricted calls to known service providers. In France, two practices help to widen the evaluation community used by any single programme. The first, for many programmes, is to issue open calls initially, and then to draw up a shortlist from the offers received through this process. The second is to encourage communication between programmes about the evaluators they have used, and so encourage a wider population to be called upon. The 'days of exchange' on evaluation experience, which are organised by the French national development agency, DATAR facilitate this process.³²

³² These days are described in detail in Case Study 11, which provides a wide-ranging illustration of methods to improve evaluation capacities more generally.

In Italy, open calls are the predominant approach - made through the relevant national or regional official journal. In order to address the issue of quality, terms of reference clearly set out the conditions which have to be met by evaluators, eg as regards their previous experience in the field of Structural Fund evaluation. In this sense, the process can tend to privilege the most experienced evaluators, which can present a barrier to newcomers. However, this can be overcome through inexperienced firms participating in joint bids with more experienced actors in 'Temporary Business Associations' (ATI - Associazioni Temporanee d'Impresa). This joint bidding allows less experienced evaluators to bid and also contributes to exchange of experience and capacity development among evaluators.

In some cases, the range of available evaluators is limited by geography. Bottom-up Structural Fund programme evaluation methods and the partnership dimension of evaluation often require evaluators to be close to the programme area. Since most Structural Fund evaluators tend to learn their skills 'on the job', and very limited opportunities exist for formal training in this area, the level of professionalisation is often relatively low, although it has developed significantly over time.³³ To expand choice to include higher quality but more distant service providers, interview evidence indicates that a frequent French solution is to contract two evaluators working as a team - one from the preferred lead institute, and one from a local organisation who undertakes the functions requiring more constant local availability. This alliance is sometimes proposed by the evaluator and sometimes by the organisation commissioning the evaluation. This process also leads to capacity building among evaluators in the provinces. A related approach has been used for Objective 1 and 2 evaluation in Germany, with a local evaluator working in collaboration with a team member from another country in order to get access to different evaluation techniques and experience.

Table 5.5: Capacity of the evaluation community

Member State	Availability of evaluators	Evaluation society
Austria	Limited – use of foreign consultants required	None
Belgium	Adequate – some use of French evaluators	Société Wallonne de l'Evaluation de la Prospective
Denmark	Limited	None
Finland	Limited and apparently declining	Finnish Evaluation Society
France	Good – active community	Société Française de l'Evaluation
Germany	Good – active community	Deutsche Gesellschaft für Evaluation
Greece	Limited	None
Ireland	Limited but high quality	None known
Italy	Good – active community	Associazione Italiana di Valutazione

³³ MEANS (1998) op. cit. p11.

Luxembourg	Limited, but additional availability in neighbouring Member States	None
Netherlands	Good – active community	None
Portugal	Limited	None
Spain	Adequate with variable quality	None
Sweden	Limited – use of foreign consultants required	None known
UK	Good – active community (limited numbers in Scotland)	UK Evaluation Society

A second indicator of the evaluation community summarised in Table 5.5 is the presence of evaluation societies which can contribute to developing the quality and quantity of the actors involved in Structural Fund evaluation. A growing number of national and international evaluation societies have been established during the 1990s, most of which aim to develop evaluation capacities across numerous disciplines. The table illustrates known evaluation societies. They are generally concerned with evaluation in its broader sense, but the Structural Funds are often among the key themes, since they represent a significant policy area, and evaluation in this field has required specific approaches and techniques. Such societies are not always seen as useful by practitioners, some being considered as too academic and concerned with theoretical issues to provide genuine support to the practical process of policy evaluation. In other cases, however, they have made a notable contribution to widening the pool of EU evaluators by encouraging debate and learning on relevant policy evaluation methods, for example in Italy and Germany. The organisation of evaluation society activities under a series of work programmes (as in France and Germany) can help to animate what could otherwise be a relatively passive structure.

A further issue is the relationship between the commissioning authority and the evaluator. There is a choice between using a single provider for the whole of a programming period (several studies being let in the same call), so establishing a longer-term relationship, or using different providers each time, selected by a new call before each study. Of course, there are again advantages and disadvantages in both approaches. Using a single provider for multiple studies is quicker (less time is lost on the commissioning process) and means that there is less need for the evaluator to build up familiarity with the programme, its management and monitoring systems, etc: more time can be spent responding to the brief. It also allows longer-term relationships to be built up. On the other hand, using a new provider for each study has the virtue of competition: it may be cheaper in cost terms (although not necessarily value for money); it enables relevant expertise to be bought in at each stage; there is less risk (if the evaluators do not perform well, then a different team can be selected next time); and it ensures that the evaluator is seen as impartial (ie. not swayed by vested interest).

As with the issue of multi-programme versus single-programme evaluations, Scottish Structural Fund evaluation practice has elements of both of the above approaches. Separate calls are made for each evaluation, but largely from the

same pool of providers, and the same evaluators have sometimes been chosen using the argument of continuity and familiarity. **North-Rhine Westphalia** provides an example where an explicit attempt has been made to cultivate a **relationship with an evaluation team** – see **Case Study 8**.

**CASE STUDY 8: CULTIVATING A RELATIONSHIP WITH AN
EVALUATION TEAM
NORTH-RHINE WESTPHALIA**

The approach to evaluation in North-Rhine Westphalia has been both extensive and sustained: extensive because the complementary services of a large evaluation team have been used for each study, and sustained because the same consortium of evaluators has been used since the 1994-96 interim evaluation.

Continuity in the evaluation team has allowed them to gain a deep understanding of the regional structure and operation. This brings the evaluators an understanding of the region which can only be acquired over time. The consortium consists of three evaluation institutes:

- a locally-based evaluator chosen for their knowledge of the region;
- a consultancy specialising in Structural Fund evaluation; and
- a Dutch evaluator who provides an international perspective.

In NRW, the programme manager believes that the combined strengths of each evaluation institute produce work which is collectively stronger than that which could be produced by a single evaluation team. The arrangement has been sustained because it has delivered worthwhile results.

Relevance to Scotland

There may be possible benefits in Scotland from programmes considering contracting a core evaluation team for multiple evaluation studies - especially given that in the 2000-06 period there will be two interim evaluations required, in relatively rapid succession (at the three and five year points). Continuity here could help in delivering good quality results rapidly, and might provide programmes with an ongoing source of advice and support.

5.4.2 Developing the capacity of the evaluation community

The development of expertise and skills through *ad hoc* initiatives has been increasing in parallel with the evolution of the concept of evaluation in itself. The increasing attention given to evaluation has been associated with initiatives at EU and national levels promoting the development of competence and know-how.

In the first programming periods, programme managers could not count on the availability of Structural Fund evaluation skills which had to be developed through a process of learning by doing. Even in countries which had a significant experience in evaluating domestic regional policies, Structural Fund programmes required new, specific knowledge, since the programmes were complex multi-annual initiatives, involving multiple actors and interests.

Interview evidence indicates that there was little specific pre-existing expertise in the area, and in many cases also no particular theoretical grounding available. This applied both to those commissioning evaluations and to evaluators.

As noted in the previous section, the process of learning by doing was progressively supported in later programming periods by training initiatives, by the publication and dissemination of developmental documents, and by a broader debate on evaluation through conferences and meetings. Various capacity building initiatives were undertaken by the EC and national authorities. At the apex of this process is the **MEANS programme** funded by the European Commission (see **Case Study 9**) which, despite publicity, is apparently still not sufficiently well-known among Scottish Executive officials, programme managers, partners and some evaluators.

CASE STUDY 9: DOCUMENTARY SOURCES ON EVALUATION PRACTICE - THE MEANS PROGRAMME AND EC RESOURCES

The most structured and sustained EC initiative for developing methodologies and other support for evaluation capacity is represented by the **MEANS** programme. The programme was started at the initiative of DG XVI at the end of 1994,³⁴ in response to the increasing demand for evaluation methodologies appropriate to the assessment of Structural Fund interventions and to address other problems related to this activity (eg. the development of a common language in a complex and evolving area).

The MEANS research programme, whose activities were reported in a quarterly MEANS Bulletin, led to the publication of a range of thematic methodological reports. In addition, under the aegis of MEANS, three international evaluation conferences took place (in Brussels 1995, Berlin 1996 and Seville 1998) representing a significant and high-profile opportunity for evaluators and other actors involved in the evaluation of Structural Fund interventions to debate and exchange experience. The conference papers and reports of conference have been made available on-line, through the Europa server (<http://europa.eu.int>) and the Inforegio web-site (<http://inforegio.cec.eu.int>), supporting the low-cost diffusion of best, distinctive and innovative practices and encouraging debate among practitioners in all Member States.

In association with the MEANS programme, an extensive series of seminars for evaluators was carried out by C3E, the consultancy firm based in Lyon which ran the MEANS programme. Targeted training interventions continue to be delivered by C3E on a commercial basis. The company has also been contracted to provide methodological support to selected regions in their work with evaluators (eg. Poitou-Charentes, Franche-Comté and Wallonia).

³⁴ MEANS (1998), *Internal Bulletin N.12 October 1998*, MEANS, Lyon, France, p. 5.

In response to requests to the EC to circulate MEANS reports more widely, the core output of the MEANS project was recently consolidated into a set of six handbooks, collectively entitled *Evaluating Socio-Economic Programmes* (cost: €120), which encompass central aspects of Structural Fund evaluation:

- *Volume 1: Evaluation Design and Management*
- *Volume 2: Selection and Use of Indicators for Monitoring and Evaluation*
- *Volume 3: Principal Evaluation Techniques and Tools*
- *Volume 4: Technical Solutions for Evaluation within a Partnership Framework*
- *Volume 5: Transversal Evaluation of Impacts on the Environment, Employment and other Intervention Priorities*
- *Volume 6: Glossary of 300 Concepts and Technical Terms*

In spite of these efforts to make MEANS reports more widely available, interview evidence indicates that they are still not sufficiently well known among evaluators or civil servants in many Member States.³⁵ These handbooks are an essential starting-point for anyone involved in the field, providing a core set of definitions and a guide to the menu of methodological choices available to any given study, illustrated by examples of where the different methods have been used.

Where they are known, MEANS materials are perceived as a source of useful basic information - although they contribute relatively little to encouraging ambitious methodological improvements. MEANS reports are particularly valuable if they are seen as a first step, and as a support to methodological decision-making rather than a methodological base on which to rely completely. The handbooks do not provide formulaic answers, but instead equip actors to be better able to structure their own decision-making.

Linked to the MEANS programme, it should be noted that there are further, more easily obtainable, *developmental documents* published by various DGs. Among them is the series 'Evaluation and documents' published by DG Regio:

- No.1 – *Counting the Jobs. How to evaluate the employment effects of Structural Fund interventions*, January 1997
- No.2 – *Learning through Best Practice. Best Practice identified in the context of the ex post Evaluations of Community Initiatives (1989-1993)*, May 1997
- No.3 – *Guide to Cost-Benefit Analysis of Major Projects in the context of EC Regional Policy*, June 1997 (1997 edition)

³⁵ CNASEA - DATAR (1998) *Les évaluations intermédiaires 1996-97. Programmes Communautaires des Objectifs 1,2 et 5b*, DATAR, Paris, notes that: 'Très peu de rapports ont fait état précisément des techniques utilisées. A ce titre, les évaluateurs ont précisé qu'ils n'ont pas eu forcément connaissance des outils d'analyse développés notamment par la Commission européenne dans le cadre du programme Means'.

- No. 4 - *Ex post evaluation of the 1989-1993 Objective 2 programmes. Synthesis report*, September 1997
- No.5 – *Understanding and Monitoring the Cost-Determining factors of Infrastructure Projects. A User's Guide*, April 1998
- No. 6 - *Environment and Sustainable Development. A Guide for the ex-ante evaluation of the environmental impact of regional development programmes*, May 1999.

Other related DG Regio reports have also been broadly diffused, including through the Inforegio web-site (eg. the 1999 publication *Better management through evaluation. Mid-term review of Structural Fund programmes: Objective 1 and 6 (1994-99)*).

Relevance to Scotland

It would be cost-effective to ensure that each Scottish Executive and Structural Fund programme management teams had full knowledge of the documents issued by MEANS and relevant DGs of the European Commission. The Scottish Executive could act as 'intermediary', providing newly issued documents and information to the PME's on a regular basis, including through the Scottish cross-programme committees which meet on a regular basis.

Most of the capacity-building initiatives at national level have tended to be *ad hoc* or related to institutional investment in evaluation units within government departments, such as in Ireland and Italy. The most structured example of capacity-building for the wider evaluation community is the so-called 'Checkpoint EVA', a **common evaluation resource in Austria** – see **Case Study 10**. A different approach is used in **France** where the emphasis is on **active and interactive exchange on evaluation**, in particular to learn lessons following evaluation exercises – see **Case Study 11**.

CASE STUDY 10: CHECKPOINT EVA - AUSTRIA

Before its accession to the EU, Austria lacked a strong evaluation culture. At first, Structural Fund evaluations were viewed as tasks which had to be undertaken to meet Commission requirements. However, over time, there has been increasing awareness of the purpose and uses of evaluation studies.

A key development has been the creation of a common evaluation resource, focused primarily on Structural Fund evaluation. Initiated by the Federal Chancellery, *Checkpoint EVA* is run by the Austrian Spatial Planning Institute (ÖIR - *Österreichisches Institut für Raumplanung*), a research institute specialising in spatial policy.

Checkpoint EVA provides a central point of contact within Austria for evaluation information and documentation. The Internet is actively used to increase the accessibility and exploitation of the facility's resources. Their site, which can be accessed at <http://www.oir.at/checkpoint-eva.html>, includes the following resources:

- a **database** of all evaluation projects in Austria, including projects which are planned, underway and complete (listed on-line). These studies are not limited to Structural Fund evaluations, but have a wider scope, including evaluations of technology, structural and labour market policies, especially where these aim to identify good practice;
- a **news page**, containing evaluation news up-dates. The scope includes the latest in Structural Fund evaluation news and news items on specific themes such as gender mainstreaming;
- a **discussion forum** where theoretical and methodological evaluation questions can be posed for wider discussion among other evaluators or interested parties.

Checkpoint EVA also offers the use of an evaluation database in the ÖIR library, which is described on the internet site as ‘probably the most extensive collection of the most important literature on evaluation in Austria’. It contains over 600 entries including handbooks, EU guidelines and studies. It also offers a directory of evaluation actors, with a list of the names, addresses and evaluation activities of relevant institutions across the EU.

The value of the service is that it brings increased visibility to the evaluation function in Austria and ensures that there is an easily accessible source of relevant information, available nation-wide.

Checkpoint EVA has required relatively modest resources since the service is housed in an existing library, and the website is a sub-section of the site of the host organisation. Additional budgetary requirements were for document purchase, personnel (one person part-time) and web site input.

Relevance to Scotland

The idea of an evaluation centre like *Checkpoint EVA*, could easily be transferred to Scotland, supporting the evaluation efforts of practitioners and other stakeholders. The resources required do not have to be significant, especially if such a facility is sited in a pre-existing organisation already offering information services.

CASE STUDY 11: LESSON-LEARNING FROM EVALUATION FRANCE

Following each major round of evaluations in France, a ‘day of exchange’ is held to draw lessons from the experience gained. The days of exchange form part of a wider structured programme of guidance and capacity building.

In common with various other capacity building initiatives, the days of exchange are co-ordinated by DATAR, which takes the lead in co-ordinating evaluation efforts, with the practical support of the CNASEA - an organisation providing support to French Structural Fund programmes under the national technical assistance programme.

As an example, one such day of exchange was held on 7 October 1997, based on the 1996/7 interim evaluations of Objective 1, 2 and 5b programmes. Run by the CNASEA, the day involved national experts, regional programme managers and/or evaluation experts and the evaluators themselves in discussing the most recent round of studies. The aims were to take stock of results, to promote future evaluation activities and to learn lessons.

In preparation for the day, a survey of those commissioning the evaluations was carried out to enable experiences to be addressed in a systematic manner. Two areas addressed were the process of evaluation and the use made of the results. In addition to the results of the survey, the published proceedings also contained a list of all the evaluators used, including details of their experience and expertise.³⁶

The proceedings provided an overview of the evaluation activity undertaken, enabling each region to place its own practices in a wider context, and potentially to identify new ideas. Among the sections covered were:

- how the evaluation function is allocated in each programme management team;
- how evaluators were selected and what studies cost;
- how studies were managed, including partnership structures used;
- the contribution of monitoring information to evaluability; and
- the realism of the objectives set for studies and the impact of this on their results and reception.

Drawing on discussions during the day of exchange, the proceedings which were published subsequently set out an explicit work programme to improve the evaluation of Structural Fund programmes in France. Among the resolutions were:

- to undertake a study of the employment effects of Structural Fund interventions (now complete);
- to distribute a short guide to the evaluation of Structural Fund programmes (done);
- to establish an Intranet exchange forum (underway); and
- to organise inter-regional workshops to support programme managers in writing their terms of reference for final evaluations.

Among the perceived benefits of such events are the following:

- a partnership-based approach, where all the main actors involved in evaluation meet to draw lessons and establish recommendations to move forward, ensures that resolutions reflect the interests of and have the commitment of all key actors;

- those procuring evaluations can meet a wider selection of active evaluators with relevant expertise, and discuss their work with organisations which have already used their services;
- evaluation comes to be seen as a part of a wider process, in which follow-up, feedback and lesson-learning are integral parts; and
- all regions benefit from the lessons learned in individual regions.

Relevance for Scotland

Institutionalising periodic ‘days of exchange’ after each round of evaluation would provide a useful opportunity to develop active debate and awareness on evaluation in Scotland. As noted in Section 3, there is already informal exchange of experience and know-how among the PMEs. More formalised and enlarged meetings, open to the participation of the main partners and involving evaluators, would give the opportunity for a more extensive debate on relevant issues. Among the strong features of French experience are the prior surveys of evaluation experience which enrich the days of exchange, and active follow-up with the publishing of proceedings and recommendations arising from the discussion. Once again, such action could be co-ordinated by the Scottish Executive, potentially with the collaboration of a further facilitating organisation.

Among other national initiatives, measures are being undertaken in Italy to develop evaluation skills with the creation of a National Evaluation Network whose functions are capacity building and dissemination. An *ad hoc* convention between the Ministry of Treasury and Budget, the institute Formez and the State-Region Conference has already set the basis for the activity of this network and a State-Region protocol has committed resources on a multi-annual time-scale. The inception meeting of the technical-scientific committee has recently been held to determine a programme of activity. Training and developmental activities in Italy were also undertaken in the 1994-99 period under the PASS programme, aiming at developing the skills and expertise of local administrators in the Mezzogiorno and of officers from national administrations operating in the Structural Funds, to increase the efficiency and effectiveness in the use of the Structural Funds and to support administrations in their re-organisation for the implementation of Structural Fund programmes.³⁷

A related activity was undertaken in Finland where a task force in the Ministry of Finance was established. In addition, with regard to the ESF, the Finnish Ministry of Labour publishes regular evaluation reports, some of them

³⁶ CNASEA/DATAR (1998) *Les Évaluations Intermédiaires 1996-97 - Journée d'Échanges du 7 octobre 1997 entre Évaluateurs et Gestionnaires, Informations Générales et Compte-rendu des Débats*, DATAR, Paris.

³⁷ More information on the PASS programme are available at the web-site of Formez, <http://formez.it>.

addressing specific problematic methodological questions, with the aim of providing guidance for programme managers and evaluators.³⁸

5.5 Improving the quality of data

The availability of reliable, homogenous and meaningful data is the pre-condition for carrying out evaluations. Past programming periods have been characterised by a general lack of information and data on the outcomes of programmes. This has meant that sometimes evaluators would undertake only 'partial evaluations', assessing only what they considered to be capable of evaluation given the data available. More often, evaluators would commit themselves to the task of establishing data sets, taking time and resources away from 'real' evaluation tasks.

The quality of data available to evaluators in Scotland has improved significantly in recent years following the investment in quantification at programme and project levels since 1994 (see section 3.2). This effort is continuing in the new programming period, with a commitment by programme managers to work more closely with project applicants to ensure that applicant data is more credible and reliable, and with a Scottish Executive-led initiative to establish core indicators.

There are no outstanding case studies from other parts of the EU in this area: improving the quality of data is an incremental, multi-faceted process where developments tend to be country-specific and even programme-specific. The following sub-section, therefore, provides a general review of recent trends and patterns among Member States, highlighting the most significant aspects.

Across the EU, monitoring systems have undergone sustained development over the last decade, especially during the late 1990s. There is no standard EU monitoring system imposed on the Member States. However, while they have flexibility in design, it is increasingly the case that systems have to fulfil a range of specific conditions, for example, compatibility for electronic data exchange with the EC to facilitate their own future analyses.

Developments in data management technology have created the possibility for improved monitoring systems and tools. Among the key trends have been shifts:

- from paper-based to computerised data management systems;
- from databases housed on single computers to networked databases accessible from multiple points; and
- towards flexible internet- or intranet-based systems which do not need special software to be installed on individual users' computers for access.

Major problems remain, most notably the fragmentation of information. Two types of fragmentation stand out: (i) data for different Structural Funds is frequently collected separately - especially for the ERDF and ESF; and (ii) systems may be specific to single regions - there is not always uniformity in

³⁸ *The Model of Evaluation for ESF Projects and Employment Effects*, Ministry of Labour, Helsinki, Finland.

even the basic parameters of the systems in use to monitor different Structural Fund programmes.

Table 5.6: Monitoring systems in EU Member States

Member State	Monitoring systems
Austria	Single, integrated, national monitoring system – EVI – available on-line to national and regional authorities.
Belgium	Separate regional monitoring systems for Wallonia and Flanders.
Denmark	Two-tiered monitoring system (national and regional) with limited integration.
Finland	Single, integrated, national monitoring system - FIMOS - for ERDF with on-line access, part publicly available, part restricted.
France	Single, integrated, national monitoring system - <i>Présage</i> – for ERDF available from 2000. Working towards integration with ESF database, but not yet achieved.
Germany	Separate regional monitoring systems for each of the <i>Länder</i> .
Greece	National monitoring system under development.
Ireland	National monitoring system under development.
Italy	Single, integrated, national monitoring systems – SIRGIS – mainly for financial data. Separate, more comprehensive regional monitoring systems in some regions. Monitoring System of Public Investments (MIP) and a related database under development. ³⁹
Luxembourg	Not known
Netherlands	Single, integrated, national monitoring system.
Portugal	National monitoring system under development.
Spain	Separate regional monitoring systems for each of the regions.
Sweden	Single integrated national monitoring system.
UK	Separate monitoring systems for each of the UK territories.

Table 5.6 illustrates the basic differences in monitoring systems among Member States. There is a basic contrast between several countries (Austria, France, Finland, Italy, the Netherlands, Sweden) which have or are moving towards a single integrated, national monitoring system and countries (Germany, Austria, Spain) where monitoring systems are region-specific. In some of the cohesion countries (Ireland, Greece, Portugal) a national monitoring system is still under development. It should be noted that even

³⁹ See Case Study 2, *Institutional arrangements for managing evaluations - Italy*.

where there is a single national monitoring system, this may only or predominantly apply to ERDF (eg. Finland, France) or to financial data (Italy). Interestingly, some of the systems are beginning to be made available to partners, and even the public, via the Internet (Finland, Wallonia). Looking at developments in more detail, some examples from Denmark, Finland, Sweden, Belgium and Italy provide insight into the problems of improving the quality of data.

- *Denmark.* The co-existence of two tiers of monitoring systems, one national and one regional, with different data and time-scales represented a significant problem in terms of co-ordination. The national system operates over the lifetime of the projects with data inputs on a three-monthly basis, while the regional system is responsible for tracking data only at the beginning and at the end of projects.
- *Finland.* The first attempt to create a comprehensive monitoring system at the national level – the Rehua system - was not immediately successful, with problems of co-ordination and reliability of information;⁴⁰ it was re-organised in 1997, under a new name, FIMOS, with significant improvements. FIMOS covers all Structural Fund awarded projects, with information on expenditure and projects summaries available on line at <http://fimos.atbusiness.com> (more detailed data and information are subject to restricted access). Project data are supplied through the PMCs (which in turn are gathered from the individual returns of project participants) to the central data collecting point within the Ministry of Interior, which maintain FIMOS. Finland has good, detailed data in a series of ‘registers’ available to evaluators (including population, housing, business information at disaggregated levels) but for confidentiality reasons there is restricted access. Some questions continue to be raised concerning the effectiveness of the data gathering: this is generally overseen by regional authorities on a monthly basis, but officials at national level, question the adequate measuring of the data inserted.
- *Sweden.* Data is considered a real problem: the national monitoring system included a lot of data and indicators that were not encompassed in project applications. As a result, the regional programme authorities did not put much effort into this task, which they considered too time-consuming and inappropriate to their needs. Instead, they developed autonomous data collection systems, normally in the form of Excel charts, which were not always consistent with the data gathered by the national database. The resulting overall picture often involved little consistency. Evaluators would therefore try to combine the two levels of data, integrating the information available with *ad hoc* survey analysis and sample visits, in order to reconcile discrepancies. Some attempts to improve the data

⁴⁰ Malinen P *et al* (1997) *Interim report of the Mainland Finland Objective 5b programme*, Kajaani Development Centre, Regional Development, University of Oulu. The report underlines that “the role of REUHA (information system) register of projects is absolutely essential from the point of view of co-ordinating the 5b programme. The fact that it has not operated reliably, as an adequate tool to serve its various users up to now, is an indication of weakness and a lack of co-ordination in matters of co-operation. New measures were initiated to improve the Project Register in the summer of 1997”.

collection are undertaken currently by the national economic development agency NUTEK, which is trying to develop a regional database.

The Objective 3 and 4 programmes in Sweden have a more effective system. For training activities, the evaluator can rely on the data of a national database on *individuals*, each one being identified by an ID code (equivalent to the UK National Insurance number). This database tracks the personal history of each person (ie. years of unemployment, access to the labour market, changes of job, salary etc) allowing a significant series of different analyses. The database is subject to restricted access for privacy reasons, but evaluators can submit specific queries and obtain the requested information on an anonymous basis. One strength of this system is represented by the longer-term tracking that this system enables: while most monitoring systems tend to record project-related data only during the life-time of the project, this system goes further, also allowing long term analysis.

- *Belgium.* The Wallonian government has a database of financial data dating back to 1996, which has been gradually extended to physical and impact data; since this system had not proved to be sufficient, a new Internet-based system has recently been launched, which can be directly fed into and consulted by programme managers. The new system is supervised by two programmers at the Direction for European Programmes, and academic supervisors that ensure that the encoding is done properly, although the ultimate and formal responsibility for the correctness of the data inserted relies on the programme managers. The Wallonian authorities have commissioned the Department of Applied Economics of the Free University of Brussels (ULB) to construct a special model for the Hainaut region, as well as supplying the data required for the model.
- *Italy.* The Objective 1 CSF for the 2000-06 period has created the basis for reorganising the pre-existing national monitoring system. For the 1994-99 programmes a computerised monitoring system had been established at the SIRGIS-IGRUE, Ministry of Treasury and Budget, linked with the IT centres of the regional and national administrations responsible for the implementation of the various Operational Programmes. The previous system tracked mainly financial data, while the re-organised system will also encompass procedural and physical data. A set of indicators (for outcomes, results and impacts) had already been identified prior to the approval of the CSF, being ready to be implemented at the start of the programmes. Also most Objective 2 regions are re-organising their monitoring systems for the 2000-06 programmes. Among others, Tuscany is perfecting an existing financial, procedural and physical monitoring system. Since 1994, the manager responsible for each measure has been periodically submitting data on measures to the central monitoring office of the Industry Service. With a computerised system, measure managers will be able to introduce data directly into the system, which will provide a clearer picture of the stage of implementation of the entire SPD. As with other regional monitoring systems, some of the information gathered will flow into the national monitoring system.

In Scotland, monitoring is perhaps one of the areas which has been improved since the past, with useful work on indicators and on information for evaluation purposes since the beginning of last programming period (see section 3.2). However, some problems persist. In particular, the reliability of data provided by project implementers has not always been optimal. In addition, for the institutional reasons highlighted in section 3, there has been little integration between the ESF and ERDF/EAGGF. In this respect, future steps in Scotland should be taken to:

- organise a coherent and integrated monitoring framework for all the Structural Funds, especially given the new role that Scotland will have for the ESF - the Finnish FIMOS system, for instance, represents a good example of a monitoring system which is easy to update and access; and
- consolidate work to date to increase the monitoring skills of implementers by dedicating increased resources to monitoring visits and to other developmental activities.

5.6 Following up evaluation studies: validation, dissemination and exploitation

There are several ways in which evaluations can be exploited, responding to different aims and objectives. However, evaluations have not always become a vehicle for learning and influencing decisions. The diffusion of evaluation findings is not just an element that provides opportunities for exchange and debate, but also involves other policy aspects, such as transparency, capacity building, legitimisation and validation. The extent to which an evaluation is disseminated and exploited among ‘users’ or ‘clients’ can significantly affect the returns that an evaluation study may have.⁴¹

One of the key recommendations of the MEANS research is for programme managers and partnerships to give a high profile to the dissemination and application of evaluation findings, including the elaboration of a communication plan to structure this process.⁴² The dissemination and use of the results are key phases of the evaluation process - regions must find a balance between transparency and confidentiality. The research makes several important points.

- Different documents should be produced for different types of audience eg. early drafts being dealt with by the evaluation sub-committee; special documents for social partners, private actors, and the media. MEANS note that the dissemination process should be structured to achieve positive reactions. For example, within the Monitoring Committee ‘special care needs to be taken to ensure that discussion of the report and decisions on its use takes place under the most favourable circumstances’.

⁴¹ Diez MA (1999) *How to improve evaluation practice? From traditional evaluation to participative evaluation*, Paper presented to the RSA International Conference ‘Regional Potentials in an integrating Europe’, Bilbao, September 1999. The paper argues that evaluation can only become an ‘open collective learning process’ if there is wide participation in studies on the one hand, with the evaluator as facilitator, and then wide and active dissemination of the evaluation results.

⁴² MEANS (1999) Evaluation design and management, *Evaluating socio-economic programmes*, Volume 1, Luxembourg, p.163.

- The dissemination process should be phased, timing the release of evaluation results to suit the deadlines of Monitoring Committee meetings, regional, national and Commission requirements.
- Dissemination should not rely exclusively on the distribution of paper documents, supplementary media potentially comprising press conferences, seminars, action days etc.

In practice, it does not appear that most regions are (or will be) adopting a communication plan as sophisticated as that suggested by MEANS. Partly this reflects the restricted scope of the evaluation studies conducted, but also the limited level of interest outside main programme partners. Some regions do regularly inform the media (including local, regional and national newspapers) but have found through experience that there is little press interest in the evaluations.

In practice, follow-up on evaluation studies takes one or more of the following approaches.

- *Internal circulation.* Most regions have a policy of circulating full copies of the reports to partners, including all relevant Committees, the Commission and implementing agencies. In some regions (in France, Italy and Sweden), it has been useful to have the evaluator present the results to the Monitoring Committee in person, which helps ensure that all key partners are aware of the findings even if they have not read the report. In one Spanish region, a forum of key partners was organised to discuss the outcomes of the interim evaluation of the Objective 1 ERDF Operational Programme, involving 15 hours of discussion in five structured sessions to give detailed consideration to the findings of the evaluation.⁴³
- *External dissemination.* Beyond internal circulation, a next step in exploiting evaluations is the dissemination of evaluation reports and findings. Whereas in Scotland, evaluation studies are not easily accessible beyond those involved in implementing programmes (it is necessary to know a study exists before it can be requested), elsewhere there is more structured and pro-active dissemination. A good example is the practice in Ireland where evaluation studies are published regularly and broadly diffused. Similarly in Finland, evaluation reports commissioned by the Ministry of Labour are published in a series of brochures, intended to be used as practical guides for programme managers and evaluators; dating back to 1997, there are now over sixty reports available. Evaluation reports are more likely to be published and to remain easily obtainable where there is an existing publications series in which they can be presented, even if only in summary form (eg a regional academic journal or official bulletin routinely used to publish current policy analysis and debate). This is the case in the examples mentioned above, and also in many German *Länder*, in Austria and in some Spanish regions, notably Catalonia.

⁴³ MEANS (1998) *Le développement de l'évaluation des interventions structurelles dans les Etats Membres*, C3E, Lyon, p.19.

- *Internet dissemination.* Complementing the dissemination of evaluation reports in text form is the use of the Internet. Although the Web is starting to be exploited for disseminating evaluation material – by the EC and some national administrations (eg. Finland) - the Internet presence of evaluations is still poor, in most cases even absent. The Web clearly has much wider possibilities than just dissemination; as noted earlier, it can be a powerful tool for assisting all aspects of evaluation.
- *Validation.* In most cases, the completion of an evaluation study signifies the end of the evaluation process. For those engaged in Structural Funds management, it is often regarded as sufficient to have thereby fulfilled the regulatory requirements. Although the evaluation process may have experienced problems, there is usually no formal process of validating the results or drawing lessons for the future. There are, however, a few examples of so-called ‘meta-evaluations’ whereby the evaluation process is itself evaluated as part of a process of establishing the credibility and validity of the evaluation methods and results. Most of such meta-evaluations have been conducted by the EC. Among national examples, are the comparative studies undertaken to support the French ‘days of exchange’,⁴⁴ but a more fully developed example is **the meta-evaluation of evaluation studies in Sweden** undertaken after the last round of interim evaluations (see **Case Study 12**).

CASE STUDY 12: LESSON LEARNING FROM META-EVALUATION SWEDEN

On behalf of the Swedish Government, the Swedish Agency for Administrative Development (SAFAD) undertook a study to assess the interim evaluations of the 1994-99 Structural Fund programmes, in terms of organisation, resources and processes.⁴⁵ The research also had to cover the procurement of the evaluations and to include a comparative analysis, encompassing the experiences of other countries.

The conclusions of the research, of which an executive summary is available in English, reported that the quality of the Swedish mid-term evaluations was relatively high, although a general lack of consideration of the domestic policies in the areas concerned and a general absence of details on the methods or on the basic data used. The report highlighted a series of targeted recommendations, among which were the following.⁴⁶

- Better preparation of evaluation assignments; for this, it was deemed necessary to reinforce the evaluation skills of the competent programme managers, with seminars, training, networks etc.
- Clearer indication, in the evaluation reports of the methods and data used.

⁴⁴ See Case Study 11: Lesson Learning from Evaluation (France).

⁴⁵ SAFAD (1999), *Mid-term evaluations of the Structural Funds – executive summary in English*, Stockholm. The study can be obtained on-line from the web-site www.statkontoret.com (only in Swedish).

⁴⁶ In fact, similar considerations could be applied to the Scottish case, underlining how the major problems experienced in relation to evaluation are often similar across countries, as the international review also demonstrated.

- Major integration, in the evaluations undertaken, of the domestic policies in the area concerned.
- More adequate attention to timing: the evaluations, in most cases, should have been postponed since the programmes had just started when they were commissioned and there were only limited elements to assess.
- Need to establish an on-going dialogue with the Commission, to clarify issues and requirements. For example, the term 'impact' had been translated in various ways and several misinterpretations of the Commission's expectations took place too. Uncertainties also persisted about how to follow up evaluation.
- Increase the number of evaluators through the procurement of more, smaller assignments, since the market of evaluators was considered too narrow.
- Better use of financial resources, through the concentration of technical assistance funds for IT support, evaluation, follow ups, and to develop the related skills.
- Need to develop a database containing all data necessary for evaluation purposes, a system that should be ready at the start of the new programming period.
- Requirement to establish a co-ordinating unit, with competence in the field of evaluation. The implementation of Structural Fund programmes in Sweden took place with a one-year delay, relative to other countries. This determined an 'urgent' approach to the implementation of Structural Fund programmes within the pre-existing administrative structures. The 2000-6 programmes will be implemented through a higher degree of decentralisation, which will imply the risk of a 'new trial period' and make it 'more difficult to put experience from the current [1994-99] period to use'. A co-ordinating unit for evaluation and procurement of Structural Fund programmes is therefore seen as fundamental. The study highlights three options: an expert unit in the Government offices, not suggested because not coherent with the Swedish administrative tradition; an agency, for example, extending NUTEK mandate for the 2000-06 period; or the appointment of a committee, a temporary organisation, appointed for a fixed period (recommended solution).

Relevance to Scotland

Meta-evaluations could be undertaken after each round of evaluations to assess the quality of the studies and address issues to follow up. This activity, undertaken or commissioned by the Scottish Executive, should ideally encompass the analysis of the studies and the assessment of these by the main actors involved (programme managers, partners, applicants), to end with a discussion forum where the findings of the studies could be illustrated and debated.

In Scotland there is still some room for improvement in terms of the validation and dissemination of evaluations.

In terms of the diffusion of evaluation documents, three main actions are suggested: first, the institution of a central evaluation point at the Scottish Executive (see paragraph 5.1); second, the more formal establishment of a communications plan at the start of evaluations; and third, the identification of routes through which evaluation studies, or at least their summaries, could be more routinely published and so made more easily obtainable.

As far as validation is concerned, in Scotland, the results of evaluation studies are presented to and have to be approved by Monitoring Committees. This is already a good practice. However, in addition to this, possible meta-evaluation could bring potential benefits - either in the developed form illustrated in Sweden above or, perhaps more cost-effectively, in the more modest form of the French surveys of programme managers/main partners and evaluators (Case Study 11). This would ideally be followed by open discussion leading to concrete resolutions on the process and methods of evaluation. An important dimension of such initiatives is that they bring an external appraisal of the evaluation process, balancing and drawing on the perspectives of the procurer and the contractor.

5.7 Assessment

In an international context, Scotland has positioned itself among the leaders in terms of its approach to the organisation and conduct of evaluation.

This review of international practice has underlined the fact that Structural Fund actors in Scotland have taken a pro-active and ambitious - yet at the same time pragmatic - stance to evaluation, which has paid off in terms of the quality and utility of the studies undertaken. In benchmarking against other Member States, Scotland offers many examples of good practice in the management of evaluation. Nevertheless, the diversity of evaluation systems in the EU provides several examples from which Scotland can reinforce its existing strengths and address its acknowledged weaknesses.

Although there has been a relatively robust framework for Structural Fund evaluation in Scotland, what can be concluded about economic impact is limited by the fact that Structural Funds are only part of the wider picture of economic development. There is scope for developing new institutional frameworks which will reduce the isolation of Structural Fund evaluation.

In Scotland, within the Scottish Executive, specific (albeit modest) resources have been devoted to managing and co-ordinating the evaluation of the Structural Funds, something which has not been always the case in other Member States. However, as discussed in section 3.7, to appraise the overall

effectiveness of Structural Funds it is necessary to place the evaluation of Structural Fund programmes in a wider context. It is often difficult to separate the impacts of different Structural Funds and domestic policy measures, and any effort to understand the contribution of the Structural Funds should consider a more holistic approach. At the least, the Scottish Executive should determine the desirability and feasibility of taking such a broader perspective on Structural Fund evaluation. As part of this process, regard should also be given to whether Scottish Enterprise evaluation procedures could be more fruitfully integrated into a wider Scottish framework.

There are several institutional implications of this approach. Paramount is the need to co-ordinate different evaluations more systematically. Two ways of achieving such a cross-agency framework have been noted in this chapter:

- the establishment of a centralised *Evaluation Unit*, whose remit would be to integrate all economic and social development policy evaluations (including the Structural Funds), as demonstrated by the National Evaluation Unit in Italy (cf. Case Study 2); and
- a ‘softer’ approach, involving the *‘nesting’ of monitoring and evaluation* at different levels, with agencies with policy development interests becoming more directly involved through partnership-based working groups, as has been undertaken in the West Midlands (cf. Case Study 3).

Scotland has systematically carried out all required ex ante, interim and ex post evaluations in compliance with the regulations. Nevertheless, the quality and effectiveness of evaluations could be improved by further developing a more integrated and on-going approaches to the cycle of Structural Fund evaluations.

Currently, the evaluation of Structural Fund programmes is undertaken as a series of separate, period-specific exercises, albeit using common guidelines and terms of reference and a similar approach to the evaluation of some programmes (given the reliance on a limited group of evaluators). There is considerable scope for organising a more integrated programme of Structural Fund evaluation in Scotland, with (see section 5.2):

- a more explicit *link between different phases of evaluation* (*ex ante*, *interim*, *ex post*);
- more use of *combined programme and thematic evaluations* in a diagnostic process, allowing programme-specific and minority issues to be considered in more detail (as in North-Rhine Westphalia in Case Study 4);
- explicit *comparative evaluation research* across programmes (as in Austria, as described in section 5.2); and
- use of *meso-level evaluation techniques* to explore programme elements in greater detail (as in Yorkshire and Humberside in Case Study 5).

Although evaluation of process issues is strong in Scotland, there is scope for strengthening its role by using evaluation outside its stipulated phases in the programming cycle.

The evaluation of *process issues* is already a strong aspect of Structural Fund evaluation in Scotland. As with the PME review conducted recently in Scotland, process evaluation can provide deeper insights and understanding into how administrative functions actually operate and inter-relate. However, process evaluation could be improved by involving evaluators in-between evaluations, employing an approach to evaluation which is more integrated into the programming cycle. The activity of the evaluator would be interpreted not simply as evaluations in the programming cycle, but more extensively as support to programme management. On a continuing basis, this would allow an on-going identification of procedural bottle-necks and structural weaknesses, specific recommendations on different stages in the design and delivery of individual measures, and establishment of greater trust between evaluators and programme managers. The example of Tuscany in Case Study 6 displays the benefits of this approach.

There is value in enhancing the involvement of partners and other stakeholders in all phases of evaluation exercises.

The involvement of partners is one of the strengths of the Scottish approach to Structural Fund implementation. Partners are active in programme management at all stages, from the design of programming documents, through project selection, to the validation of the decisions adopted, via their participation in the Management and Monitoring Committees. Over successive programming periods, this has led to progressively greater partner activity in the evaluation process and a gradual (if partial) recognition of the potential benefits of evaluation as a programme management tool and mechanism for feedback.

However, with regards to evaluation in particular, further improvements could be introduced. As seen in Finland (as reported in Case Study 7), a more explicit planning of partner involvement – allocating roles and objectives – at different stages in the process could be introduced successfully. It is suggested therefore:

- that the PMEs institutionalise *evaluation Steering Groups* with the participation of partners on an on-going basis - these Groups should meet regularly, not only in the proximity of evaluation exercises, with the roles of different actors clarified in a more transparent way; and
- over the longer term, that a more *permanent dialogue* on evaluation issues be introduced, perhaps through a sub-group of the proposed Scottish Structural Funds Forum or the Scottish Co-ordination Team, which would focus more exclusively on evaluation issues – this could draw on good practice from the EU, such as the Finnish ESF Steering Group (as described in Case Study 6).

The evaluation community in Scotland is deemed as well-prepared and responsive, but also too restricted in terms of numbers and diversity. Enlarging the community of evaluators will ensure higher quality and greater understanding of the potential role of evaluation over the long term.

Scotland has some of the leading Structural Fund evaluators in the UK (and indeed the EU), but the evaluation community is small, sometimes making it difficult to generate interest and involvement in evaluation outside a small group of private sector consultants. This has largely been due to the specialist nature of Structural Fund evaluation, as well as the time constraints under which evaluations have to be carried out. At present, little is done to invest in developing the evaluation community: interaction between those commissioning studies and potential evaluators is limited, and there is no central resource support for Structural Fund evaluation. Even the Scottish Executive lacks a complete library of Scottish programme evaluation studies and relevant evaluation literature (eg. the output of the MEANS programme, evaluation books and articles). At the PME level, furthermore, often little time can be devoted specifically to evaluation, as their often passive re-proposal of the terms of reference suggested by the Scottish Executive suggests.

Enlargement of the evaluator community in Scotland can be difficult because of the absence of official journals for making calls for tender (as, for example, in Italy and Spain) and suitable alternatives for publicity. However, the Scottish Executive could establish a register of available evaluators – with systems for extending the register over time and guaranteeing its transparency – which would encourage greater evaluator participation. Moreover, the register could be expanded to include evaluators from outside of Scotland. This could be usefully linked to the recommendation for the establishment of a web-site, as discussed in more detail below.

Capacity building should also include greater investment in increasing the range of evaluation skills held by the Scottish evaluation community.

The Scottish evaluation community has some skills gaps, particularly in the evaluation of particular themes or horizontal themes within the Structural Funds, such as social inclusion. The Scottish Executive and programme managers should invest in developing capacities for Structural Fund evaluation by promoting:

- more awareness, interest and involvement in Structural Fund support by developing a publicly accessible resource library on evaluation, and a ***point of reference*** for practitioners and partners on evaluation, similar to the Austrian Checkpoint Eva (as noted in Case Study 10); and
- ***exchange of experience*** and lesson-drawing, before and after evaluation exercises, similar to the ‘days of exchange’ organised by DATAR in France (in Case Study 11).

In particular, as regards the *Evaluation Point* – which could be co-ordinated by the Scottish Executive, but possibly out-sourced - it could include:

- a comprehensive, organised and open *library* on evaluation, encompassing evaluation documents (at least the Scottish ones, but possibly also UK and other countries), developmental documents (such as MEANS materials, EC guidance, methodological papers) and conference materials;
- a *web-site* with:
 - documents and literature available on-line
 - a list of evaluators and their profiles, as described above - this would not be a ‘closed’ register, but an easily accessible list with on-going submissions by the consultants to the office responsible for the maintenance of the evaluation point and its web-site
 - the calls for tender regarding programmes and thematic evaluations, for Structural Funds as well as domestic policies
 - a ‘what’s new’ page, with the announcement of forthcoming conferences, seminars and other similar initiatives, and highlighting of new documents/information available on the site
 - links with the evaluation societies and other relevant institutions (eg. the inforegio web-site) and academics
 - an interactive window, where users could submit queries to the staff of the evaluation point
- permanent, though not extensive, *staff*.

The ‘*days of exchange*’ could be organised by the Scottish Executive after each round of evaluation and the conduct of studies to assess the strengths, weaknesses and quality of the evaluations undertaken. Such meetings would be open to partners and stakeholders and publicised broadly. To make possible the creation of a constructive debate, the organisation of small thematic working groups and a final plenary session would help in managing the high number of participants.

A lot of work has been conducted in Scotland on the availability and reliability of data. However, further improving the quality and range of data is essential.

Although the quality of data has been to date more than acceptable, further work should be undertaken in several areas.

- ***Re-organise the monitoring framework*** and set of indicators in a coherent and integrated framework for all Structural Funds, especially in consideration of the new role that Scotland will have on ESF. The accurate design of an on-line monitoring system, such as the Finnish FIMOS, for all Scottish programmes would simplify the monitoring and evaluation activities significantly (see section 5.6).

- ***Train project applicants*** on the monitoring requirements they are asked to fulfil. This activity is already on the agenda of the PMEs, but within a centralised monitoring system, guidance meetings with all PMEs could be organised by the Scottish Executive to provide them support in identifying the approach to follow in this training activity.

Disseminating evaluation studies and findings regularly and to a broad audience is critical, particularly in going beyond the partnerships.

Evaluation findings in Scotland have not been diffused adequately. Broader dissemination of evaluation documents and results will contribute to promoting interest and strengthening the credibility of evaluations. It would also enhance the possibility of building a more democratic consensus around programmes and focusing future interventions on the needs of beneficiaries and other stakeholders.

Evaluation studies should be actively publicised and disseminated by programme managers beyond the programme partners in order to promote interest, awareness and understanding of Structural Fund evaluation methods, processes and results. This could be facilitated:

- at the programme level, by the design, for each Structural Fund programme evaluation, of a basic '***communications plan***', to ensure appropriate dissemination to different target audiences; and
- at the Scottish level, by the establishment by the Scottish Executive of an ***Evaluation Report series***, similar to that of DG Regio or those of the governments of Ireland and Finland, containing all Structural Fund evaluation studies as well as relevant reports on methodological issues.

The functioning of the evaluation process should be independently assessed once a group of evaluations has been completed, either by commissioning meta-evaluations, or more informal surveys of the main participants in the group of evaluations. Such exercises enable lessons to be learned and shared, especially if they are followed by open debate.

To date, no formal assessment has been undertaken on the quality and effectiveness of the evaluation studies commissioned. A meta-evaluation, similar to the one undertaken in Sweden (cf. Case Study 12), and subsequent discussions would improve the quality of evaluation studies, contribute to greater focus on new themes for future evaluation activity and raise the degree of competence of practitioners and other involved actors. It is suggested that:

- the Scottish Executive undertake or commission, after each round of evaluations, ***second-degree evaluations***, to identify the strengths and weaknesses of the evaluation reports, the degree to which they have been exploited and new themes for future evaluation exercises; and

- the outcome of these suggested meta-evaluations be discussed by civil servants, programme managers, partners, evaluators and other interested parties to encourage learning and lesson-drawing among the wider community – one option for facilitating debate and exchange of opinions could be the organisation of *open discussion meetings*, similar to the French ‘day of exchange’.

6. INTERNATIONAL REVIEW AND CASE STUDIES: METHODS OF EVALUATION

The earlier critique of Structural Fund evaluation methods in Scotland noted that interim and *ex post* evaluation studies of the Scottish programmes have generally adopted ‘bottom-up’ approaches to evaluation, often based on surveys of programme partners. Some, but not all, studies included a ‘top-down’ element with surveys of the overall regional economic context or economic performance of the area. Issues such as additionality and deadweight were not considered in depth and, until now, there has been no real assessment of the impact of Structural Fund programmes.

It was suggested in Section 3 that methods of Structural Fund evaluation in Scotland would benefit from consideration of the following issues.

- Improvements in the methodological rigour of the evaluation design and execution.
- More serious evaluation of the impact of Structural Fund programmes.
- The identification of specific methodologies to address distinctive issues more roundly and to place greater emphasis on the quality of outcomes and impacts as well as their quantity.
- The evaluation of Structural Funds co-financed activity in a Scotland-wide context.

The following section reviews international experience with methods for evaluating Structural Fund programmes and presents case-study information that could help to address the above issues.

The studies considered embrace both ‘official’ evaluations of programmes and those undertaken by academic experts. It is worth noting, however, that this section does not provide a comprehensive literature survey of all evaluations of Structural Fund programmes in Europe. An overview is provided but the primary focus is to highlight best practice in the light of Scottish experience. The section is in four parts. The first part discusses some of the key issues relevant to the evaluation of the impact of the Structural Funds. Thereafter, we summarise under each heading the issues arising from the Scottish experience, compare with international experience, highlight best practice and consider its relevance to Scotland. Issues surrounding the application of bottom-up approaches are considered first and this is followed by a discussion of the application of top-down approaches.

6.1 Evaluation issues

At the outset, it is useful to consider the key issues relevant to evaluating the impact of the Structural Funds. It is also of value to outline a conceptual structure within which the various broad approaches to evaluation can be compared. In Figure 2.1 of the Commission document, *Evaluating EU Expenditure Programmes*⁴⁷, Nagarajan and Vanheukelen represent the

⁴⁷ Nagarajan N and Vanheukelen M (1997) *Evaluating EU Expenditure Programmes: A Guide. Ex Post and Intermediate Evaluation*, Commission of the European Communities, Brussels p. 25.

relationship between needs, objectives, inputs, activities, outputs, results and outcomes as follows.

$$N \rightarrow [B \rightarrow I \rightarrow A \rightarrow P] \rightarrow R \rightarrow C$$

They identify that part of the path in between the square brackets, that is the part of the path linking *objectives* (B) to *inputs* (I) to *activities* (A) to *outputs* (P), as being within the domain of the programme. However, the link between *needs* (N) and *objectives* (B) at the beginning of the process, and between *outputs* (P) and *results* (R) and *outcomes* (C) at the finish, is in the domain of socio-economic problems. The distinction that is made here between these two domains is important.

6.1.1 *The domain of the programme*

Beginning with that part of the process that is in the domain of the programme, this has two important characteristics.

- The evaluation of this part of the programme can be undertaken without considering the overall *utility* of the programme as a whole. In particular, where a programme's objectives are expressed in terms of outputs, then project *efficiency* can be measured as a ratio of outputs to inputs. An example would be cost per gross job.
- If it is common knowledge that an evaluation process is to take place, then this in itself can change the efficiency of the programme currently in progress.

Essentially, the part of the evaluation process that occurs in the domain of the programme is in the nature of monitoring. It improves the delivery of the programme by reducing the initial asymmetry in information between programme funders and providers. In so far as future funding can be adversely affected by delivery identified to be sub-standard, the existence of evaluation should improve present delivery. Also, it can be used as a management tool if implemented at the project level in that it identifies those types of project which are relatively efficient. Therefore, if the programme's objectives are expressed in terms of outputs, then the programme manager can use the evaluation information to augment the appraisal process for individual projects.

6.1.2 *The domain of socio-economic problems*

In order to evaluate the effectiveness of a programme, or its utility or sustainability one requires to know how the programme outputs translate into impacts on the wider economy and society at large. In the terminology employed in the Commission document, *Evaluating EU Expenditure Programmes*, initial impacts are known as results, longer term impacts as outcomes. This information is also needed to evaluate the utility of a programme, that is the extent to which programme outcomes meet economic or social needs.

The central point to make here is that once we move into the domain of socio-economic problems, evaluation necessarily implies the imposition of a particular theory of the operation of the local economy. This theory might

well be implicit. However, it must be there. It is simply impossible to do an atheoretical evaluation of a programme's impact on the relevant economy. As soon as one moves into the domain of socio-economic problems, the identification of the appropriate model of the region and that model's parameterisation is a crucial issue for evaluation because reliable models of the regional economy are generally difficult to come by. This issue is picked up in the discussion of both bottom-up and top-down models.

A second important point is that if the utility of the programme is to be assessed, the evaluator needs to know the relevant Social Welfare Function against which the impacts of the programme are to be measured. Where the needs targeted by the programme are limited, for example where the programme is designed primarily to increase employment, this is not a serious problem in principle. However, where a wider range of needs, such as sustainability and equal opportunities, are to be addressed in the single programme, the problem becomes more pressing. A single measure of programme effectiveness involves an explicit or implicit weighting of the needs of the society or economy.

6.1.3 *Alternative broad approaches*

The *MEANS Handbook No. 3*⁴⁸ suggests two overarching methodological approaches (bottom-up and top-down) to the evaluation of impact (particularly on employment) of the Structural Funds. Moreover, the *MEANS* guidelines favour the application of a bottom-up approach to the estimation of gross employment effects. In our review and critique of the approaches adopted in Scottish Structural Fund evaluation, we noted the predominance of bottom-up approaches. A few studies contained a top-down element but they largely amounted to surveys of the economic context in the area within which the programme operated. However, in the final evaluation of the earlier Structural Fund programmes for 1989-93, statistical analysis, in the form of the shift-share technique, was also used in an attempt to control for non-programme influences on overall area employment performance.

Our research suggests that, in the other EU countries, if there is an overall pattern, it is that top-down evaluation is most prevalent where the largest programmes are operating, eg. the nation-wide Objective 1 programmes of the cohesion countries. Bottom-up approaches are frequently relied on more in those Member States where programmes are predominantly small and fragmented (eg. the UK, Denmark, Germany, France and Sweden). This is not to deny that there are some top-down assessments in these countries, but both bottom-up and top-down approaches appear to be used more frequently, although not necessarily in the same evaluation, in Italy, Finland – for EDRF, but bottom-up for ESF - the Netherlands and Spain.

a. *Bottom-up evaluations*

In reviewing Structural Fund evaluations across Europe, we did not find any that were wholly 'bottom-up'. By this we mean that there were no methods which calculated aggregate effects by simply summing the measured impact

⁴⁸ MEANS (1999) *Evaluating Socio-economic Programmes. Principal Evaluation Techniques and Tools*, MEANS Collection, volume 3, p. 24.

on individual aided and unaided individuals or firms. Rather, what is meant by a 'bottom-up' approach might be better called a hybrid approach. The direct outputs of the programme are identified by some microeconomic analysis. As outlined in Chapter 3 for the Scottish evaluations, this usually means adopting some form of industrial-survey method using information derived directly through interviews and/or questionnaires. This direct output is then transformed so that it can be treated as an exogenous disturbance in a model of the region to calculate net impacts.

The advantages of using the bottom-up approach for measuring the outputs of the programme are that:

- it provides a better opportunity to trace through the routes which each measure, or individual programme component, contributes to the overall programme output; and
- some programme elements are difficult to formally model.

Disadvantages are that:

- there may be double counting if two elements of the programme combine to generate a particular output; and
- synergy effects might be missed by concentrating on the components of the programme rather than the whole.

In the measurement of net employment impact, the MEANS guidelines argue that to move from gross jobs to net jobs requires the identification and quantification of additionality, deadweight, substitution and displacement effects. Where evaluations attempt to calculate net employment effects (as well as gross effects) these also generally incorporate multiplier effects too. (See Ernst and Young *Evaluating Structural Funds Employment Effects* for the prevalence of these practices and Table 3.2 and 3.3 for the use of these methods in Scottish evaluations). These processes affect the size of the link between programme outputs and impacts - that is, the steps between $P \rightarrow R \rightarrow C$ in the schematic representation of the evaluation process above, as one moves from the domain of the programme to the domain of socio-economic problems.

Additionality and substitution measure the extent to which the programme effects the behaviour of the direct beneficiaries. Deadweight quantifies the extent to which aid is above the minimum that would have been necessary to achieve that change in behaviour. Whilst it is not always made explicit how additionality and displacement have been calculated, it is clear from our discussion with evaluators in Scotland that an implicit, very simple, regional export-base model is being used. The net employment impact associated with the programme (E_N), is derived in the following way. First the direct employment associated with additional regional exports is added (E_X) and this value is then multiplied by the marginal employment export-base multiplier (m_E) so that:

$$E_N = E_X m_E$$

The employment required to produce the additional regional exports generated by the programme, E_X , is calculated by adjusting the gross employment to

accommodate additionality, substitution and displacement effects. Additionality and substitution measures the extent to which the programme affected the employment behaviour of firms directly benefiting from the programme. Displacement measures the negative impact on employment in other firms in the target area. But displacement is measured only as product market displacement: the labour market is taken to react passively as is characteristic of traditional Keynesian export-base models. Also, the only output that is thought not to displace existing employment is where export markets are being supplied (or import substitution is taking place).

The multiplier value is a standard Keynesian consumption/Input-Output (I-O) multiplier. Because typically I-O tables do not exist for the relevant policy regions, these multipliers are often rather crude but they are based upon simple demand-driven models.

There are a number of generic problems faced when using the simple export-base, demand-side model in attempting to calculate the economic impacts in the conventional bottom-up approach.

- Those areas receiving Structural Fund assistance are generally thought to have supply-side problems.
- The policies pursued under the Structural Funds are supply-side policies.
- The passive supply side that characterises these models is not regarded as acceptable in national macro-models.

This means that the impact on the economy is not evaluated through the modelling of the supply-side effects that the programme is attempting to bring about (the impact of changing efficiency, competitiveness, relative factor prices, transport costs etc). Rather, calculating the employment impact of the Structural Fund policies requires that the supply-side stimulus has to be converted in an *ad hoc* way into a demand-side stimulus, and then a straightforward demand-side multiplier applied. However, it is known from work with the Scottish Computable General Equilibrium model, AMOS, both that the employment multiplier associated with supply-side policies can be quite different from those stimulated by demand-side policies and that the multipliers in supply-constrained economies will differ from those in economies not experiencing such a constraint.⁴⁹

b. Top-down evaluation

Top-down evaluations do not go through the process of tracking the effects of individual projects, aggregating to get the programme outputs and then calculating impacts by multiplier and displacement adjustments. Rather, these methods attempt to identify impacts at a more aggregative level. One broad method is to construct a counterfactual position. This means identifying the level of activity that would have occurred in the area without the Structural Funds. The difference between actual and the counterfactual is then attributed to the impact of policy. There are two ways of constructing the counterfactual:

⁴⁹ Armstrong H and Taylor J (2000) *Regional Economics and Policy*, 3rd Edition, Section 2.4, Harvester Wheatsheaf, New York.

- choose a comparator, non-aided region that is identical in all important respects to the region in receipt of Structural Funds; and
- model the behaviour of the region before the region received Structural Fund assistance and then run the model forward into the period where funding operates.

With reference to the evaluation path that was outlined at the beginning of this section, it is clear that this counterfactual method skips two steps. It is useful to know the inputs used in the programme so as to be able to measure efficiency. However, this approach does not separately identify activities or outputs and focuses wholly on impacts. As such, it does not go through the process of calculating gross jobs and then adjusting for additionality, displacement, substitution and multiplier effects. Rather, it focuses solely on the net effect. Theory is clearly involved where the counterfactual position is provided by a model. However, even where the counterfactual is derived through a comparator, non-aided area, theory is required to identify the key characteristics that need to be matched.

There are a number of weaknesses with the counterfactual approach:

- the accuracy of the counterfactual prediction;
- the lack of detail on the impacts of individual elements of the programme;
- as with all 'residual' identification techniques, any errors in the data or the model get attributed to policy effects - if the expected policy impacts are small, their estimation through this method is likely to contain a large percentage error; and
- the counterfactual approach can only be used for *ex post* evaluations.

A second type of top-down method seeks to formally model the policy effects themselves. This implies that the section of the evaluation path identified as being in the domain of the programme is formally modelled. However, again, the focus is almost certainly the impacts of the programme so that separate additionality, displacement and multiplier figures will not be calculated or required. The explicit modelling of policy effects is conceptually more satisfactory than the corresponding counterfactual approach. Moreover different elements of the programme can be separately modelled and such a modelling approach facilitates *ex ante* evaluation. Attempts can be made to simulate the *ex ante* impacts of completely novel policies with this kind of a model. However, there are again difficulties with this form of evaluation:

- cost and data requirements of building a formal model;
- the need for the model to be able to deal with supply-side effects; and
- the ability to explicitly model many of the policy measures associated with the Structural Funds.

A final top-down approach is one that deals with policy effects through direct econometric estimation. This could be in the form of a single equation or a system of equations. From a methodological view, economists normally favour formal econometric approaches that test the statistical significance of the impact of independent variables on dependent variables. In this case, the key independent variables are the policy variables and the dependent variable

would be a measure of policy impact, for example employment. The policy impact is determined through using the econometrically estimated coefficients on the policy variables to calculate the effect of reducing these policy variables to zero. If data are available, this approach should calculate the impact of policy with the greatest degree of accuracy. Problems here are:

- data availability;
- the simultaneous operation of a large number of projects within a programme whose effect might be difficult to identify individually;
- the small overall impact of the Structural Funds might not be picked up by these econometric methods; and
- this method can only be used for *ex post* evaluations.

6.1.4 *Complementarity*

Perhaps the main point to make is that the designation of top-down and bottom-up suggests too great a gap between the various methods. As is argued above, so-called bottom-up approaches in practice are hybrid approaches where detailed micro and ‘industrial survey’ methods have to be embedded in an overall model of the regional economy. That such a model is rather rudimentary does not make it any less a model. Similarly, top-down models are likely to have difficulty modelling the way that policy operates. For example, where policy takes the form of a uniform subsidy to all firms in a particular sector, the operation of the policy is reasonably easy to model. However, discretionary policies such as business support, innovation grants and assisted training are more difficult to capture in a set of formal equations. One possibility is to embed the bottom-up evaluation in a rather more sophisticated regional model. At the very least, the top-down and bottom-up approaches should be seen as complementary, not competitive.

6.2 **European comparative overview**

The purpose of this part of the report is to examine European experience in the application of methods to the evaluation of the Structural Funds. We are particularly concerned to identify the issues arising from Scottish experience, consider the European experience, highlight best practice and assess its relevance to Scotland.

6.2.1 *Bottom-up approaches*

In principle, as noted in the first part of this section of the report, bottom-up approaches seek to trace the consequences of programme measures and activities, through:

- the gross effects (eg. supported jobs, or new firms assisted),
- to the net additional effects (after allowance for additionality, substitution and deadweight),
- to the overall net impact on the targeted results and outcomes in the socio-economic domain of the programme area (after allowance for displacement and positive secondary effects in both the short and long run).

Each of these is now at looked in turn.

a. *Gross effects*

Section 5 of this report noted that a key requirement for the identification of gross effects is the availability of appropriate data at the project and beneficiary level. We noted in Section 3 that *in Scotland investment in Programme monitoring systems has significantly improved the quantification of Programmes and the consistency of data available.*

In most European countries, monitoring systems have also improved over the period of the Structural Fund programmes, but, as in Scotland, there is still further to go.

Section 5 also provided examples of best practice monitoring systems elsewhere in the EU. We noted in Section 3 that in Scotland, despite significant improvements, monitoring systems were not always sufficiently comprehensive to enable the processes or routes to impact of each measure to be tracked through appropriate indicators and data. So, for example, the impact of a project on the employment of local residents cannot be established if the domicile of those obtaining jobs is not recorded.

This establishes the general point that *an ideal monitoring system should provide a rich source of project/beneficiary information relevant to all the priorities and targets of the programme.*

So, for example, if one of the priorities or targets of a programme is job creation, then all projects and assisted activities should provide job creation - or, at the very least, jobs-associated information. The same argument applies to other objectives such as: equal opportunities, where data on the gender balance of jobs should be collected; sustainability, where *inter alia* data on expected and actual job duration, as well as job forecasts, is required; and social inclusion, where information is at least required on whether the person gaining a job was drawn from unemployment.

In view of the importance of the quantification of targets and the existence of an appropriate monitoring function to effective evaluation we come to the following conclusion that

- there is a case for ensuring that an analysis of assessability is properly incorporated in the prior appraisal of Structural Fund programmes.

The Commission requires that this type of assessment is incorporated in the prior appraisal, but European and Scottish experience suggests the shortcomings of monitoring for evaluation purposes are only fully realised when interim and even final evaluations are undertaken. Prior appraisals have tended to focus more on the fit between the programme priorities and the socio-economic situation in the programme area than the appropriateness of the functional and organisational arrangements for evaluation. The following Case Study provides a best practice example from southern Spain of an analysis of assessability.

CASE STUDY 13: ANALYSIS OF ASSESSABILITY - SPAIN

Avila Cano and Cirera undertook an analysis of assessability in the context of the intermediate evaluation of the 1994-99 operational programme in the Objective 1 region of Andalusia (as reported in a paper presented to the *European Conference on Evaluation Practices in the Field of Structural Policies*, Seville, 16th and 17th March, 1998). The analysis examined the quality of programming, the information system and management, the three factors upon which the possibility of evaluating a programme depends. The appropriateness of the three factors was examined through a combination of desk analysis and interviews with officials involved in implementation. Specifically, the suitability of the three factors was assessed as follows.

- *Quality of programming*: an assessment of the rationality between needs and programme objectives as well as the internal and external coherence of the objectives.
- *The information system*: an assessment of the 'quality' of proposed indicators, data collection methods and the monitoring system as a whole.
- *Management*: an assessment of the quality of human resources allocated to monitoring and evaluation.

The outcome of the analysis was that the following deficiencies were highlighted:

- shortcomings of the information system;
- inadequately defined indicators; and
- a failure to distinguish between output and impact indicators.

Critique

The advantage of a formal assessability exercise is that it ensures that the requirements for proper evaluation are considered early in the programme and allows for the possibility of change to programme indicators, data collection methods and management before the formal evaluation process begins. The weakness of the Case Study is that the exercise involved significant subjective elements, so that it is not clear how, for example, 'quality' is assessed.

Relevance to Scotland

This has limited relevance for Scotland in that general perception is that the assessability of Programmes through monitoring procedures is not a major problem, though, as noted above, there is always room for some improvements.

Given a suitable monitoring system then the estimate of the gross effects of a programme becomes feasible, although subject to the pitfalls that project-based data may produce double counting and ignore synergy effects. From our desk research and interviews the following conclusions can be drawn on the estimation of gross effects at the *programme* level.

- In Scotland, not all evaluation studies tend to establish the gross effects at the programme level, with the line often being drawn at the measure level.
- The experience of France and Spain appears to be much worse than Scotland where programme effects were infrequently calculated, if at all.
- Austrian, German, Swedish and rest of UK studies generally did aggregate to the programme level.

As in Scotland, difficulties in the availability of appropriate monitoring data have led evaluators elsewhere in the EU to undertake surveys of project managers and beneficiaries in an attempt to establish gross effects, although this is not the case in all countries. In France, for example, of the 15 evaluation studies reviewed in Ernst & Young, only two - Bourgogne and Region Centre – undertook survey work to estimate employment effects in greater depth. Similarly, German evaluations appear to rely primarily on monitoring data. But our interviews reveal that in Denmark, Finland, Spain and Sweden survey work is given greater weight, while according to Ernst & Young “... UK evaluators appear to have placed more emphasis on survey work than in other countries covered by the research”⁵⁰. It is not clear though whether the greater emphasis on surveys in the rest of UK Structural Fund evaluations is a product of necessity, reflecting the deficiencies of programme area monitoring systems, or a desire to put together a richer information set. However, one thing is certain, the collection of a range of data through surveys allows a more varied description of gross effects.

Specific comments can be made about the treatment of different types of gross effects in Scottish and other UK studies:

- They generally fail to distinguish between permanent and temporary employment⁵¹, despite UK Government guidelines offering more detailed guidance than MEANS on how the durability of employment effects can be measured.
- Scottish studies make little distinction between jobs of differing quality, despite suggested measures in the UK guidelines.
- The sectoral breakdown of gross outputs is also often ignored, which may be an important indicator of sustainability, degree of local integration and a factor influencing the extent of displacement.

Ernst & Young cite several German studies as offering best practice examples of more sophisticated approaches to the identification of types of gross effects.

- On the permanent/temporary effects issue, the Brandenburg interim evaluation undertook survey work to establish whether jobs created by Structural Fund assistance still existed at the time of survey.
- In Saxony, employment effects were translated into ‘job-years’ – jobs existing over the lifetime of projects - as a way to measure durability.
- ‘Meso’ level analysis was also a common feature of the German Objective 1 1994-99 interim evaluations involving:

⁵⁰ Ernst & Young (2000) *op. cit.* p. 18

⁵¹ Ernst & Young (2000) *op. cit.*

- sectoral distributions of gross effects;
- spatial distributions by sub-regions within priority areas; and
- size-band distribution of job creation in assisted firms.⁵²

In the UK, the *ex post* evaluation of the Thanet Objective 2 programme took projections of construction jobs from programme monitoring data converted them into full-time equivalents, and then into years of work. The number of ‘permanent’ jobs was then estimated using the UK Treasury assumption that such a job comprises ten years’ worth of employment. Back in Germany, the Saxony evaluation provides a detailed sectoral and company size breakdown of gross job creation, while the Berlin interim evaluation provided a sectoral disaggregation of new jobs. Finally, overview work undertaken by the IFO Institute in Munich, noted above, provides a good example of an analysis to compare the capital intensity required to create new jobs in different sectors, with reasons for variations, using data from the monitoring records of the five Objective 1 areas⁵³.

All of which leads us to conclude that *the identification and measurement of different types of gross effects is one area where Scottish practice can learn from EU experience.*

b. *Net additional effects*

In Section 3, we reported on the treatment of additionality and deadweight within the Scottish evaluations. We found little direct attempts to estimate additionality at both interim and final evaluation stages. Only one Scottish study, *Interim Assessment of the Highlands and Islands 1994-99 Objective 1 Programme*, undertook a detailed survey of projects where questions were asked on whether the project would have gone ahead, wholly or partially, without programme funding. In the remaining studies, evaluators either used ‘benchmark’ data derived from a survey of Objective 2 programme experience or other sources, or simply used their judgement. However, many projects receiving Structural Fund support in Scotland had Scottish/Highlands & Islands Enterprise and their network partners as principal sponsors. We believe that most of these projects were subject to fairly rigorous internal evaluation using, for example, the Output Measurement Framework. These prior evaluations were of considerable benefit to evaluators in forming their judgements about additionality and deadweight. However, in general terms, the basis for the evaluator’s judgement tended not to be explained, nor the relevance and applicability of the chosen benchmark data.

Nevertheless, our research suggests that programme evaluations elsewhere in Europe appear to treat the need to estimate additionality less seriously than in Scotland.

In the Ernst & Young review of Structural Fund evaluations, 75 studies were considered. Of these, only 18 percent of the 1989-93 studies made adjustments for additionality, while only 36 percent of those covering the 1994-99 period did so. Of course, the issue here is not just whether

⁵² Ernst & Young (2000) citing IFO (1997).

⁵³ IFO (1997).

adjustments for additionality were made but the quality of the adopted adjustment.

Ernst & Young's analysis of comparative European experience suggests that official evaluation studies in the UK, Germany and Sweden are more likely to attempt to adjust for additionality than their counterparts in France and particularly in Spain. This finding is confirmed by our own interviews with European experts, from which it appears that the northern European countries, including Finland and Denmark, which were not covered in the Ernst & Young study, are more likely to seek to adjust for additionality.

But what is the best approach to assessing additionality? European experience suggests that there are four alternatives that might be adopted in the context of a bottom-up approach:

- sole reliance on monitoring data;
- the application of coefficients or benchmarks;
- the use of control groups; and
- the adoption of survey techniques.

Reliance on *monitoring data* alone is clearly problematic. In principle, all projects should be additional, so the monitoring data in itself is not sufficient. The evaluator can use the data to weed out double counting, by bringing together complementary projects, but in the last analysis, conclusions about additionality rely on the evaluator's judgement, a judgement that should be made explicit.

The explicit application of *coefficients or benchmarks* is a feature of some Scottish studies, as noted in section 3, and the approach has also been adopted elsewhere in Europe. One of the best examples, which covers the application of coefficients for secondary effects as well as for additionality, is the Evaluation of the Yorkshire and Humberside Objective 2 Programme 1994-96. This Evaluation was presented earlier as Case Study 4.

As we noted in the first part of this section, well-chosen benchmarks can play a useful role in evaluation studies. However, at best, application of such benchmarks can only provide a general guide and, given that they are drawn from widely differing measures, projects and areas, are unlikely to be sufficient to accommodate the particularity of the assisted projects or the economic circumstances of the programme area. Further, the use of benchmarks for additionality clearly does not identify the skill or expertise of the Programme Management Executive (or the relevant programme committee) in the initial screening process for choosing projects suitable for Structural Fund aid. Moreover, this procedure was regarded as problematic in a number of the interviews with consultants and programme managers in Scotland.

The use of *control groups* has tended not to feature in Scottish evaluations, with the exception of some earlier applications of shift-share analysis, which offers a quasi-control group approach. There are some examples of their use elsewhere in Europe but the adoption of such an approach is equally rare. The Ernst & Young review cites the evaluation of the Objective 1 industrial OP in

Ireland as the only example of control group methods being applied to identify ERDF net effects. However, the review notes that good examples of such methods to assess ESF employment effects can be found in the Objective 3 interim evaluations in the Netherlands and the UK.

In Ireland, the interim assessment of the 1994-99 Objective 1 CSF involved the development of a 'Regional Industrial Employment Differentials' (REID) index to assess the net contribution of Structural Fund measures to employment creation. The index was developed to measure the employment differential between firms assisted under the Industry Operational Programme and non-assisted firms. Essentially, the employment performance of the assisted firms is compared to a matched sample of non-assisted firms. Differences in employment performance are ascribed to the Structural Fund assistance. The approach can be applied at the sub-regional level. Application to the Irish OP found that employment in assisted firms rose by 11 percent between 1993-95, while employment in non-assisted firms fell by 7 percent.

One of the difficulties in implementing the control group approach is not only to construct the 'control' so that it matches the 'experimental' group as well as possible but also to ensure that the control is truly independent. That is, it should not be influenced by the assisted behaviour of the experimental group. Clearly, if the control is subject to substitution, displacement and positive secondary effects, then the usefulness of the exercise is much reduced. In the first part of this section, we noted that the selection of a comparator area or group requires a theory to select the key characteristics that need to be matched. The relevant attributes will vary according to the type of group to be matched but in general terms the control should be:

- independent of the experimental group; and
- similar to the experimental group in those attributes that are likely to be a factor in mediating the effects of the assistance provided by the Structural Funds - so, for example, in the payment of financial assistance to firms, one should probably control for sector or product, size, and capital intensity, since these characteristics are likely to have a strong bearing on the eventual outcome.

An interesting example of the control group approach as applied to ESF assistance in the UK is given in Case Study 14.

CASE STUDY 14: USE OF CONTROL GROUPS TO ASSESS THE IMPACT OF ESF FUNDING - UNITED KINGDOM

The Pieda interim evaluation of the UK ESF Objective 3 programme, 1994-96, is a good example of the measurement of the net impact of ESF projects, where the recipients are often not firms but individuals. For part of the evaluation of net impacts, Pieda made use of a control group, where participants in one of the programme-supported schemes were compared with a control group of non-assisted individuals with 'similar' characteristics. The control group was constituted using the JUVOS database, which provides a continuous record for a group of claimant unemployed at particular junctures.

Drawing two cohorts of people who had been unemployed for one year at two different points in time for each cohort, the control group could be tracked in terms of their employment history and compared with the performance of those receiving training through the ESF scheme. The use of two separate time periods for the control group cohorts allowed for the effects of varying national unemployment circumstances to be taken into account. Two comparison dates were used to compare the control groups with those benefiting from the training: six months and twelve months after completion of the training programme.

Critique

It is not clear how well members of the control group were matched to the control group. Moreover, the transferability of such an approach to the programme area level is limited by the absence of appropriate databases, which make it much more difficult to identify a suitable control group.

Relevance to Scotland

This is relevant in so far as ESF will now be evaluated at a Scottish, rather than UK-wide, basis.

Survey techniques have been used in Scotland in an attempt to measure additionality. However, great care requires to be taken in the application of this approach. We noted in Section 3 that surveys of programme partners in Scotland were not truly bottom up, in that each programme partner was often associated with several projects. Ideally, the survey should be undertaken at the project and/or beneficiary level. Care should also be taken over the sampling procedure to be adopted eg. the sampling fraction, whether the sample is to be randomly selected or stratified, or both, and if so on what criteria. Something more is required than simply contacting a number of sponsors, project managers or firms. Finally, there is the issue of the medium through which the survey is conducted: by telephone; postal questionnaire; or face-to-face interviews. In our interviews with Scottish evaluation practitioners, there appears to be consensus that face-to-face interviews offer the best opportunity to establish additionality. One interviewee commented that with postal questionnaires the accuracy is poor with respondents reluctant to write down that a project is non-additional. Direct interviews allow the interviewer to explain and probe, which increases the probability of an accurate response. Clearly, direct interviews are more costly than a postal questionnaire and only limited use can be made if the resources allocated to EU evaluation studies are not to be increased. As to the sampling fraction, we note the comments of one interviewee from Italy who argues that it is best to assess the impact of aid by focusing on a limited number of assisted enterprises, then undertake an in-depth analysis of their budgets, their employment situation etc before and after the grant. For him, this is better than “an omni-comprehensive and therefore superficial overview”. We might add that there may also be resource cost advantages in adopting such a focus.

The Ernst & Young review suggests that best-practice examples of surveys to identify both additionality and displacement effects can be found in France, Germany and UK. In France, the surveys conducted in Aquitaine and

Bourgogne are highlighted. Here, deadweight effects were estimated by asking beneficiaries of schemes supported by the Structural Funds whether or not they would have proceeded with their plans without assistance. This is very similar to *the approach adopted by SQW in their interim assessment of the Highlands and Islands Objective 1 programme, which in our view is itself an example of best practice in this area.*

The Thüringen 1994-99 Objective 1 interim evaluation can be considered as an example of best practice in the use of surveys to identify net effects. The survey findings suggested that in four percent of assisted firms there was full employment additionality, while in 53 percent of firms the additionality was partial. The evaluators (IFO) assumed that firms that supplied more than 50 percent of the finance involved would have gone ahead with their projects anyway, producing a 20 percent estimate for additionality. In the UK, the evaluation of the West Midland's Objective 2 programme involved a survey of final beneficiaries, some 200 businesses and individuals. The estimate of additionality here ranged from 35 to 70 percent of gross jobs. Another best practice example of the survey approach using a combination of different types of interview along with the evaluator's own judgement is given in Case Study 15.

CASE STUDY 15: USE OF INTERVIEWS AND EVALUATORS JUDGEMENT TO ESTIMATE ADDITIONALITY AND DISPLACEMENT - GIBRALTAR

In the final evaluation of the 1994-96 Objective 2 programme and the interim assessment of the 1997-99 programme for Gibraltar, face-to-face and telephone interviews were conducted by ECOTEC using a sample of projects to measure additionality among other impacts. The researchers' questionnaires had a set of four boxes for different degrees of additionality: 'exhibiting low additionality <25 percent', '25-49 percent', '50-74 percent' and 'very high additionality 75 percent+'. The researchers were requested to probe interviewees about the impacts of their projects and the extent to which they could be attributed to SF assistance: the researchers were then to use their own judgements to decide which category best described the project. Similar judgements were employed in determining displacement effects but not in the actual interviews.

Critique

The evaluators based their views only on whether displacement was likely to occur because of the targeting of particular sectors in the Gibraltar programmes: no quantification or explicit methodology appears to have been used.

Relevance to Scotland

The accuracy of the measured additionality and displacement effects of Structural Fund assistance is important for Scotland. Perhaps as a result of improvements in monitoring aspects of Structural Fund evaluation, there is a perception by both evaluators and Programme managers that the identification of the impacts on the economy of the Programme areas is relatively weak.

Of course, the objectives of programmes embrace more than job or output targets. A more problematic area for evaluation is to assess the effect of measures devoted to the stimulation of technological development and innovation and the associated additionality. Here the outcomes are likely to be more qualitative and so are difficult if not impossible to quantify. An interesting example of the treatment of the effects of RTDI-related actions is to be found in Austria, which is discussed in Case Study 16.

CASE STUDY 16: MEASURING NET EFFECTS OF TECHNOLOGY AND INNOVATION SUPPORT IN THE EVALUATION OF RTDI RELATED ACTIONS - AUSTRIA

As part of a study by Joanneum Research to evaluate the approach of Austria's Objective 2 regions towards RTDI issues, the evaluation undertook an 'impact survey' of a sample of 25-30 RTDI projects in the case-study regions. A questionnaire was drawn up which was designed to elicit a 'self-evaluation' of net RTDI impacts by the project leaders. The survey asked about the effects of the projects – including changes in productivity, sales revenue, new market developments and employment – by using a 0 to 4 ranking system (with 0 as 'not applicable' and 4 as 'important increase'). Where objectives were not met, project leaders were asked to rate a range of different explanatory factors, including skills deficiencies, time lag in receiving funds, intellectual property rights problems and contractual issues. More importantly, project leaders were asked whether the impacts were restricted to the company alone, other companies within the sector, or other companies outside the region. Lastly, the project leaders were invited to assess whether their projects would have gone forward without the Structural Fund support.

Critique

This approach offers an example of how useful information can be obtained about the impacts of SF interventions that are difficult to measure. Measurement is shifted from a cardinal to an ordinal scale using a ranking system. The weaknesses of the approach lie in the sole reliance on the views of project leaders and the risk of moral hazard, with project leaders having an interest in overstating impact. Ideally, an independent check should be provided, say through desk analysis of company performance before and after assistance, or through the selection of a control group.

Relevance to Scotland

This example is relevant for attempting to measure those variables which are hard to quantify.

c. Overall net impact

The establishment of the overall net impact of Structural Fund measures using the bottom-up approach requires the further identification of the displacement and positive secondary effects on the socio-economic domain (see first part of this section) of the programme area. Our discussion of the Scottish evaluation studies in Section 3 revealed that this was probably the weakest part of the

evaluation. The interim evaluations, following the Scottish guidelines, did not consider such effects, while the final evaluations established the secondary effects by the application of ‘off the shelf’ discount factors for displacement and multipliers for the positive secondary effect.

It is important to stress that there might well be circumstances where the use of well-chosen benchmark values gives both a cost-effective and accurate measure of these secondary effects. This would especially be the case for multiplier values. However, in practice, little or no explanation was offered as to the source of such benchmark values or their particular applicability to the programme area. Moreover, as the first part of this section noted, the approach adopted both in Scotland and elsewhere in Europe involves an implicit Keynesian ‘export base’ type approach where only demand and quantity adjustments matter. Accordingly, no account is taken of sectoral differences or the possibility of displacement occurring through the labour market via wage adjustments, or of other supply-side effects occurring through the effect of, for example, price changes on competitiveness.

However, the conclusion from our desk research and interviews is that *the identification of overall net impact is generally less well done elsewhere in Europe than it is in Scotland and the rest of UK.*

The Ernst and Young review reveals that, of the 75 official evaluations considered, only 21 (28 percent) sought to identify displacement effects, while 28 (37 percent) allowed for positive indirect effects. UK studies tended to take secondary effects more seriously, with 11 from 26 studies (42 percent) seeking to allow for both displacement and positive secondary impacts. Of the larger countries with several studies, no Spanish studies allowed for product market displacement – from 14 considered – and only four (29 percent) sought to allow for positive secondary demand effects. In France, three out of 15 (20 percent) adjusted for displacement and six (40 percent) allowed for positive secondary effects, while in Germany, where 12 studies were reviewed, the percentage was 17 percent in both cases (two studies out of 12). It is evident from these results that the treatment of displacement is more of a problem than positive secondary effects. We judge that this is because the Keynesian heritage of secondary multiplier analysis has produced many studies with multiplier values, while the treatment of displacement at a theoretical level is much less well understood and there are many less empirical studies.

While in principle there may still be examples of best practice even though there are only a small number of studies elsewhere in Europe, we consider that *there is little for Scotland to learn from official evaluations of secondary effects outside the United Kingdom.*

An example of one of the better studies undertaken elsewhere in Europe – the Netherlands – is discussed in Case Study 17.

CASE STUDY 17: IDENTIFICATION OF NET EMPLOYMENT EFFECTS FOR A WHOLE PROGRAMME - NETHERLANDS

This work was part of a larger report on the *ex post* evaluation of industrial core regions in northern EU regions for the 1989-93 period undertaken by Research voor Beleid. The evaluation for this Dutch region made use of a methodology developed by Ernst & Young for the assessment of the net employment effects of a programme.

A sample of projects under different priorities was selected and surveyed about different employment effects. For additionality, respondents were asked if the projects were ‘wholly’ additional, ‘partly’ additional – where additionality refers to Structural Funds induced changes in the scale and timing of projects – or ‘displaced’. Jobs resulting from these projects were then calculated using shares (1 for wholly additional, 0 for non-additional and an intermediate fraction for partly additional). At the same time, multiplier effects were calculated, distinguishing between ‘direct’ effects (supplier and income related) and ‘indirect’ effects (the contribution of projects to the creation of ‘an attractive business climate’).

Critique

The evaluation could only directly quantify secondary multiplier effects and did establish the degree to which an attractive business climate was created. It is also not clear whether a standard multiplier was employed to estimate the net employment figures, or whether the effects were estimated on a project-by-project basis within the sample. Nonetheless, the net effects were aggregated by applying net employment/project ratios for each priority’s sampling fraction for the priority, and then grossing the figures up for the programme’s different priorities. For the UK, the use of the PACEC discount and multiplier factors within studies such as the Yorkshire and Humber Region evaluation – considered in Case Study 4 – provide the best examples of the treatment of overall net impacts. This is because the benchmarks were derived from a review of 250 evaluation studies, which allowed different additionality and displacement benchmarks to be derived for different types of measures and projects. In applying the benchmarks, consideration was also given to differences in project aims, project populations and the stage of the business cycle.

Relevance to Scotland

This is an example of good practice using the approach most often used in Scottish evaluations although as can be seen from the critique even this evaluation was not ideal.

For the UK, the use of the PACEC discount and multiplier factors within studies such as the Yorkshire and Humberside Region evaluation – considered in Case Study 4 – provide the best examples of the treatment of overall net impacts. This is because the benchmarks were derived from a review of 250 evaluation studies, which allowed different additionality and displacement benchmarks to be derived for different types of measures and projects. In

applying the benchmarks, consideration was also given to differences in project aims, project populations and the stage of the business cycle.

6.2.2 *Top-down approaches*

Scottish Structural Fund evaluations have almost completely eschewed the application of top-down approaches. The main perceived drawback of the application of a top-down approach is the belief that such an approach can only effectively be used where the programme area is coincident with a Member State or larger region. This reason is cited in the Ernst & Young report.

In the first part of this section, we briefly considered the issue of spatial scale and evaluation. On the basis of that discussion, we would not agree that geographic scale is the principal determining factor of the appropriateness of a top-down approach. Clearly, Member States and large regions are more likely to have a full range of data covering the principal aggregates of the macro-economy. In addition, larger geographical areas may benefit from the prior existence of appropriate models that may be suitable for the evaluation of the Structural Funds. However, the more sophisticated top-down approaches, at least, involve a tested view of how the economy works, including its principal market interactions and spillovers. Applying such models to Structural Fund evaluation, whether it be *ex ante* or *ex post* assessments, necessitates that some thought be given to how specific measures might affect the aggregate economy and offers the prospect of controlling for other non-Structural Funds influences on the programme area.

Finally, from our survey of EU experience it is clear that top-down and explicit modelling approaches are somewhat more important for Structural Fund evaluation in the rest of the EU than in both Scotland and the UK. While there are some examples of the use of similar non-assisted areas or regions to form a counterfactual, we believe that the most fruitful best practice can be found in those studies that adopt an explicit modelling approach to the estimation of the impact of the Structural Funds. Four such examples are discussed in Case Studies 18-21.

We can find no examples of explicit modelling approaches that deal with the problems of modelling a fragmented partnership within the context of a devolved region. But the problem of focusing on sub-regions within a larger region is essentially one of reduced data availability and the possible absence of extant models. Moreover, there is a case for conducting complementary evaluations for the spatial unit of account over which the devolved government has jurisdiction. Hence, top-down evaluations at the Scottish level would be able to check whether estimated effects using either bottom-up or top-down approaches at the programme-area level have been at the expense of other non-assisted areas within Scotland. Another point of relevance here is that the more aggregative modelling approaches are not necessarily more expensive in resource terms than properly-conducted bottom-up approaches. Clearly, the construction of a full economic model for a programme area could be expensive but perhaps not as expensive as might at first be thought. There are several examples of multiplier and I-O models that have been constructed for sub-regions in Scotland using non-survey based techniques to adjust

existing Scottish models. In addition, several sub-regions that form part of current EU programme areas have constructed I-O tables, using survey techniques. Examples include Orkney, Shetlands, and the Western Isles. The Scottish Office sponsored work on tourism multipliers also would appear to be relevant here.

The following case studies have been chosen because of their quality but also because of their relevance to Scotland. For example, the use of input-output analysis in Case Study 18, is highly relevant to Scotland because Scotland is a leader in the development and application of I-O at both the Scottish and sub-Scottish levels. Such an approach offers the potential for a sectoral disaggregation of effects, something that is presently lacking in Scottish evaluation studies. The identification of Structural Fund impacts through direct econometric testing is illustrated in Case Study 19 which describes the PARADISE model. Case Studies 20 and 21 illustrate the use of formal macro-economic modelling techniques for *ex post* and *ex ante* evaluations. Case Study 20 is of the Irish HERMIN model, Case Study 21 is for the Mezzogiorno area of Italy.

CASE STUDY 18: DYNAMIC INPUT-OUTPUT ANALYSIS - BEUTEL

Jorg Beutel developed an evaluation system for DG XVI. This system uses dynamic Input-Output (I-O) analysis to evaluate the impact of Structural Funds on the economies of Greece, Spain, Ireland and Portugal. There are two elements to this system. First, Beutel develops a method for updating Input-Output tables. Second, he has used the relevant national I-O tables to simulate the impacts of Structural Fund expenditure in both *ex ante* and *ex post* evaluation exercises.

Updating I-O tables

One problem in using I-O analysis is that the construction of I-O tables is costly and time consuming. This means that I-O tables are typically available only for particular years and are published only with a time lag. Therefore, one can be working with a very dated table. A mechanism for using available information to update existing tables is clearly very valuable and improves the credibility of the analysis. However, in the Scottish context this is less important in that, at the moment, the Scottish Executive is producing annual I-O tables and these are relatively up to date.

Simulating the impact of the Structural Funds

Beutel uses I-O analysis to quantify the impact of the Structural Funds on the following economic variables:

- the rate of growth;
- other economic aggregates and the industrial structure;
- employment; and
- capital stock.

The I-O system is used to identify three separate effects. These are:

- the demand impacts generated by the construction of the infrastructure financed by the Structural Funds;
- the wages and salaries supported by the Structural Fund; and
- the additional induced investment generated as a result of the expansion in local activity.

Infrastructure

The expansion in infrastructure is the most straightforward to deal with. Here, this expenditure is simply treated as an expansion in final investment demand, and its impact calculated in the normal way via the Leontief inverse.

Wages and salaries supported by the Social Fund

The impact of the Social Fund expenditure seems to be treated as allowing an expansion in resources to particular resource-constrained sectors. This will lead to subsequent expansion in indirect and induced outputs and incomes. Whether this treatment is strictly compatible with the way in which the I-O model is used to analyse the other elements of Structural Fund expenditure is questionable.

Induced investment

In this case, the model is dynamised through linking net investment to changes in output via an accelerator mechanism. This means that sectoral capital stocks adjust, through investment, to changes in sectoral output, giving an added boost to the economy in terms of further increase in final demand. The predicted change in local activity now evolves over time as the economy more gradually adjusts to the injection delivered by the Structural Funds.

Once the changes in gross output have been determined, the subsequent impact on total employment and capital stock can be calculated through the use of the capital and employment matrices. Employment is a key policy variable and changes in capital stock might be taken to generate supply side improvements in the local economy. Structural changes in the economy can also be tracked, as can variables such as the balance of payments.

Strengths and weaknesses of the dynamic I-O approach

Strengths

- It is based upon a formal modelling approach with a clear analytical framework.
- It stresses sectoral disaggregation.
- It incorporates the interaction between economic sectors and elements of final demand.
- It uses existing I-O tables and is relatively low cost.

Within a Scottish context, an I-O based technique has the advantage that the economy is relatively well-served in terms of available I-O tables and certainly the existing information embodied in those tables could be more effectively utilised for evaluation purposes.

Weaknesses

- I-O is not a supply-side model. It is a demand driven model that has a passive (permissive) supply side. This has two key implications.
- First, the supply-side stimuli that the Structural Funds aim to produce have to be converted to a demand side stimuli to be captured by I-O. This has to be done in an essentially *ad hoc* way.
- Second, there are no supply constraints built into the model as it normally operates.
- I-O tables are not available for existing programme areas so that evaluation at that level is more problematic.

Whilst there are strengths to this type of modelling it is an unsatisfactory way, if used in isolation, to model the impact of Structural Fund expenditure. It is unable to capture the effect of improved competitiveness and efficiency that are central aims of Structural Fund intervention. However it is the case that at present the bottom-up Scottish (and UK) approach is really a hybrid. ‘Industrial survey’ methods are used to derive the direct outputs, and an export-base multiplier is employed to calculate the full impact. In this respect, at the very least, Scottish evaluation could make more use of the national and sub-national I-O tables that now exist in Scotland. This would imply a sectoral disaggregation of outputs which at present generally is not produced in evaluation studies. This in itself would be a beneficial change.

Relevance to Scotland

Scotland has available detailed I-O tables. However, the strong demand-driven nature of the standard I-O framework means that any solely I-O based approach is far from ideal for the evaluation of supply-side policies such as those pursued in the Structural Funds.

CASE STUDY 19: ECONOMETRIC TECHNIQUE - THE PARADISE MODEL

Eddy Blaas and Peter Nijkamp have developed an econometric technique for evaluating the impact of the European Regional Development Fund. The technique has the acronym PARADISE, Policy Assessment of Regional Achievements and Developments Induced by Stimuli of ERDF. The aim is to generate an evaluation method that can be applied in a consistent manner across regions in all the EU countries. This would give the Commission an indication of the effectiveness of ERDF expenditure as a whole, and also reveal the countries and types of region in which this expenditure had been particularly effective.

Over the time period covered by Blaas and Nijkamp, infrastructure takes the major share of ERDF expenditure, and they take the major economic variable affected by this expenditure to be private investment. They are, therefore, attempting to identify and quantify statistically the link between the levels of ERDF expenditure and private investment in the same region. They are mindful that employment is the variable that has the strongest policy relevance and, therefore, also attempt to measure the relationship between private investment and employment.

The method involves a two-step approach. The first step is an ‘exploratory frequency analysis’. Imagine that the relevant impact variable is A and the policy (or control) variable is B and that we have observations on these variables across all regions. These variables are standardised for regional size, so that they are expressed, for example, per head of population. The average values of A and B across the whole population under consideration are A^* and B^* . Each region is then classified as having an above or below average value for A and B. This places the region in one of the four categories identified in the table below. If there is expected to be a positive relationship between the policy and impact variable, we expect the bulk of the observations across regions to lie in the quadrants I and IV.

Cross-classification frequency table:

	B>B*	B<B*
A>A*	I	II
A<A*	III	IV

The frequency analysis is taken to be a flexible tool that is used in an exploratory way to identify possible relationships. Where a time series of cross-sectional (regional) data is available this analysis can be done year by year, with and without lags or for moving averages. Given the usual time lags that are thought to occur between improved infrastructure and increased private activity plus the lumpy nature of these investments, these options are valuable. Also a number of variables can be tried out. Further, if the number of regions permits, statistical testing (for example χ^2 analysis) can be undertaken at this level of analysis.

The model that Blaas and Nijkamp test to determine private investment, I_p , has additional variables alongside the ERDF expenditure, I_c . These are the expenditure on other public investment, I_{o-c} , and the change in gross value added, ΔGVA . The model therefore takes the form:

$$I_p = I_p(I_c, I_{o-c}, \Delta GVA)$$

The first step in the PARADISE procedure is to undertake extensive forms of this pairwise frequency analysis between the dependent and independent variables, taken one at a time. These are interpreted as revealing a positive relationship between the level of private investment in a region and all the independent variables.

The second stage of the PARADISE method is to perform a full-blown multivariate econometric analysis. Within this framework a lot of experimentation was done regarding appropriate time lags, etc. For an experiment with Dutch data, using moving averages the coefficient on the ERDF expenditure term was positive and significant, though in many of the other regressions it was not.

Strengths and weaknesses of the PARADISE model

Strengths

- It is econometric analysis, which is, in practice, the most rigorous form of testing with economic data.
- It develops an evaluation method that can potentially be used across all the EU regions with commonly-available data.

The statistical testing of the relationship between an independent policy variable and a dependent impact variable is in principle a powerful tool and this is a general methodological approach most strongly favoured in economics. The advantages that the regression results can be compared across European regions is perhaps less relevant in domestic evaluations.

Weaknesses

- It is only an *ex post* technique. Further, in so far as the technique depends upon time series data, results are generated some time after programmes are completed.
- The evaluation method does not distinguish explicitly between individual elements of the programme. Potentially if ERDF expenditure is differentially effective between regions and countries, inferences might be drawn linking this with the differential nature of the programmes in the different spatial areas, but this is very indirect.
- This approach does not focus on individual partnerships.
- The rigour of the existing econometric analysis is not clear. For example, no diagnostic statistics are reported. Here the model specification is restricted by the data availability for the least data-rich country.

This approach is appropriate for more long-run strategic decision taking at the EU level. It does not fulfil any of the short and medium run monitoring functions of present evaluations. It is, therefore, properly seen as complementary to many of the alternative bottom-up and top-down approaches.

Relevance for Scotland

The idea of attempting to identify the aggregate effects of Structural Fund policy through direct econometric methods is appealing though a more sophisticated model that this should be attempted.

CASE STUDY 20: THE HERMIN MODEL

Hermin is a development of the Hermes model. The Hermes model was designed by the Commission in the late 1970s primarily to identify and quantify the impacts of the oil price shocks on the individual economies of the EC. The Hermin models were subsequently built for the peripheral economies of Greece, Ireland, Portugal and Spain. Their main focus was to capture the impacts of the CSF and specifically the effects of policy aimed at enhancing infrastructure, human capital formation and other elements of industrial policy. Here we concentrate on the Irish Hermin model: the corresponding models for the three other peripheral economies are similar but differ in some details.

In the Hermin model there are four sectors:

- the traded sector (manufacturing);
- the non-traded sector (market services, utilities and construction);
- the public sector (public administration, education and health); and
- the agricultural sector (agriculture, forestry and fishing).

The traded sector

In the Irish Hermin, the output of the manufacturing (= traded) sector is driven by external demand. Specifically, the Irish traded sector is a pure price taker on world markets. Output is therefore dependent on the exogenous world price and the endogenous cost competitiveness.

The non-traded sector

The output of the non-traded sector (= sheltered or market-services sector) is determined by final demand. In this case, there is a wider range of sources of final demand, and the price is set as a mark-up on unit costs.

The public sector

Output in the public sector is treated as exogenous (ie. driven by government policy). However, the treatment of borrowing and debt accumulation is endogenous with the option of enforcing a PSBR borrowing constraint.

The agricultural sector

Here again output is primarily policy-determined through the operation of the CAP.

In all sectors, a KLEM (capital, labour, energy and materials) constant-returns production function is used. Factor demands are a function of sectoral output and relative factor prices. Also, in each sector capital is fixed in each period but is updated between periods through net investment.

In the Hermin model, the traded sector plays a crucial role in determining the performance of the whole economy. The small country assumptions are imposed so that exchange rates and interest rates are fixed exogenously. In the traded sector, the wage rate is determined by a bargaining procedure. The wage is therefore a function of commodity prices, direct and indirect tax wedges, the unemployment rate and productivity in the traded sector.

This wage is then transferred to the non-traded sectors (that is to say, the non-traded sectors experience the same percentage changes in the wage rate as does the traded sector).

Over the longer run, the unemployment rate, and thereby the manufacturing wage rate, will be affected by migration, which is assumed to take a Harris-Todaro form. This simply means that net migration flows are driven by variations in Irish – UK real wage and unemployment rate differentials.

Externalities

The Structural Funds focus on three types of expenditure which are thought to have externalities which will improve the overall performance of the recipient economy and encourage regional convergence. These are:

- factor productivity externalities;
- industrial composition externalities; and
- labour market externalities.

Factor productivity externalities

These are externalities experienced by the private sector which take the form of increased productivity. They are linked to two types of programme:

- public infrastructure; and
- education and training.

Evidence suggests that efficiency in the private sector is positively related to the stock of public infrastructure. Similarly, private sector productivity is positively related to the stock of human capital. Therefore in the Hermin model, the efficiency parameters on the private-sector production functions are taken to be increasing in the stock of public infrastructure and the number of trainees. Any increase in the efficiency of production has a positive impact on the level of economic activity through an increase in competitiveness and a higher real wage (encouraging immigration). However, an increase in labour productivity can reduce the demand for labour if the stimulus to activity is insufficiently low.

Industrial composition externalities

This stands for the stimulus to the level of foreign direct investment that is thought to be linked to expenditure on the Structural Funds. In the Hermin model, the proportion of the world's output of traded goods (here manufactures) produced in Ireland depends upon Irish cost-competitiveness. However, it is also argued that the attractiveness of Ireland for foreign-owned plants will additionally depend on the level of infrastructure capital and the skill level of the local labour force. Therefore, domestic output in the traded sector is again made dependent upon the stock of public expenditure and the number of trainees.

Labour market externalities

Programmes under the European Social Fund focus on the long-term unemployed and school leavers. These groups have only a small impact on the operation of the labour market and unemployment amongst these sections of the labour market has little impact on wage determination. If expenditure by the Structural Funds make some of these workers more active participants in the labour market, this will have an effect on the wage bargaining mechanism, making the wage more sensitive to variations in the unemployment rate. Again this is incorporated in the Irish Hermin model.

Strengths and weaknesses of the Hermin approach**Strengths**

- It generates an explicit model of the recipient economy. This model is open to scrutiny and based on prior theoretical reasoning so that the causal mechanisms are transparent.
- It captures economy-wide impacts.
- It incorporates supply-side effects, such as competitiveness and labour market effects.
- Once the model has been set up, additional simulation is inexpensive, particularly as against extensive face-to-face interviews.

These are very significant strengths. Any evaluation process requires some theoretical framework. Often these are both implicit and simplistic. The formal approach of the Hermin model makes the underlying model explicit (and therefore subject to challenge and debate) and also tailors the model to address the key evaluation issues raised by the Structural Funds. The model will allow sensitivity analysis to reveal the extent to which results change with a change in the values of key parameters.

Weaknesses

- Formal economic modelling involves a high set-up cost.
- The data requirements are high.

It would seem unlikely that one would set up a formal economic model of the Hermin type simply to carry out evaluations of Structural Funds. Hermin is a national model and therefore can be used in many other circumstances so that the fixed costs involved in setting up the model is spread over a wider range of work. Also if such a model is to be primarily econometrically parameterised, there probably is insufficient data at the regional level and certainly there would be difficulty applying this approach to individual partnerships. However, a similar model does now exist for Scotland. This is the Fraser of Allander Institute's AMOS Computable General Equilibrium model, a version of which has been used to evaluate Scottish Enterprise policy at the Scottish and UK level.

Relevance to Scotland

A model with an explicit, active supply side, such as Hermin would be very desirable for assessing the Scottish-wide impacts of supply-side policies, such as the Structural Funds. Whilst there is less data at the Scottish level, than in Ireland, for parameterising the model, a Computable General Equilibrium model, the AMOS model, is available for Scotland. This model has very similar characteristics to Hermin.

CASE STUDY 21: THE CSF EX ANTE EVALUATION FOR THE MEZZOGIORNO REGION OF ITALY

The Objective 1 CSF aims to encourage economic growth through its impact on the productivity, and therefore volume, of investment. The intention of the programme is to generate ‘breaks’ (rottore) in the supply- and demand-side behaviour of key economic agents. A traditional supply-side model, in which the productivity of private capital is expressed as a function of the stock of public capital, has been deemed inappropriate. This is because the parameters of this function would reflect what has been characterised as inefficient past public intervention in the Mezzogiorno. Indeed, the future improved quality of public investments has been interpreted as a direct additional effect on the productivity of the system. The macro-economic model therefore has been built around a total factor productivity function, which has as arguments the ‘break variables’ that the programme targets. Four variables are endogenous to the model: export, investments, net imports and the activity rate. The other variables are exogenous and simulate the positive externalities that the programme will have on the economy of the Mezzogiorno. Through these variables, the positive externalities on the economy of the Mezzogiorno are simulated and measured under alternative hypothesis.

To some extent the model is similar to the Hermin model, but is differentiated along the following lines:

- It does not distinguish between different types of public investments
- The explicit generation of prices is not considered
- It has no sectoral disaggregation
- It estimates the elasticity of some of the determinants of the function of the total productivity of factors.
- It is an explicitly regional model and deals separately with trade to and from the rest of Italy and the rest of the World.

The approach adopted by the Italian government has allowed it to verify the coherence of the objectives of the programme in relation to three main elements. These are:

- the compatibility with the overall macro-economic framework,
- the existence of, and transmission mechanisms through which, the hypothesised externalities act and,
- lastly, the size and dynamics of these externality effects on the productivity and, consequently on the growth, of the economic and social system of the Mezzogiorno.

Strengths and weaknesses

These are best measured against the Hermin model.

Strengths

- This is a regional model, which takes into account regional interaction
- Parameterised on regional data

Weaknesses

- Less sectoral disaggregation
- Less-well articulated supply side
- The model is particularly orientated toward the evaluation of infrastructure expenditure

Relevance to Scotland

This is relevant to Scotland in that it represents a regional, rather than national model. However, the lack of sectoral disaggregation and the rather narrow policy focus is a weakness. It is doubtful whether this would be the optimal choice of supply-side model for Scotland given the choice available.

6.3 Assessment

Scottish Structural Fund Evaluations primarily adopt the MEANS-recommended bottom-up approach. Gross effects are quantified almost wholly through industry survey techniques, with net impacts identified by calculating additionality, substitution, displacement and multiplier effects.

The bottom-up approach taken in Scottish Structural Fund evaluations makes it easier to trace the impacts of the Programme to individual Measures and even projects. Moreover it more directly tracks exactly how the elements of the programme work.

- The purely monitoring elements of evaluation have improved over time with greater consistency between objectives, targets and indicators and a heightened awareness of evaluation issues within PEs.
- Scottish evaluations, are amongst the best in the EU in providing additionality, displacement, substitution and multiplier estimates.

- However, there is legitimate concern, expressed by programme managers and consultants, over the accuracy of these estimates and therefore the ability of the present evaluations to give the true economic impact of the Structural Funds on the Programme areas. Concern applies to both practical and conceptual issues.

In the bottom-up approach followed in Scottish evaluations, identification of additionality and displacement strictly requires direct contact with recipients. But this is expensive and with limited budgets other short-cut methods have been adopted.

The accurate identification of additionality and displacement effects requires direct contact, preferably face-to-face, with recipients. This is expensive. Therefore in order that evaluations can be undertaken within budget, often Scottish evaluations use either benchmark values or the results of independent evaluations undertaken by a partner which is co-financing a particular project..

- The use of benchmarks to assess the efficiency with which an individual Programme is run (for example the efficiency with which projects are initially screened) is clearly questionable.
- The use of benchmark multiplier values requires care and discrimination. Multiplier values will vary across sectors and will also depend on the area size and types.
- The use of project evaluations, even where undertaken by independent consultants, raises moral hazard problems.

More resources are required if more accurate and Programme-specific measures of the effectiveness of spending on the Structural Funds is to be achieved.

Consideration should be given to the more intensive use of existing I-O tables for Scotland and sub-regions within Scotland for determining multiplier impacts.

There should be careful consideration of the role of Scottish/Highland Enterprise as a major partner in many programmes and whether their own evaluations of projects cannot be more formally and systematically integrated into the Structural Fund evaluation procedure.

The conceptual basis which links gross outputs and net impacts in the bottom-up approach is conceptually flawed. An implicitly demand-driven model is used to evaluate supply-side policies. This is not a criticism aimed specifically at Scottish evaluations. However, “cutting-edge” practice would link “industrial-survey” and a model with a more active supply side.

Essentially the identification of additionality, displacement and multiplier effects in the bottom-up approach favoured in Scotland and recommended by

MEANS implicitly uses a standard export-base model. This procedure is conventional, convenient and easy to grasp. Unfortunately it is clearly conceptually limiting where the focus is supply-side policy. Even for bottom-up approaches, a greater understanding of the wider impact of supply-side policy requires a background model that has an active supply-side. Such a model, the AMOS CGE model is available for Scotland and has been used for policy evaluation.

One area where Scottish Structural Fund evaluations appear relatively weak against other EU evaluations is in identifying disaggregated impacts, especially in the labour market

- If issues such as equal opportunities and sustainability gain more prominence, more disaggregated impacts are required.
- Disaggregating by sector will also improve the accuracy of displacement and multiplier estimates.

Top-down and bottom-up methods are not mutually exclusive. However, top-down methods are more practical at the economy-wide (Scottish level). The modelling of Structural Fund impacts at the national level might be more appropriately dealt with in a wider programme of economic modelling.

Top-down and bottom-up methods are in some respects complementary and at the least are alternative takes on the same problem. Given the difficulty in quantifying policy impacts there is room for both approaches. However, it is likely that top-down methods will be more appropriate at the Scottish level due to data availability and the ability to spread the fixed costs of model calibration and specification across a number of economic evaluation exercises, not limited to the Structural Funds. The availability of Input-Output (I-O) tables for Scotland would appear to support an I-O based approach, such as the Beutel model. We do not recommend this route. This is because I-O is ill designed to deal with policy-induced supply-side shifts in the economy.

- The impact of aggregate Structural Funds in Scotland should be assessed using an economy-wide supply side model of the Hermin type.
- Such a framework offers maximum flexibility and can be used for a wide range of policy purposes.
- Other direct attempts to identify policy impacts econometrically should be encouraged.

7. CONCLUSIONS AND RECOMMENDATIONS

The following section presents the conclusions and recommendations of the report, covering the basic premises for evaluation of regional development policies (and the Structural Funds in particular), the organisation and management of evaluation studies and the methodologies that could be applied to the Scottish context.

These conclusions and recommendations arise from taking into consideration Scottish past experience and the practice in other EU countries. They also reflect the continuing discussion with the study Steering Group, thus representing the point of view of the writing team and not the position of the commissioning institution.⁵⁴ This final part of the report is structured around five main issues: approach, partnership, capacity, methods and use of results.

APPROACH

It is necessary to place Structural Fund co-financed activities in a wider context, integrating Structural Fund and domestic policies also with regards to evaluation.

In Scotland, within the Scottish Executive, specific resources have been devoted to managing and co-ordinating the evaluation of Structural Funds. This has not always been the case in other Member States, as is reflected by the fact that Scotland has a strong reputation in evaluation. However, to appraise the overall effectiveness of Structural Funds it is necessary to place the evaluation of Structural Fund programmes in a wider context.

Recommendation 1: A common framework for the integration of Structural Fund and domestic policies should be developed. The construction of such a framework could be facilitated by:

- the institution of a permanent Evaluation Unit (cf. Case Study 2 – Italy), encompassing and co-ordinating the evaluations of Structural Funds and other economic and social development policies; or
- the development of a ‘nested’ approach to monitoring and evaluation (cf. Case Study 3 – West Midlands), with the inputs from various socio-economic actors involved in the implementation of regional policies, guaranteed by the organisation of meetings of partnership-based working groups.

Furthermore, a more careful consideration should be given to the role of Scottish Enterprise as a major partner in many programmes and whether their own evaluations of projects could be formally and systematically integrated into the Structural Fund evaluation procedure. In both options suggested, moreover, other Scottish institutions and, UK government departments, specifically with reference to some national policies, should be involved too.

Addressee: The Scottish Executive.

⁵⁴ The conclusions and recommendations of the report represent the point of view of the EPRC/FAI research team and are not the official position of the Scottish Executive. The usual disclaimers apply.

The quality and utility of evaluation could be improved further by developing a more integrated and on-going approach to evaluation.

Scotland has systematically carried out all the *ex ante*, interim and *ex post* evaluations in compliance with the regulations. However, the evaluation of Structural Fund programmes has been undertaken as a series of separate exercises, albeit using common guidelines and terms of reference and, given the reliance on a same group of evaluators, a similar approach to the evaluation of some programmes. There is considerable scope for organising a more integrated programme of Structural Fund evaluation in Scotland.

Recommendation 2: Those commissioning and steering evaluations could consider the following as a means to further reinforce the integrated approach to monitoring and evaluation:

- establish a more explicit link between different phases of evaluation (*ex ante*, interim, *ex post*);
- use more combined programme and thematic evaluations in a diagnostic process, allowing programme-specific and minority issues to be considered in more detail (cf. Case Study 5 - North-Rhine Westphalia);
- utilise more explicit comparative evaluation research across programmes (cf. Paragraph 5.2 - Austria);
- use broader meso-level evaluation techniques to explore programme elements in greater detail (cf. Case Study 4 - Yorkshire and Humberside).

If an Evaluation Unit was be created, this could potentially provide guidance on such issues.

Addressees: The Scottish Executive, the PMEs, the evaluators.

Recommendation 3: Process issues could be addressed more roundly by involving evaluators on an on-going basis and establishing a continuing dialogue between evaluators and programme managers (cf. Case Study 6 – Tuscany).

Addressees: The PMEs and the evaluators.

PARTNERSHIP

A more explicit partnership approach to evaluation could be encouraged, explicitly recognising the priorities and objectives of different stakeholders.

Partners' involvement is one of the strengths of the Scottish approach to Structural Fund implementation. Partners are involved in the management of programmes at all stages, from the design of programming documents, through project selection, to the validation of the decisions adopted as they participate in the Management and Monitoring Committees. For evaluation in particular, there is evidence of progressively greater involvement, over successive programming periods, and also of a gradual (though partial)

recognition of the potential benefits of evaluation as a programme management tool and mechanism for feedback. However, in some cases the involvement of partners has been difficult and in general there is still much scope for improvement. The involvement of partners and other stakeholders therefore should be enhanced, allowing dialogue and exchange at all phases of evaluation exercises.

Recommendation 4: At the programme level, the PME should continue to use Evaluation Steering Groups which already involve partners. However, these could meet on an on-going basis, in parallel with the implementation of the programmes and not only in the narrow context of evaluation exercise timetables. Moreover, the provision of an activity plan for such Steering Groups, defined and agreed at the outset of their creation, would contribute to ensuring the effective involvement of all parties and the undertaking of all relevant activities.

Addressees: The PMEs.

Recommendation 5: At the multi-programme level, and over the longer term, a more permanent dialogue on evaluation issues could be developed, perhaps through a permanent sub-group of the proposed Scottish Structural Funds Forum or the Scottish Co-ordination Team. The institutionalisation of an inter-programmes evaluation working group would ensure that the person responsible for evaluation in each programme attended the meetings: this was not always the case in the past, when the wide ranging agendas of the meetings between PMEs could determine the participation of other programme managers, even when evaluation was on the agenda.

Addressees: The Scottish Executive and the PMEs.

CAPACITY

Evaluation would benefit from improving capacities and skills of PMEs and partners.

At the programme level, often little time can be devoted specifically to evaluation, as the often passive re-proposal of the terms of reference suggested by the SE proves. The PMEs are frequently overwhelmed by the amount of tasks that they have to fulfil because of Structural Fund regulations, and this contributes to evaluation being relegated to a secondary position, and limits the possibility of time and resources being devoted to the development of specific expertise and skills in this field. It is therefore important to develop evaluation skills and expertise among PME staff, so that the PMEs can be actively involved from the design stage of evaluations, and are able to focus consultants on the specific needs they want to target. Similarly, to allow an active participation in the evaluation process of the partnerships, it will be necessary to stimulate their interest and develop their competence too.

Recommendation 6: the Scottish Executive and programme managers should invest in developing capacities for Structural Fund evaluation by promoting awareness, interest and involvement in Structural Fund support. This could be achieved by developing a publicly accessible resource library on evaluation, and a point of information for practitioners and partners on evaluation, similar

to the Austrian Check-point Eva (see Case Study 10). The Scottish Executive should therefore investigate the possibility of setting up such a centre, that, ideally, should include:

- a comprehensive, organised and open library on evaluation, encompassing evaluation documents (at least the Scottish ones, but possibly also UK and other countries), developmental documents (such as MEANS materials, EC guidance, methodological papers), conference materials;
- a web-site with:
 - documents and literature available online
 - a list of evaluators and their profiles, also from outside Scotland - this would not be a 'closed' or obligatory register, but an easily accessible list: submission by the consultants to the office responsible for the maintenance of the evaluation point and its web-site would allow the list to be up-dated
 - the calls for tender regarding programmes and thematic evaluations, for Structural Fund and domestic policies
 - a 'what's new' page, with the announcement of forthcoming conferences, seminars and other similar initiatives, and highlighting of new documents/information available on the site
 - links with evaluation societies and other relevant institutions (ie. the inforegio web-site) and academics
 - an interactive window, where users could submit queries to the staff of the evaluation point
- permanent, though not extensive, *staff*.

Such an evaluation info-point would help policymakers to take stock and exploit the existing work on evaluation (ie. the MEANS materials), more effectively than it has been to date.

Addressee: The Scottish Executive.

Recommendation 7: The Scottish Executive and programme managers should invest in developing capacities for Structural Fund evaluation by promoting exchange of experience and lesson-drawing before and after evaluation exercises, similar to the 'days of exchange' organised by DATAR in France. These 'days of exchange', could be organised by the Scottish Executive, after each round of evaluation and after meta-evaluations have been conducted to assess the strengths, weaknesses and quality of the evaluations undertaken. Such meetings should be open to partners and stakeholders and publicised broadly. The organisation of small thematic working groups, followed by a final plenary session, would help in managing the high number of participants and make possible a constructive, useful debate.

Addressee: The Scottish Executive.

It would be beneficial to widen the evaluation community and develop evaluators' skills and expertise for specific, distinctive issues.

The evaluation community in Scotland is deemed as well-prepared and responsive, but also too restricted in terms of numbers and range of skills. Investing in the development of skills and expertise will be essential to raise the profile and effectiveness of evaluations. The enlargement of the community of evaluators and the creation of a common evaluation culture will ensure higher quality and greater understanding of the potential impact of evaluation and, in the longer term, improve their use. Scotland has some of the leading Structural Fund evaluators in the UK (and indeed the EU), but the evaluation community is small and it has sometimes proved difficult to generate interest and involvement outside a small group of private sector consultants. This is largely due to the specialist nature of Structural Fund evaluation, as well as the time constraints under which evaluations have to be carried out. At present, little is done to invest in developing the evaluation community: interaction between those commissioning studies and potential evaluators is limited, and there is no central resource support for Structural Fund evaluation. Even the Scottish Executive lacks a complete library of Scottish programme evaluation studies and relevant evaluation literature (eg. the output of the MEANS programme, evaluation books and articles).

Moreover, the development of specific skills for the evaluation of specific issues is desirable too, since the PMEs have sometimes experienced difficulties in finding competent evaluators from whom highly specialised thematic evaluations can be commissioned. This would make it easier to address specific issues, such as the horizontal themes or social inclusion, that are gaining more and more importance in the context of the Structural Funds.

Recommendation 8: It is important to ensure broader knowledge on the available evaluators, beside an informal system of referral between PMEs. The provision of an open list of evaluators in the suggested info-point's web-site would be a effective solution to this. The web-site would also allow a broad diffusion of tender calls, which would certainly facilitate the participation of newcomers to the bids. Further publicity to the calls should be made through a broader use of traditional mass media, such as newspapers, PME bulletins and so on.

Addressees: The Scottish Executive and the PMEs.

Recommendation 9: Consultants and evaluators should be encouraged to undertake Structural Fund evaluations by publicising in advance a plan of activity on the future evaluations that the PMEs and the Scottish Executive intend to undertake on a certain timescale (ie. on an annual basis), as the DETR already does.

Addressees: The Scottish Executive and the PMEs.

METHODS

Improving the quality and range of data and information available is essential to the reliability and profitability of evaluations.

A lot of work has been conducted in Scotland on the availability and reliability of data. Although the quality of data has been relatively good, often better than

in other Member States, further work should be conducted to improve the reliability, adequacy and range of information available.

The purely monitoring elements of evaluation have improved over time with greater consistency between objectives, targets and indicators and a heightened awareness of evaluation issues within the PMEs. Scottish evaluations are amongst the best in the EU in providing additionality, displacement, substitution and multiplier estimates. However, there is legitimate concern, expressed by programme managers and consultants, over the accuracy of these estimates and therefore the ability of the present evaluations to give the true economic impact of the Structural Funds on the Programme areas. Concern applies to both practical and conceptual issues.

In the bottom-up approach, followed in Scottish evaluations, furthermore, the identification of additionality and displacement strictly requires direct contact, preferably face-to-face, with recipients. But this is expensive and with limited budgets, other short-cut methods have been adopted. Therefore in order that evaluations can be undertaken within budget, often Scottish evaluations use either benchmark values or the results of independent evaluations undertaken by a partner which is co-financing a particular project. The use of benchmarks to assess the efficiency with which an individual programme is run (for example, the efficiency with which projects are initially screened) is clearly questionable. The use of benchmark multiplier values requires care and discrimination, because these will vary across sectors and will also depend on the area size and types. The use of project evaluations, furthermore, even where undertaken by independent consultants, raises moral hazard problems.

Recommendation 10: The Scottish Executive should take step to re-organise the monitoring framework and set of indicators in a coherent and integrated framework for Structural Funds and, ideally, also non-co-financed domestic policies. As regards the Structural Funds, this issue is particularly relevant for ESF, which in the new programming period will in the be full competence of the Scottish Executive and may require a gearing-up in terms of resources and expertise. The Finnish FIMOS monitoring system, for instance, represents a good example of an easily maintainable and accessible data source.

Addressee: The Scottish Executive.

Recommendation 11: The work undertaken by the PMEs to increase the monitoring skills of implementers could be consolidated by dedicating increased resources to monitoring visits and other developmental activities.

Addressees: The PMEs.

Recommendation 12: More resources are required if more accurate and programme-specific measures of the effectiveness of spending on the Structural Funds are to be achieved.

Addressee: The Scottish Executive and the PMEs.

Scottish Structural Fund evaluations primarily adopt the MEANS-recommended bottom-up approach. Gross effects are quantified almost wholly through industry-survey techniques, with net impacts identified by calculating additionality, substitution, displacement and multiplier effects. However, the

accuracy of the estimates and the ability of the adopted approach to give the true economic impact of Structural Funds are questionable.

Recommendation 13: In the bottom-up approach, an accurate estimate of additionality and displacement requires direct contact with the recipients of Structural Funds assistance.

Addressee: The Scottish Executive.

The procedure conventionally adopted to link gross outputs and net impacts is convenient and easy to grasp. However, these procedures at present fail to capture the full complexity of the relevant economic relationships.

Recommendation 14: Where multiplier values are used to identify indirect effects, great care should be taken in the choice of multiplier. This is because multipliers vary across sectors, areas and the size and type of assisted activity.

Addressee: The PMEs

Recommendation 15: Consideration should be given to the more intensive use of existing Input-Output tables for Scotland and sub-regions within Scotland for determining multiplier impacts. The above-mentioned Evaluation Unit could have a leading role in promoting this.

Addressee: The Scottish Executive.

A more fundamental problem is that the bottom-up approach is conceptually flawed. An implicitly demand-driven model is used to evaluate supply-side policies. This is not a criticism aimed specifically at Scottish evaluations. A better understanding of the wider impact of supply-side policy requires greater disaggregation and a background model that has an active supply-side.

Recommendation 16: More attention should be paid to the estimation of disaggregated impacts, embracing industrial sectors, labour market impacts such as gender and environmental impacts that have implications for sustainability.

Addressee: PMEs

Recommendation 17: Industrial surveys typically associated with the bottom-up approach should be linked to an evaluation model with a more active supply side of the Hermin type. Such a model is available for Scotland – the AMOS computable general equilibrium model - and this model has been used for policy evaluation.

Addressee: The Scottish Executive and the PMEs

Recommendation 18: Top-down and bottom-up approaches should be treated as complementary and not competitive. Top-down and bottom-up methods are not mutually exclusive. However, top-down methods are more practical at the economy-wide (Scottish) level. The modelling of Structural Fund impacts at the national level might be more appropriately dealt with in a wider programme of economic modelling.

Given the difficulty in quantifying policy impacts, there is room for both top-down and bottom-up methods. However, it is likely that top-down approach will be more appropriate at the Scottish level due to data availability and the ability to spread the fixed costs of model calibration and specification across a

number of economic evaluation exercises, not limited to the Structure Funds. The availability of Input-Output (I-O) tables for Scotland would appear to support an I-O based approach, such as the Beutel model. We do not recommend this route. This is because I-O is not designed to deal with policy-induced supply-side shifts in the economy. However, the detailed sector-specific data available from the I-O table greatly facilitates disaggregated supply-side modelling.

Addressee: The Scottish Executive.

Recommendation 19: There should be more formal and systematic integration of Scottish Enterprise and Highlands & Islands Enterprise evaluations into Structural Fund evaluations at the Programme Area level. It is often the case that Scottish Enterprise and Highlands & Islands Enterprise undertake separate evaluations of projects that they support which also receive aid through the Structure Funds. These evaluations should be incorporated more formally into the Structure Fund evaluations.

Addressee: The Scottish Executive and PMEs.

Recommendation 20: The encouragement of the use of statistical and econometric methods to identify policy impacts both at the Programme Area and Scottish levels. Approaches similar to, but ideally more sophisticated than, the PARADISE model. The proposed Evaluation Unit could have a leading role in stimulating the development of innovative econometric models.

Addressee: The Scottish Executive.

USE OF RESULTS

Evaluations in Scotland could be diffused more effectively.

Broader dissemination of evaluation documents and findings will contribute to promote interest and strengthen the credibility of evaluations. It would also enhance the possibility of building a more democratic consensus around programmes and focusing future interventions on the needs of beneficiaries and other stakeholders. Such dissemination would promote interest, awareness and understanding of Structural Fund evaluation methods, processes and results.

Recommendation 21: An Evaluation Report Series could be set up, similar to that of DG Regio or those of the governments of Ireland and Finland, containing all Structural Fund evaluation studies as well as relevant reports on methodological issues, should be established.

Addressee: The Scottish Executive.

Recommendation 22: At the programme level, each evaluation should be planned accurately, including at the design stage, a basic ‘communication plan’, to ensure appropriate dissemination to different target audiences.

Addressees: The PMEs.

There is further room to strengthen the validation and exploitation of evaluation findings.

No formal assessment to date has been undertaken of the quality and effectiveness of the evaluation studies commissioned. Evaluations were perceived and undertaken as one of the many compulsory requirements of the regulations and, once the tasks were completed and no formal objections were formulated by the European Commission, no further work was undertaken to exploit them fully. This attitude has improved significantly over time, however there is still much to do to maximise the utility of evaluations. In particular, the conduct of ‘meta’ or second-degree evaluations, would improve the quality of evaluation studies and contribute to enhancing the consensus and acceptance of the main findings.

Recommendation 23: The Scottish Executive could undertake or commission, after each round of evaluation, meta-evaluations, to identify the strengths and weaknesses of the evaluation reports, the degree to which they can be and/or have been exploited and new themes for future evaluation exercises.

Addressee: The Scottish Executive.

Recommendation 24: The outcome of the suggested meta-evaluations could be discussed by civil servants, programme managers, partners, evaluators and other interested parties to encourage learning and lesson-drawing among the wider community. The suggested open discussion fora, such as the ones in France, would represent a good opportunity for debate and exchange.

Addressee: The Scottish Executive.

ANNEX 1: RESEARCH SPECIFICATIONS

RESEARCH INTO METHODOLOGIES USED IN THE EVALUATION OF THE EFFECTIVENESS OF EUROPEAN STRUCTURAL FUNDS

1. INTRODUCTION

1.1 This specification sets out the requirements of the Scottish Executive Development Department for research into methodologies used in the evaluation of the effectiveness of the European Structural Funds, drawing on experience here and in other European countries.

1.2 The Scottish Executive is currently developing its framework for the evaluation of European Structural Fund spending in Scotland. We are in the process of identifying the core data required for evaluation in order that it can be collected timeously. This research project will contribute to the development of that framework, by drawing on the experience elsewhere. The purpose of the research is to identify the evaluation methods used in other countries or by other researchers, identify the strengths and weaknesses of each method, and identify examples of best practice. It may also identify possible improvements to our framework.

2. BACKGROUND

2.1 Currently arrangements for the level and distribution of future European Structural Funding are being finalised for Scotland. The programmes will run over the next 6 years and the funding will be allocated to 5 programmes. Four of these programmes are spatially limited (East, West and South of Scotland and Highlands & Islands) and one covers all of Scotland outside the Highlands and Islands (Objective 3). These programmes have been allocated around 1.5 billion Euro and spending will be focussed on promoting employment, social justice, equal opportunities, sustainable development, training and local economic development.

3 RESEARCH AIMS

3.1 The overall aim of this research study is contribute to the development of the Scottish Executive's framework for the evaluation of the effectiveness of structural funds, by drawing on experience elsewhere. Working towards this we have the following aims:

- identify the evaluation methods used by other countries, or by other researchers, to evaluate the effectiveness of structural fund expenditure
- identify the strengths and weaknesses of each method
- identify examples of best practice
- identify improvements to our evaluation framework, e.g. gaps in our core data

3.2 In achieving the above aims the research should identify the relevant lessons from other countries and therefore fulfil the following requirements.

- Include a description and analysis of how evaluation and monitoring feed back into the running and funding of projects and programmes in the short, medium and long term
- Consider both top down and bottom up approaches
- Consider how bottom up approaches can be linked to macro economic effects
- Distinguish between gross and net effects
- Consider whether it is possible to assess the effectiveness of structural funds in relation to macro economic variables
- Consider the use of baselines and counter factuals
- Identify if and how other countries identify what good value for money is
- Consider how impact is measured when the effect is not easily quantifiable
- Consider the spatial level at which evaluation approaches are appropriate
- Provide a clear justification for and summary of views offered, particularly for the identified examples of Best Practice

3.3 In order that the above can be considered in context the research will also

- Identify resources associated with the evaluation methodologies previously identified, including manpower and IT, both internally and externally.
- Relate the findings on evaluation methodologies to what should go into the developing Scottish framework for monitoring and evaluation
- Identify difficulties that might be associated with the evaluation of overall effectiveness of structural funds in the context of a devolved region

4. RESEARCH METHODS

4.1 The successful contractor will be required to fulfil the aims as set out in this specification. We would anticipate that it would include

- A desk based study
- Discussions with practitioners and researchers, where appropriate

4.2 The tender document will cover, if possible, the following issues:

- Which regions/countries will be studied
- Which sources the study will draw on
- The methodology, which will be used to collate and analyse the material
- The likely layout and thematic content of the interim and final report

5. COSTS

It is envisaged that the total cost of research will be between £35,000 and £40,000 (including fees, expenses and VAT). Payment of 25% of the contract value will be made on satisfactory submission of the interim report and payment of 50% will be made on satisfactory submission of the draft final report with the balance payable on completion of the study.

6. RESEARCH OUTPUT

6.1 The main output of the study will be a final report, which meets the aims and requirements set out in this specification. The project outputs are as follows

- An interim report, to be submitted to the project manager by 1 June 2000
- The draft final report, to be submitted to the project manager on or before 30 June 2000
- A presentation of the draft final report to the Steering Committee, on 5 July 2000
- A final report of publishable standard to be submitted to the project manager by 14 July 2000.

6.2 Eight copies of each of the above reports will be provided on the dates given above. It would also be helpful, but not essential to provide an electronic version of these reports compatible with Scottish Executive systems (Microsoft Windows NT 4.0 and Microsoft Word, Excel and PowerPoint 97 SR-2).

7. TIMESCALE

7.1 The study is expected to be completed within a 3-month period commencing on 4 April 2000.

7.2 A detailed research schedule will be agreed with the successful tenderer and will form part of the contract. For guidance, however, the desired timetable will be:

4 April 2000	Award of contract.
4-12 April 2000	Consultation and research design stage
12 April 2000	Inception meeting (Steering Group Meeting)
12 April-1 June 2000	Gathering of material and interview stage
1 June 2000	Interim report
To be agreed	Progress meeting (Steering Group Meeting)
1-30 June 2000	Analysis and report writing stage
30 June 2000	Submission of draft final report
5 July 2000	Presentation of results (Steering Group Meeting)
5 July-14 July 2000	Redrafting as necessary
14 July 2000	Submission of final report

8. PROJECT MANAGEMENT

8.1 A Steering Group covering the interests from within the Scottish Executive will be set up for this project. The contractor will be expected to provide an agreed team to attend 3 steering group meetings as indicated above.

8.2 The project manager will be updated on the progress of the study on a fortnightly basis, but must be informed immediately of any problems, which may jeopardise the timing of the project. Fabian Zuleeg from SE-DD-EAS will act as project manager for this research project.

8.3 The contract will be governed by the Scottish Executive's Standard Conditions for Consultancy Contracts, a copy of which can be obtained from the project manager.

9. OPEN GOVERNMENT POLICY

9.1 All information supplied by tenders will be treated as Restricted-Commercial except that:

(i) references may be sought from banks, existing or past clients, or other referees submitted by the tenderers, and

(ii) disclosure may be made of such information relating to the outcome of the procurement process as may be required to be published in the Official Journal of the EC or elsewhere in accordance with EC directives or UK government policy on the disclosure of information regarding government contracts.

10. OWNERSHIP AND PUBLICATION OF THE REPORT

10.1 The ownership of the research material including the final report, the presentation material associated with the final report and any data produced as a result of the research lies with the Scottish Ministers.

11. ACCESS

11.1 Access to all necessary agencies will be negotiated or delegated to the tenderer by representatives of the steering group. Tenderers should not approach any organisations until the contract is let.

12. SUBMITTING A TENDER

12.1 Eight copies of the tender document should be submitted to Mike Pow at the address given in point 12.2 by 4 p.m. on Monday 27 March 2000 in accordance with the stipulations given in the covering letter. The tenders should include detailed information on how the consultants propose to carry out the research in order to meet the aims and objectives given in these research specifications. A detailed timetable should be included. The breakdown of costs (including daily rates and time allocated for each member of staff involved) should also be given.

12.2 Responses should be based on the information given in the Specification and upon the professional knowledge and expertise of the tenderer. Clarification of specific points, preferably by e-mail, can, however, be sought from:

Fabian Zuleeg

OR Mike Pow

Fabian.Zuleeg@scotland.gov.uk

Mike.Pow@scotland.gov.uk

Scottish Executive
DD-EAS-4b
Victoria Quay, Area 3-G
Edinburgh
EH6 6QQ
Tel. 0131/2440887
Fax 0131/2440888

Scottish Executive
DD-ESF-Team 4
Victoria Quay, Area 2-G
Edinburgh
EH6 6QQ
Tel. 0131/2440732
Fax 0131/2440738

12.3 A shortlist of tenderers, to be notified on 31 March, will be invited to give a short presentation of their proposal before a decision on letting the contract is made. This presentation will take place on Monday 3 April 2000.

12.4 Tenders should be submitted in a sealed envelope marked 'Tender Documents: ESF Evaluation Methodologies' to Mike Pow. Tenders should arrive by **4 p.m., 27 March 2000.**

All tenders will be acknowledged.

Fabian Zuleeg
**The Scottish Executive
Development Department
March 2000**

ANNEX 2: EVALUATION DOCUMENTATION ANALYSED FOR THE SCOTTISH REVIEW

Past programmes:

Programme	Evaluation document
Eastern Scotland CSF 1989-93	<i>Ex post</i>
Dumfries & Galloway CSF 1989-93	<i>Ex post</i>
Borders Obj. 5b 1994-99	Interim
Rural Stirling and Upland Tayside Obj. 5b 1994-99	Interim
North and West Grampian Obj. 5b 1994-99	Interim
Dumfries and Galloway Obj. 5b 1994-99	Interim
Highlands and Islands Obj. 1 1994-99	Interim
Eastern Scotland Obj. 2 1994-96	<i>Ex post</i>
Eastern Scotland Obj. 2 1997-99	Interim
Western Scotland Obj. 2 1994-96	<i>Ex post</i>
Western Scotland Obj. 2 1997-99	Interim
Objective 3 1994-96 (overview summary)	Interim

New programmes:

Programme	Evaluation document
Highlands and Islands 2000-06 phase out	<i>Ex ante</i>
South of Scotland Objective 2 2000-06	<i>Ex ante</i>
Objective 3 for Scotland	<i>Ex ante</i>

Thematic evaluations:

Evaluator	Evaluation document
SQW	Thematic evaluation of Business development
Eglinton	PME review
Colin Roxburgh	D&G European Awareness Project

Among other documents examined are the following:

- Framework for the evaluation of area-based European Structural Funds intervention, Scottish Office, European Structural Funds Division, May 1996
- List of ESF Indicators
- List of ERDF Indicators
- ESF evaluation strategy and work programme (2000-03)

- DfEE Pathways to employment. The final evaluation of ESF Objective 3 in Britain (1994-99)
- DfEE Appraisal and evaluation methods for assessing net value added, final report by Ernst & Young, August 1995
- Appraisal and evaluation methods for assessing net value added, workshop by Ernst & Young 23 November 1995
- DfEE Guide to appraisal (November 1999) on cost-benefit analysis (by Jacqui Hansbro)
- Guidelines for *ex post* evaluation of measures under regulation (EC) 951/97
- Documents from the DETR (framework for the evaluation of area-based European Funds, Terms of reference for the evaluation of O2 in the North-West Region, terms of reference for the *ex ante* evaluation of North East of England O2)
- 1994-99 selection criteria for monitoring visits used by D&G
- Western Scotland O2 1997-99 review of baselines, indicators and targets
- Western Scotland O2 1997-99 Revised indicator study, final report by EKOS, Jan 1999
- Evaluation of rural development programmes 2000-06 - Guidelines
- Method statement by EKOS and Fraser Associates for the *ex ante* evaluations of UK SF Programmes 2000-06
- Report of the Steering Committee of the review of the Programme Management Executives, 6 March 2000
- List of the steps of the *ex ante* evaluation 2000-06 in D&G.

ANNEX 3: LIST OF INTERVIEWEES – SCOTTISH REVIEW

<i>Organisation</i>	<i>Interviewees</i>	<i>Period covered</i>	<i>Programme</i>	<i>Date</i>
Scottish Executive	Elisabeth Williamson Mike Pow Fabian Zuleeg	Overview of all programmes across the various programming periods		15.05.2000
DfEE	Jacqui Hansbro	Overview of the O3 and O4 programmes up to date		15.05.2000
Northern Ireland Department of Finance	Tony Simpson	Overview on the Northern Irish approach to evaluation up to date		15.05.2000
South of Scotland Partnership	Caroline Lintsay	1994-99 2000-06	D&G Obj. 5b SS Obj. 2	12.05.2000
Highlands and Island Partnership	Dennis Malone Iain Campbell	1992-93 1994-99 2000-06	H&I OP H&I Obj 1 H&I phase out	17.05.2000
ESEP	Gordon McLaren	1994-96 1997-99 2000-06	Eastern Scotland Obj.2 Programmes	02.06.2000
SEP	Laurie Russell	1994-96 1997-99 2000-06	Western Scotland Obj. 2 Programmes	19.05.2000
Objective 3 Partnership	Jennifer Ezzi Stephen Herbert	1994-99 2000-06	Objective 3 programmes	16.05.2000
SQW	Alan Brazewell	Methodological issues		13.06.2000
EKOS	John McCready	Methodological issues		21.06.2000

The Steering Group Meetings and workshops on 22 May provided an opportunity to consider feed-back. Further contributions have been collected through telephone and e-mail contacts.

ANNEX 4: LIST OF INTERVIEWEES – INTERNATIONAL REVIEW

Country	Interviewee	Institution or Company	Function
Austria	Markus Gruber	InTeReg - Joanneum Research	Senior Evaluator
	Eduard Kunze	ORÖK	Civil servant
Belgium	Luc Vandendorpe	Ministry of the Walloon region	Civil servant
Denmark	Ulla Christensen	North Jutland County Council	Deputy Head of Department
	Henrik Halkier	Aalborg University - European Research Unit	Associate Professor
Finland	Pentti Malinen	University of Oulu	Researcher
	Susanna Piepponen	Ministry of Labour	Civil Servant
	Pasi Rantahalvari	Ministry of Interior	Civil servant
France	Céline Issard	CNASEA	Facilitator
	Pierre Metge	ACT Consultants, Paris	Evaluator
Germany	Herbert Jakoby	Nord Rhein Westphalen Government	Civil servant
	Ulrike Kugler	Government of Bremen	Civil servant
	Kathleen Toepel	DIW	Evaluator
Greece	No interviews		
Ireland	David Hegarty	CSF Evaluation Unit	Senior evaluator
	John Bradley	ESRI	Director
Italy	Stefania Cenciarelli	NOVA, Rome	Senior Evaluator
	Manuela Crescini	RESCO, Ancona	Senior Evaluator
	Massimo Florio	University of Milan	Professor
	Giampiero Marchesi	Ministry of Treasury and Budget, Evaluation Unit	Civil servant
Luxembourg	No interviews		
Netherlands	Pieter van Run	Research voor Beleid	Expert
	Rosemarie Wijnands	Ministry of Economic Affairs	Civil servant
Portugal	Nuno Vitorino	ILHA	Evaluator
Spain	Maria Angeles Diez	Universidad del Pais Vasco	Evaluator
	Antoni Soy	Universidad de Barcelona	Evaluator
Sweden	Hallgier Aalbu	Nordregio	Director
	Goran Hallin	Swedish Institute for Regional Research	Director

UK	Chris Hubbard	DETR	Civil servant
	Steve Nead,	DTI	Civil servant
	John Bedingfield Richard Crofts Alistair Reekie	Government Office for the West Midlands	Civil servants
EC	Andrea Mairate	DG Regio	Evaluation expert

All the interviews were undertaken during May and June 2000. In addition, the research drew on additional interviews on the subject of monitoring and evaluation undertaken in the course of past EPRC research, notably under the IQ-Net project (1997) and European Regional Incentives project (1999). Among them, not repeating interviewees mentioned above, were the following:

Country	Interviewee	Institution or Company	Function
Austria	Elisabeth Holzinger	ÖIR	Evaluation expert
	Dr Kunze and Herr Schodl	ÖROK	Civil servants
	Mag. Dr. Schwarz and Mag. Kräftner	NÖ Landesregierung	Civil servants
Denmark	Henning Christensen	North Jutland County Council	Civil servant/ programme manager
	Ebbe Poulsen	National Agency of Industry and Commerce	Civil servant
Finland	Mr Veijo Kavonius	Ministry of Interior	Civil servant
France	Martine Lévy	DATAR	Civil servant
	Isabelle Sordel	SOFRES	Evaluator
	Olivier Heres	Préfecture de Rhône Alpes	Civil servant/ programme manager
	Véronique Bardin	Mission FSE, Ministry of Employment	Civil servant
	Marc Challéat	Préfecture d'Aquitaine	Civil servant/ programme manager
Germany	Klaus Gerstner	Saarland Ministry of Economics	Civil servant/ programme manager
	Martin Zwick	ISOPLAN	Evaluator

Sweden	Dan Hjalmarsson,	EuroFutures Consultancy	Evaluator
	Tomas Ekberg	Fyrstad Objective 2 Secretariat	Programme Manager
	Lars Gunner Rönqvist	Ångermanlandskusten CAB	Programme Manager
UK	Ian Gibson	Welsh Office	Civil Servant
	Andy Macguire	Glasgow City Council	Civil Servant
EC	Jose Palma Andres David Sweet Jack Engwegen Richard Harding Michael Kiro Singh Miek van der Wee	DG XVI	
Other	Eric Monnier	C3E, Lyon	Evaluation expert

ANNEX 5: GLOSSARY

Definitions of the core evaluation terminology used in this report are given in this Annex, in alphabetical order, in order to clarify the meaning attributed to terms which could be interpreted in various ways. The definitions are taken from the MEANS 'Glossary of 300 concepts and technical terms'.⁵⁵ MEANS have worked to develop a common language for Structural Fund evaluation, and the definitions they propose are used increasingly widely among international Structural Fund evaluation circles.

Additionality

The fact that Community support for economic and social development is not substituted for efforts by national governments; in other words, the fact that the beneficiary State's own financing remains, globally, at least equal to that which existed before the Structural Funds' contributions.

Verification of the implementation of this principle is carried out at the national level in the context of financial control and not of evaluation as such. This term must not be confused with additionality which applies to the evaluation of the net effects of an intervention.

The 1997-99 Western Scotland Single Programming Document, furthermore, on ERDF, specifies that additionality means that 'ERDF grants can be awarded only to finance eligible projects which could not be carried out by the applicant without the additional support. The ERDF requested, therefore, should be the minimum necessary for the project to proceed'.

Baseline

State of the economic, social or environmental context, at a given time (generally at the beginning of the intervention), and from which changes will be measured.

The basic situation is described by context indicators which describe the economy, socio-economic environment, concerned groups, etc.

Benchmarking

Qualitative and quantitative comparison of the performance of an intervention, with that which is reputed to be the best in the same domain of intervention or in a related domain.

Benchmarking is facilitated when, at the national or regional level, there are league tables of good and not so good practice.

Context indicator

Measurement of an economic, social or environmental variable concerning an entire region, sector or group in which public intervention takes place (e.g. per capita GDP, annual number of jobs created in the region).

Context indicators may describe a basic situation before an intervention and a desired situation after intervention. They are generally quantified on the basis

⁵⁵ MEANS, *Glossary of 300 concepts and technical terms*, CEC (1999), *Evaluating Socio-economic programmes*, volume 6.

of data from statistics offices or statistical teams. They apply to an entire territory or group, unlike programme indicators which apply only to addressees actually affected by an intervention.

Cost-benefit analysis

Evaluation tool for judging the advantages of the intervention from the point of view of all the groups concerned, and on the basis of a monetary value attributed to all the positive and negative consequences of the intervention (which must be estimated separately).

When it is neither relevant nor possible to use market prices to estimate a gain or loss, a fictive price can be set in various ways. The first consists of estimating the willingness of addressees to pay to obtain positive impacts of avoid negative impacts. The fictive price of goods or services can also be estimated by the loss of earnings in the absence of those goods or services (e.g. in cases of massive unemployment, the fictive price of a day's unskilled work is very low). Finally, the fictive price can be decided on directly by the administrative officials concerned or the steering group. Cost-benefit analysis is used mainly for the ex ante evaluation of large projects.

Cost-effectiveness analysis

Evaluation tool for making a judgement in terms of efficiency.

This tool consists of relating the net effects of the intervention (which must be determined separately) to the financial inputs needed to produce those effects. The judgement criterion might, for example, be the cost per unit of impact produced (e.g. cost per job created). This unit cost is then compared to that of other interventions chosen as benchmarks.

Counterfactual situation

A situation which would have occurred in the absence of a public intervention.

For example, a firm was assisted so that its employees could be retrained in new technologies. No redundancies were recorded in the following two years. It is estimated that without the assistance (counterfactual situation) 50 jobs would have been lost. By comparing the counterfactual and real situations, it is possible to determine the net effects of the public intervention. Various tools can be used for the construction of the counterfactual situation: shift-share analysis, comparison groups, simulation using macro-economic models, etc. At the baseline, the real situation and the counterfactual situation are identical. If the intervention is effective, they diverge while it is underway.

Deadweight

Change observed among direct addressees following the public intervention, or reported by direct addressees as a consequence of the public intervention, that would have occurred, even without the intervention.

For example: a farmer received assistance for the building of a self-catering cottage. In the survey he stated that the support had enabled him to create better quality facilities, but that he would have built the cottage, even without

support. Thus, there is a deadweight since the construction of the cottage cannot be imputed entirely to the intervention. Deadweight can account for as much as 50% of the gross effect. The estimation of deadweight necessitates a survey of direct addressees, preferably with a comparative analysis of non-participants.

Demand-side effect

Effect which spreads through growing intermediate consumption of enterprises (supplier effect) and through the income distributed within the assisted region, and which, in turn, generates spending by households (multiplier effect). When public financing ceases, demand side effects disappear.

Direct effect

Effects of a public intervention on its direct addressees, excluding all repercussions on other groups.

For example: investment support has direct effects on assisted businesses (production costs and capacity; creation or maintenance of jobs). New infrastructure has a direct effect on the people and enterprises which use it.

Displacement effect

Effect obtained in an eligible area at the expense of another area.

Displacement effects may be intended (e.g. displacement of a public administration from the capital to a region undergoing redeployment) or unintended (e.g. 10% of the jobs created by a regional development programme resulted in the disappearance of jobs in other eligible regions; a firm used programme assistance to move its premises from the centre to the outskirts of a town). When they are not intended, displacement effects must be subtracted from gross effects to obtain net effects.

Effect

Socio-economic change resulting directly or indirectly from an implemented intervention.

Effects include the results and impacts of an intervention, whether positive or negative, expected or not. In certain cases, the term “effect” is wrongly used to include outputs.

Effectiveness

The fact that expected effects have been obtained and that objectives have been achieved.

Effectiveness can be assessed by answering the following questions, for example: “Could more effects have been obtained by organising the implementation differently?” or “Which are the most successful operators or measures?”. An effectiveness indicator is calculated by relating an output, result or impact indicator to a quantified objective. For example: the objective in terms of number of firms created was as high as 85%; the placement rate

obtained by operator A is better than obtained by operator B. For the sake of clarity, it may be useful to specify whether one is referring to the effectiveness of outputs, results or impacts.

Efficiency

The fact that that effects were obtained at a reasonable cost.

Efficiency may be assessed by answering the following questions, for example: “Could more effects have been obtained with the same budget?” or “Have other interventions obtained the same effects at a lower ?”. An indicator of efficiency is calculated by dividing the budgetary inputs mobilised by the quantity of effects obtained. For example: the average cost of training a person who have been jobless for a long time is 2,000 euro; the intervention should achieve a cost per job created of less than 30,000 euro.

For the sake of clarity, it could be useful to specify whether the efficiency referred to relates to outputs, results or impacts. The efficiency of outputs is called the unit cost.

Externality

Effect of a private action or public intervention which is spread outside the market.

For example: a firm pollutes a river and causes an economic loss for a fish farm downstream; an engineer leaves the firm in which he or she was trained and applies his or her know-how in a new firm which he or she creates. By their very nature, externalities trigger private choices which cannot be optimised through the mechanisms of market competition. Only collective and often public decisions are able to promote positive external effects and prevent negative ones. A large proportion of financial support allocated within the framework of European cohesion policy is aimed at promoting positive external effects which businesses do not seek to create themselves spontaneously

Gross effect

Change observed following a public intervention, or an effect reported by the direct addressees.

A gross effect appears to be the consequence of an intervention but usually it cannot be entirely imputed to it. The following example shows that it is not sufficient for an evaluation merely to describe gross effects: Assisted firms claimed to have created 500 jobs owing to the support (gross effect). In reality, they would in any case have created 100 jobs even without the support (deadweight). Thus, only 400 jobs are really imputable to the intervention (net effect).

Impact

A consequence affecting direct addressees following the end of their participation in an intervention or after the completion of public facilities, or

else an indirect consequence affecting other addressees who may be winners or losers.

Certain impacts (specific impacts) can be observed among direct addressees after a few months or in the longer term (e.g. the monitoring of assisted firms after two years). In the field of development support, these impacts are usually referred to as sustainable results.

Some impacts appear indirectly (e.g. turnover generated for the suppliers of assisted firms). Others can be observed at the macro-economic or macro-social level (e.g. improvement of the image of the assisted region); these are global impacts. Evaluation is frequently used to examine one or more intermediate impacts, between specific and global impacts. Impacts may be positive or negative, expected or unexpected.

Indirect effect

Effect which spreads throughout the economy, society or environment, beyond the direct addressees of the public intervention.

Indirect “internal” effects, which are spread through market-based relations (e.g. effect on suppliers or on the employees of an assisted firm), are distinguished from external effects or “externalities” which are spread through non-market mechanisms (e.g. noise pollution; cross-fertilisation within an innovation network).

Input-output analysis

Tool which represents the interaction between sectors of a national or regional economy in the form of intermediate or final consumption.

Input-output analysis serves to estimate the repercussions of a direct effect in the form of first round and then secondary effects throughout the economy. The tool can be used when a table of inputs and outputs is available. This is usually the case at the national level but more rarely so at the regional level. The tool is capable of estimating demand-side effects but not supply-side effects.

Macro economic model

Tool used to stimulate the main mechanisms of a regional, national or international economic system.

A large number of models exist, based on widely diverse macro-economic theories. This type of tool is often used to simulate future trends, but it may also serve as an evaluation tool. In this case, it is used to simulate a counterfactual situation, and thus to quantitatively evaluate net effects on most of the macro-economic variables influenced by public actions, i.e.: growth, employment, investment, savings, etc. The models are generally capable of estimating demand-side effects far more easily than supply-side effects.

Multicriteria analysis

Tool used to compare several interventions in relation to several criteria.

Multicriteria analysis is used above all in the ex ante evaluation of major projects, for comparing variants. It can also be used in the exp post evaluation of an intervention, to compare the relative success of the different components of the intervention. Finally, it can be used to compare separate but similar interventions, for classification purposes. Multicriteria analysis may involve weighting, reflecting the relative importance attributed to each of the criteria. It may result in the formulation of a single judgement or synthetic classification, or in different classifications reflecting the stakeholders' different points of view. In the latter case, it is called multicriteria-multijudge analysis.

Net effect

Effect really imputable to the public intervention and to it alone, as opposed to apparent changes or gross effects.

To evaluate net effects, based on gross effects, it is necessary to subtract the changes which would have occurred in the absence of the public intervention, and which are therefore not imputable to it since they are produced by confounding factors (counterfactual situation). For example, the number of employees in assisted firms appears to be stable (change or gross effect equal to zero). However, it is estimated that without support there would have been 400 redundancies (counterfactual situation). Thus, 400 jobs were maintained (net effect).

Output

That which is financed and accomplished (or concretised) with the money allocated to an intervention.

Outputs may take the form of facilities or works (e.g. building of a road, rehabilitation of an urban wasteland; purification plant; tourist accommodation). They may also take the form of immaterial services (e.g. training, consultancy, information).

Programme indicator

Indicator which concerns the inputs and outputs of the intervention as well as the results and impacts on its direct and indirect addressees.

Programme indicators may include derived measuring efficiency, effectiveness or performance. They are quantified by monitoring information systems and also be evaluation when it produces primary data. When they concern effects, programme indicators measure only those which affect direct and indirect addressees.

Result

Advantage (or disadvantage) which direct addressees obtain at the end of their participation in a public intervention or as soon as a public facility has been completed.

Results can be observed when an operator completes an action and accounts for the way in which allocated funds were spent and managed. At this point

s/he may show, for example, that accessibility has been improved due to the construction of a road, or that the firms have received advice claim to be satisfied. The operators may regularly monitor results. They have to adapt the implementation of the intervention according to the results obtained.

Secondary effect

All the impacts produced by an intervention, over and above its first round effects, that is, apart from direct addressees and the first circle of indirect addressees.

Potentially, the mechanisms of secondary effects stretch in concentric circles throughout the entire economy. They are generated by supplier effects, income multiplier effects, impulsion effects, etc. The evaluation of secondary effects is based on macro-economic reasoning and techniques.

Shift-share analysis

Tool for evaluating regional policy, which estimates the counterfactual situation by projecting national economic trends onto the economy of a given region.

The basic assumption of this technique is that, in the absence of regional policy, the evolution of economic variables in the region would have been similar to that of the country as a whole. Comparison between the policy-off and policy-on situations is concluded with an estimation of the macro-economic impact of regional policy. The optimum conditions for using this tool rarely exist.

Substitution effect

Effect obtained in favour of a direct addressee but at the expense of a person or organisation that does not qualify for the intervention.

For example, a person unemployed for a long time found a job owing to the intervention. In reality, this job was obtained because someone else was granted early retirement. If the objective was the redistribution of jobs in favour of disadvantaged groups, the effect can be considered positive. An evaluation determines, with regard to the objectives of the intervention, whether the substitution effect can be considered beneficial or not. When it is not beneficial, the substitution effect must be subtracted from gross effects.

Supply-side effect

Secondary effect which spreads through the increased competitiveness of businesses and thus of their production.

Sustainability

The ability of effects to last in the middle or long term.

Effects are sustainable if they last after the funding granted by the intervention has ceased. They are not sustainable if an activity is unable to generate its own resources, or if it is accompanied by negative effects, particularly on the environment, and is that leads to blockages or rejection.

Weighting

Used to state that one criterion is of more or less importance than another one in the formulation of a global judgement on an intervention.

The weighting of criteria can be formalised by expressing it as a percentage (the total being 100%). Multicriteria analysis also uses weighting.