

University of Liverpool
Department of Civic Design

**URBAN POLICY IMPACT EVALUATION:
TOWARDS A SYSTEMATIC APPROACH**



Thesis submitted in accordance with the University of Liverpool
requirements for the Degree of Doctor in Philosophy

By

Ashraf H. Bakr
January 1996

**PAGE
NUMBERING
AS
ORIGINAL**

To my mother

Acknowledgements

I am greatly indebted to my supervisor, Professor Peter Batey, for his invaluable advice and continuous help and support. I was really very fortunate to work with him. Indeed, there are elements of the research that would not have been possible to accomplish without his support and close involvement. To give but two examples, the placement at the Government Office was basically his idea and personal initiative; the field trip to the Netherlands would have proven impossible if not for his support (indeed, I would have been in a dire financial situation if not for his immediate action).

I also owe a great deal to all those who, literally, went out of their way to help me at various stages of the research; unfortunately, many of whom cannot be named for reasons of confidentiality. In a strict chronological order: Mr John Dodgson and Dr Helga Drummond, University of Liverpool; the SRB Team and Working Group at the Government Office - in particular, the Head of the SRB Office who, to the very last stage of the research, was very keen on providing me with the needed information; Mr Peter Mackie, ITS, University of Leeds; Professor Phil Goodwin, TSU, University of Oxford; Professor Martin Wachs, UCLA; Professor Peter Hall, UCL; Mr Richard Evans, John Moores University, Liverpool; Professor Henk Voogd, University of Groningen, The Netherlands; Geke Kuiper, Johan Woltjer and Jos Arts, University of Groningen; Professor Pete Reitveld, Free University, Amsterdam; the Head and Team of the Planning Department, Municipality of Groningen; the senior planning officer, the Province of Groningen; the Head of Policy Analysis Division, the Dutch Ministry of Finance; Mr Maarten Mulder, the Dutch Ministry of Transport; the senior member of Liverpool City Challenge Team; the senior member of Victor Hausner and Associates and, Professor Naomi Carmon, Technion, Israel. To each and all of them, I am deeply grateful for their help and the valuable information they willingly shared with me. However, the opinions expressed in this research are strictly mine and are **not** intended to be critical of any person, team or organization.

Last, but not least, a special kind of thanks to Maria.

ABSTRACT

The main objective of this research is to develop a systematic approach for comprehensive ex-post evaluation of urban policy. In broad terms, urban policies are central government initiatives applied in a spatially targeted manner within urban areas where specific needs have been identified. 'Comprehensive' is defined as the attempt to answer *all* the questions policy-makers and stakeholders are likely to raise at the various stages of the planning and implementation process. In moving towards the achievement of this objective a number of different, yet supplementary, sources of knowledge and experience are examined.

Part 1 reviews the underlying assumptions and strengths and weaknesses of existing appraisal methods and examines their applicability in ex-post evaluation and the choice among them. It also examines the different views and models of both monitoring and implementation analysis and the role each can play within a comprehensive evaluation approach.

Part 2 examines evaluation methods adopted in a number of case studies in various fields. It starts with the first hand experience in urban policy evaluation within a governmental context. Then, it critically reviews the methodology adopted for evaluation in a number of case studies in the field of urban policy. Evaluation traditions in the fields of regional (economic) policy in Britain and, trunk road and motorway schemes in the UK, USA and the Netherlands are also reviewed.

These different strands are brought together in the form of a flexible systematic approach for comprehensive ex-post evaluation of urban policy. The choice of the 'components' of the approach is based, to a large extent, on the first part of the research. However, the organisation of the approach and the exact role each tool can play are greatly informed by the second part.

TABLE OF CONTENTS

Acknowledgements	i
Abstract	ii
List of Tables	vii
List of Figures	viii
CHAPTER 1: INTRODUCTION	
1.1 THE MAIN OBJECTIVE	1
1.2 THE METHODOLOGY	6
1.3 INFORMATION SOURCES	8
1.4 THE STRUCTURE OF THE RESEARCH	12
..... PART 1: FORMAL EVALUATION METHODS	
<i>INTRODUCTION TO PART 1</i>	17
CHAPTER 2: EVALUATION - AN OVERVIEW	
2.1 INTRODUCTION	19
2.2 DEFINITIONS, TYPES AND FUNCTION	22
2.3 THE POLITICAL AND ADMINISTRATIVE CONTEXT	27
2.4 UTILIZATION OF EVALUATION RESULTS	31
2.5 THE INTEGRATED EVALUATION APPROACH	33
2.6 CONCLUDING REMARKS	42
CHAPTER 3: CBA, CEA & VFM	
3.1 INTRODUCTION	47
3.2 COST-BENEFIT ANALYSIS	51
3.2.1 Definition	51
3.2.2 CBA & Public Decision-Making	52
3.2.3 Financial vs Social CBA	52
3.2.4 Elements of the Analysis	54
3.2.5 Public Decision-Making Criteria and Distributional Effects	64
3.3 COST-EFFECTIVENESS ANALYSIS	69
3.3.1 Definition	69
3.3.2 Decision-Making Criteria	71
3.4 VALUE FOR MONEY	72
3.4.1 Definition	72
3.4.2 The Practice	73
3.5 CBA, CEA, VFM AND EX-POST EVALUATION	77
CHAPTER 4: PBS, GAM & MCE	
4.1 INTRODUCTION	81
4.2 THE PLANNING BALANCE SHEET	84
4.2.1 The Methodology	85
4.2.2 The Pros and Cons of PBS	91
4.3 GOALS-ACHIEVEMENT MATRIX	95
4.3.1 The Methodology	96

4.3.2 The Pros and Cons of GAM	99
4.4 MULTI-CRITERIA EVALUATION METHODS	102
4.4.1 The Methodology	103
4.4.2 The Pros and Cons of MCE Methods	106
4.5 PBS, GAM, MCE AND EX-POST EVALUATION	108
CHAPTER 5: MONITORING	
5.1 INTRODUCTION	115
5.2 MONITORING: A BACKGROUND	118
5.2.1 Definition and Role	118
5.2.2 Why Monitor?	119
5.2.3 Monitoring and Review	120
5.2.4 Types of Monitoring	121
5.3 MONITORING: TWO DIFFERENT VIEWS	122
5.3.1 The Traditional View	122
5.3.2 The Extended View	123
5.3.3 Performance Indicators	126
5.4 MONITORING AND THE POLITICAL CONTEXT	127
5.5 INFORMATION: A REQUIREMENT AND A PROBLEM	130
5.5.1 The Diverse Information Base	130
5.5.2 Information Sources	131
5.5.3 Information Problems	133
5.5.4 Information Systems	135
5.6 CONCLUDING REMARKS	138
5.6.1 Monitoring and Strategic Planning	138
5.6.2 Monitoring and Comprehensive Evaluation	140
CHAPTER 6: IMPLEMENTATION ANALYSIS	
6.1 INTRODUCTION	145
6.2 DEFINITION AND SCOPE	147
6.3 IMPLEMENTATION: THEORY AND ANALYSIS	151
6.3.1 Top-Down Approaches	152
6.3.2 The Pros and Cons of Top-Down Models	159
6.3.3 Bottom-Up Approaches	162
6.3.4 The Pros and Cons of Bottom-Up Models	165
6.3.5 Synthesis of Approaches	167
6.4 THE CHOICE OF A MODEL	169
..... PART 2: CONTEXT-DERIVED METHODS	
<i>INTRODUCTION TO PART 2</i>	177
CHAPTER 7: THE SINGLE REGENERATION BUDGET	
7.1 INTRODUCTION	181
7.2 THE SINGLE REGENERATION BUDGET	183
7.2.1 The Initiative	183
7.2.2 Objectives	185
7.2.3 Partnerships	186
7.2.4 Assessment of Bids	186
7.2.5 Bidding and Decision-making Arrangements	188
7.2.6 Further Guidance	188

7.3 THE APPRAISAL PROCESS	192
7.3.1 Assessing Individual Bids	192
7.3.2 Choosing Among Bids	193
7.3.3 Budget Constraints	194
7.3.4 Influential Factors	195
7.4 THE AFTERMATH	203
7.4.1 Before the Final Decision	203
7.4.2 The Final Decision	205
7.4.3 After the Final Decision	206
7.4.4 The Second Bidding Round	208
7.5 CONCLUDING REMARKS	210
7.5.1 The Appraisal Process	210
7.5.2 The Evaluation of the SRB	212
CHAPTER 8: URBAN POLICY EVALUATION - 3 CASE STUDIES	
8.1 INTRODUCTION	221
8.2 RETROSPECTIVE EVALUATION OF A SINGLE POLICY -PROGRAMME FOR THE VALLEYS	227
8.2.1 The Initiative	227
8.2.2 Terms of reference	229
8.2.3 Methodology	229
8.2.4 Discussion	231
8.3 COMPREHENSIVE EVALUATION - ASSESSING THE IMPACT OF URBAN POLICY	238
8.3.1 Objectives and problems	238
8.3.2 Sample and Methodology	239
8.3.3 Outcome Measures	241
8.3.4 Discussion	242
8.4 INTERIM EVALUATION OF INNOVATIVE POLICY - CITY CHALLENGE	253
8.4.1 The Initiative	253
8.4.2 Monitoring and Evaluation Guidelines	255
8.4.3 National Evaluation of City Challenge	256
8.4.4 Local Evaluation of City Challenge	258
8.4.5 Discussion	261
8.5 CONCLUDING REMARKS	264
CHAPTER 9: ASSESSING THE IMPACT OF REGIONAL POLICY IN BRITAIN	
9.1 INTRODUCTION	273
9.2 REGIONAL POLICY IN BRITAIN	278
9.2.1 A Brief Background	278
9.2.2 The Choice of Criteria	280
9.2.3 Evaluation Methodologies	282
9.3 PARTIAL ECONOMETRIC METHODS	284
9.3.1 The Effects of Regional Policy	285
9.3.2 The Effects of Individual Policy Instruments	288
9.4 COMPREHENSIVE ECONOMETRIC EVALUATION	290
9.4.1 Cost-Benefit Analysis	290
9.4.2 Statistical Modelling	293
9.5 THE MACRO APPROACH (INDUSTRIAL SURVEY)	294
9.6 CONCLUDING REMARKS	295
9.6.1 Assessing the Effects of Regional Policy	295
9.6.2 Research Implications	310

CHAPTER 10: TRUNK ROADS & MOTORWAYS APPRAISAL AND EVALUATION

10.1 INTRODUCTION	319
10.2 APPRAISAL METHODOLOGY	323
10.2.1 Economic Appraisal - COBA 9	324
10.2.2 Environmental Appraisal	326
10.2.3 Towards a Comprehensive Appraisal Framework	331
10.3 EX-POST EVALUATIONS	336
10.3.1 Experience in the UK and USA	336
10.3.2 The Dutch Experience	347
10.4 DISCUSSION	355
10.4.1 Appraisal Methodology	355
10.4.2 Ex-post Evaluation	360
10.4.3 Research Implications	365

..... CONCLUSIONS

CHAPTER 11: TOWARDS A SYSTEMATIC APPROACH

11.1 THE MAIN OBJECTIVE	369
11.2 THE PROPOSED APPROACH	370
11.3 DIFFERENT MODES OF THE APPROACH	381
11.4 THE WAY FORWARD	384

BIBLIOGRAPHY 387

..... APPENDICES

Appendix A: A Monte Carlo Extension to CBA	415
Appendix B: Output Measures - The SRB	417
Appendix C: Outcome Measures - The SRB	420
Appendix D: SRB Guidance Notes	422
Appendix E: Core Impact Indicators - City Challenge	429
Appendix F: Baseline Data and Impact Indicators - City Challenge	430
Appendix G: A Chronology of British Regional Policy	432
Appendix H: An Abstract Example of the 'Framework' Approach	436
Appendix I: Illustrative Example of Environmental Impact Tables - Stage 3	440
Appendix J: The Letter Sent to 'Invited Commentators'	445
Appendix K: The Choice of A Demonstration Area	446

LIST OF TABLES

Table 2.1: Initial Components of the Integrated Evaluation Approach	37
Table 3.1: CBA in Private and Public Sectors	53
Table 3.2: Appropriate Decision Criteria	63
Table 5.1: Types of Monitoring	121
Table 5.2: Some Common Sources of Information	134
Table 6.1: Decision-Making and Implementation Modes	172
Table 7.1: Programmes Contributing to the SRB, 1994/95 (£m)	185
Table 8.1: The Socio-Economic Indicators	243
Table 9.1: Principal Social Benefits and Costs of Regional Policy	308
Table 10.1: Classification Summary of Economic Impacts	334
Table 10.2: Classification of the Sample by Type of Impact and Methodology	341
Table 10.3: Classification of Economic-Related Research by Type of Proxy	342
Table 10.4: Classification of the Sample by Geographical Coverage	343
Table D.1: An Illustration of Schedule A - Strategic Objectives	424
Table D.2: An Illustration of Schedule B - Funding Profile	425
Table D.3: An Illustration of Schedule C - Milestones	426
Table D.4: AN Illustration of Schedule D - Quantifiable Outputs	427
Table K.1: Criteria for Choosing Demonstration Area	447

LIST OF FIGURES

Fig 1.1: The Structure of the Research	14
Fig 2.1: The Planning Cycle	23
Fig 2.2 a & b: An Overview of Some Major Types of Evaluation	26
Fig 3.1: An Approach to Organizing for VFM in Local Authorities	76
Fig 4.1: An Illustration of The Standard Format of the Planning Balance Sheet	86
Fig 4.2: An Abstract Form of the PBS	87
Fig 4.3: An Illustration of the Summary Table of PBS	90
Fig 4.4: An Abstract Form of the Summation Table of PBS	91
Fig 4.5: Goals-Achievement Matrix	96
Fig 4.6: The Evaluation Matrix	103
Fig 4.7: The Priority Matrix	104
Fig 4.8: The Appraisal Matrix	104
Fig 5.1: The Context of Monitoring Activities	124
Fig 6.1: A Model of the Policy Implementation Process	153
Fig 6.2: Skeletal Flow Diagram of the Variables Involved in the Implementation Process	157
Fig 6.3: Direct and Indirect Impacts on Implementation	158
Fig 6.4: General Overview of Conceptual Framework of Policy Change	168
Fig. 7.1: An Illustration of the SRB Bid Checklist	190
Fig. 7.2: The Overall SRB Bid Assessment Sheet	191
Fig 9.1: The Response of Labour Demand and Supply to Job Creation Schemes	306
Fig 10.1: COBA Evaluation System	325
Fig 10.2: Evaluation Framework	333
Fig 10.3: Business Expansion Analytic Framework	336
Fig 10.4: A Sample of the EIM	350
Fig 10.5: Analysis Steps and Their Interrelationships	351
Fig 11.1: The Proposed Approach	371
Fig 11.2: Some Alternative Modes of the Proposed Approach	382
Fig A.1: Monte Carlo Extension of CBA	416
Fig K.1: Scorecard for Choosing Demonstration Area	447

CHAPTER 1: INTRODUCTION

1.1 THE MAIN OBJECTIVE

If it does not sufficiently inform the decision-makers and the public so that they can use the information provided in order to arrive at more rational decisions, evaluation is an academic exercise. For this purpose, evaluation will have to be more context responsive (Hill, 1985a: 31)

The main objective of this research is to develop a systematic approach for comprehensive ex-post evaluation of urban policy. In broad terms, urban policies are central government initiatives applied in a spatially targeted manner within urban areas where specific needs have been identified. 'Comprehensive' is defined as the attempt to answer *all* the questions policy-makers and stakeholders are likely to raise at the various stages of the planning and implementation process. The need for such a framework should be self-evident. A public policy (on the face of it) is enacted to achieve a particular set of objectives - to have a certain impact. It follows that policy-makers are interested, at least to an extent, in assessing the achievement of these objectives (regardless of their reasons). Targeted population(s) and interested groups share the same interest; each group, however, is motivated by different reasons and is likely to place varying emphasis on the same objective.

The complexity of urban problems, however, means that policy impacts will take some time to materialise. It also means that evaluation research aimed at assessing policy impacts is likely to require a long time to be

conducted. Policy-makers, on the other hand, have a much shorter time-frame. The built-in emphasis on regular monitoring in recent urban initiatives (eg the Single Regeneration Budget) indicates policy-makers' interest in regular feedback on policy outputs (and to a lesser degree, outcomes). This information is also a crucial input to the management and implementation process of the policy, ie at the lower-levels of the policy-making process.

A public policy is likely to affect different groups of the society in different ways. This will give rise to equity considerations: who gains and who loses. A public policy may require substantial public resources. The growing emphasis on 'value for money' and efficiency at different governmental levels clearly indicates a concern with the 'economics' of public policy. Such concern is also driven by the growing trend towards public sector accountability and the scarcity of public resources combined with people's increasing demands for better services.

In brief, a wide range of questions is likely to arise at the various stages of the planning and implementation process. These questions will relate to the output, impact, economy, efficiency, effectiveness, process and equity of the policy at hand. If evaluation is to inform policy-makers and stakeholders, it follows, *timely* answers will have to be provided to *each* and *all* of these questions. This is how comprehensive evaluation is defined in this research.

The idea for this research came in 1992 while the author was studying for the Degree of MA in Metropolitan Planning at the Department of Civic Design in the University of Liverpool. The project chosen for the dissertation was an "Evaluation of the Greater Cairo Master Scheme". The initial aim was to assess the achievement of a number of key objectives of the scheme after almost eight years of implementation and (substantial) political support. The literature review then indicated a dearth of research on systematic ex-post evaluation, comprehensive or otherwise. Indeed, that was one of the main problems encountered during the project. The approach had, unfortunately, to be confined to a comparison between what was planned and what was then achieved in regard to the Ring Road and the New Settlements Programme.

Despite the limited nature of that literature review, the hypothesis for the Ph.D. research was formulated as: "There is as yet no systematic approach for the comprehensive ex-post evaluation of urban policy". Therefore, and as stated above, the main objective of the research became to develop such a framework. It is well understood that circumstances differ from one case to the other. Any attempt to arrive at a 'universally applicable' framework is, to say the least, unrealistic. The framework, consequently, has to be a flexible one that is capable of being modified; rigidity is a drawback that must be avoided. The aim then is to develop a framework that is ready to 'build upon' not necessarily to apply in its present form.

The initial task was to validate that hypothesis. In order to do so, the first step of the research was a wide-ranging literature review of the state-of-the-art of evaluation research. It emerged that, within the field of town and regional planning, evaluation methodologies, eg the planning balance sheet, have been conceived and practised primarily in an ex-ante context. These methods focus on the (anticipated) outcome of the policy and aim to facilitate the choice between various alternatives. This attempt to estimate policy impacts raised a question about the applicability, and role, of such methods in ex-post evaluation. This is the first starting point of the research.

The literature review also revealed a different approach to comprehensive evaluation within the field of social policy. Comprehensive evaluation was based on the 'sequential' application of different analytical tools; the need for a subsequent tool is justified by the findings of the previous one(s) (eg Rossi *et al.*, 1979; Posavac and Carey, 1989 and Rossi and Freeman, 1993). For instance, it is argued that impact assessment should be carried out only when monitoring indicates that outputs have actually been delivered on the ground (Rossi and Freeman, 1993: 167). However, this approach ignores the fact that sometimes the mere intention of implementing a policy may result in (substantial) costs. Moreover, assessing the impact of policy is heavily dependent on the experimental and quasi-experimental approaches which are almost impossible to apply in the field of urban and regional policy. Nonetheless, this literature gave the first indication of the need for a comprehensive evaluation to draw on several traditions at once.

Perhaps the most important finding of the literature review was the dearth of research on comprehensive ex-post evaluation of urban policy, as defined above. To answer *all* questions, evaluation will clearly have to draw on several traditions at once. Such research is severely lacking in the literature. Being ex-ante oriented, existing evaluation methods ignore questions of output and process. At the same time, monitoring, economic evaluation and implementation analysis appear all to have been developed in isolation from each other. Monitoring is a vital source of regular feedback on policy outputs and performance. Economic evaluation, self-evidently, addresses questions of economic efficiency and effectiveness. Implementation analysis seeks to identify the factors and forces influencing the implementation process. Comprehensive evaluation, it is believed, will have to bring these traditions together; that is what is severely lacking in the literature.

The notable exception that can be traced is the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984). This approach aims to synthesise several evaluation traditions so as to enable a comprehensive, decision-oriented evaluation with a multi-group perspective. It has four components: monitoring, economic evaluation, implementation analysis and goals-achievement matrix. Although in this particular instance it was "particularly tailored to the evaluation of broad-aim social programs" (Alterman *et al.*, 1984: 381), there is no reason, in principle, why it cannot be applied to other types of programme, including urban ones. This approach became the second starting point of the research. It relates to the first starting point in the sense

that two of the approach's components were existing evaluation methods: cost-effectiveness analysis and goals-achievement matrix.

1.2 THE METHODOLOGY

The question however remained: How to achieve the main objective of the research? There was a number of alternative methodologies to choose from. One alternative was to develop a framework for comprehensive evaluation on the basis of the two starting points mentioned above. This framework would then be applied using available information generated in a reported case study; that is, a re-working of a previous study using the proposed framework to test its applicability.

This alternative, however, carried with it several inherent restrictions. First, it would be confined to formal evaluation methods and would, thus, ignore both practical experience and difficulties and contextual constraints within which evaluation takes place. The result might therefore be an unrealistic approach; one that would face severe difficulties or could not be applied altogether. Second, it would also be restricted by the particular characteristics of the chosen study. Given the fact that circumstances differ, sometimes dramatically, from one setting to the other, the approach might therefore be inapplicable in other cases. Third, and related, the choice of the case study would be a biased one; it would be based almost entirely on the availability of information. This alternative was, therefore, rejected.

A variant on this alternative was to develop the framework through both the two starting points and the evaluation of a case study. Although this alternative had the advantage of taking contextual influences into consideration, it would still be restricted by the characteristics of the case study. It would also ignore any lessons that could be drawn from practical experience in evaluation. Perhaps more importantly, it would have been almost impossible to conduct. It required resources (time, personnel and funding) far beyond the scope of this research.

Another alternative was to "extract" the framework from practical experience. Despite the noted dearth of research, ex-post evaluation has been, and is being, carried out in numerous case studies. This practice may have found a way of overcoming the lack of research and *perhaps* an unwritten 'code of practice' exists. One reason to reject this alternative was that a large sample of studies would be required in order to ensure a satisfactory degree of representativeness. This sample might prove to be un-manageable. A second reason to reject this alternative was that there was simply no guarantee that a consistent definition of comprehensive evaluation would exist. Indeed, there might be no such thing as comprehensive evaluation in practice. Moreover, this alternative would simply ignore any theoretical developments in the field.

It was therefore concluded that the methodology of the research will, itself, have to draw upon several approaches at once. Research developments

and practical experience have both to be taken equally into account. The methodology adopted, accordingly, is a hybrid of the above alternatives; an attempt to draw on their advantages and avoid their drawbacks. It may be described as a 'meta-evaluation'; a variant on 'meta-analysis', though no attempt is made to use statistical methods⁽¹⁾. This approach is an analysis and a 're-evaluation' of the findings of previous research; an attempt to pull these findings together to arrive at a cumulative result in the form of a comprehensive evaluation approach. This is not a traditional literature review. The research does not aim to arrive at a taxonomy of findings. Neither does it aim at a simplistic count of studies where particular methods have, or have not, been applied. Rather, and throughout the research, an explicit attempt is always made to relate the outcome of the analysis to the primary objective of the research.

1.3 INFORMATION SOURCES

In search for analytical tools, and on the basis of the afore-mentioned starting points, formal evaluation methods are examined to assess the role each may play within the comprehensive evaluation framework. To account for practical experience, a number of case studies in the field of urban policy are examined. In recognition of established tradition in other disciplines, evaluation experience in the fields of regional economic policy in Britain and trunk road and motorway schemes is also reviewed. The primary focus is on

⁽¹⁾ "Meta-analysis is the name given to a set of techniques for reviewing research in which the data from different studies are statistically combined." (Cook and Leviton, 1980: 449)

the questions asked, the methodology adopted and the problems encountered and how they may have been overcome, if at all.

However, the research is not entirely reliant on secondary sources of information. Successful attempts have been made to secure primary material. The research benefits considerably from a first hand experience in urban policy evaluation. A voluntary placement was arranged for 5 weeks (September - October, 1994) in one of the then newly re-structured Government Offices for the Regions. This period was chosen to coincide with GOs' appraisal of bids submitted for funding under the first bidding round of the Single Regeneration Budget (SRB). As part of a team, the author helped in assessing individual bids and was allowed to attend the Working Group's meeting in which the choice among submitted bids was made. Access was also given to (some) correspondence between the GO and the DoE and discussions were held with several members of the Working Group.

The contact with the GO has been maintained since then and through the subsequent stages of the first round: the preparation of monitoring and evaluation guidelines, the preparation of Delivery Plans and the first quarterly assessment of progress reports and payment of grant. This experience gave an invaluable insight into both the political and administrative context and constraints within which evaluation takes place and, the types of information policy-makers seek and the questions they ask. These, no doubt, are crucial considerations in any evaluation exercise.

In addition, a field trip to the Netherlands was undertaken for three weeks (May - June, 1995) to gain an insight into the theory and practice of policy evaluation. Given the major contribution of Dutch researchers to the field of multi-criteria evaluation, it was assumed that the practice of policy evaluation in the Netherlands has certain characteristics that are worth investigating (which proved largely to be true). Meetings were held with researchers in the field of urban and regional planning, including two meetings with Professor Piet Rietveld at Free University, Amsterdam. A presentation, about the research and its objectives and methodology, was given at the Faculty of Spatial Sciences in the University of Groningen, which stimulated a very useful, and long, discussion. Prior to the trip, contact was established with Professor Henk Voogd who provided some enlightening views about the contextual influences on evaluation. Professor Voogd, by introducing the author to colleagues in the Faculty of Spatial Sciences, contributed considerably to the success of the trip.

Meetings were held with senior planning officers at both the Municipality and the Province of Groningen where the discussion related mainly to their understanding and practice of policy evaluation and the problems they encounter. At the central level of the Dutch government, a meeting was held with the Head of the Policy Analysis Division at the Ministry of Finance in the Hague. Two meetings were also held with senior officers in the Ministry of Transport. The discussions focused on the role and practice of policy evaluation from the viewpoint of central government. The contact with one

of the senior officers in the Ministry of Transport has been maintained since the visit. The officer has subsequently visited the Department of Civic Design in September 1995 where an extended meeting was held with Professor Peter Batey, Dr Peter Brown and the author. Among other things, the discussion served to confirm the findings of the field trip.

Valuable material was also obtained during the field trip. One of the most important reports to be acquired was the demonstration evaluation of the Dutch national transport plan. Contact was then initiated with a member of the team who carried out this research to clarify a number of issues regarding the methodology adopted.

Furthermore, the research benefits from contacts initiated with a range of researchers and practitioners in various fields. During the literature review, both Mr John Dodgson and Dr Helga Drummond, from the University of Liverpool, were consulted about relevant literature in regard to cost-benefit analysis and decision-making models, respectively. Recently, contact has been established with Professor Naomi Carmon; one of the three authors of the 'Integrated Evaluation Approach'. Professor Carmon re-asserted the fact that the approach requires substantial resources to be conducted and that is very likely the reason why it has thus far been applied once.

Personal contacts play a central role in the second part of the research - the case studies. The review of the evaluation of the Programme for the

Valleys benefits from a meeting with a senior member of the evaluation team. In examining the national and local evaluation of City Challenge (CC), meetings were held with a member of the evaluation team and a senior member of the City Challenge team. Several key figures in the field of transport planning and evaluation were also invited to comment on the findings of examining evaluation experience in this field and provide their explanation of these findings.

A query was also placed at the Regional Science Association ServeList on the Internet in regard to the evaluation of trunk road and motorway schemes. Of the responses received that of Professor Martin Wachs, UCLA, was the most important. Professor Peter Hall, UCL, was also approached to provide a wider perspective of the issue, from the view point of town and regional planning.

1.4 THE STRUCTURE OF THE RESEARCH

The structure of the research is shown in Fig 1.1. Part 1 is devoted to the analysis of 'formal evaluation methods'; these are the result of research that has been conducted with the sole aim of developing evaluation methods and analytical tools that can be applied in (almost) any setting. Although these methods aim to inform the decision-makers, they have been developed in a 'context-free' way; they attempt to be free from any contextual constraints to their application. In contrast, 'context-derived methods', the subject of Part 2, are those methods that evaluators have developed within the context of the

case in question. These are the evaluators' attempt to carry out the task presented to them within the contextual constraints imposed upon them; the evaluators' approaches to real-world case studies.

The two parts are not in isolation of each other. The discussion of formal methods in Part 1 informs the review of case studies in Part 2. For instance, the discussion of the conceptual and practical difficulties of cost-benefit analysis (chapter 3) became of considerable value in reviewing the experience of trunk roads and motorways appraisal (chapter 10) which is based on a computerised form of CBA. Many of the problems with governmental guidelines for monitoring the SRB and CC (chapters 7 and 8) were identified on the basis of an understanding of the 'extended view' of monitoring; one of the aspects covered in chapter 5.

Part 1 begins with a summary of the first stage of the research - the literature review. This review gave rise to the two starting points mentioned above. The first point was the existing evaluation, or rather appraisal, methods. In examining these methods, the aim is, first, to assess their applicability in ex-post evaluation and, second, to choose among them, if proven applicable. These methods are examined in this particular order because of the evident 'line of development' among them. Cost-benefit analysis (CBA) is perhaps the first systematic method for project evaluation to be established, being derived from economic theory. Cost-effectiveness analysis (CEA) is an attempt to overcome some of CBA's limitations.

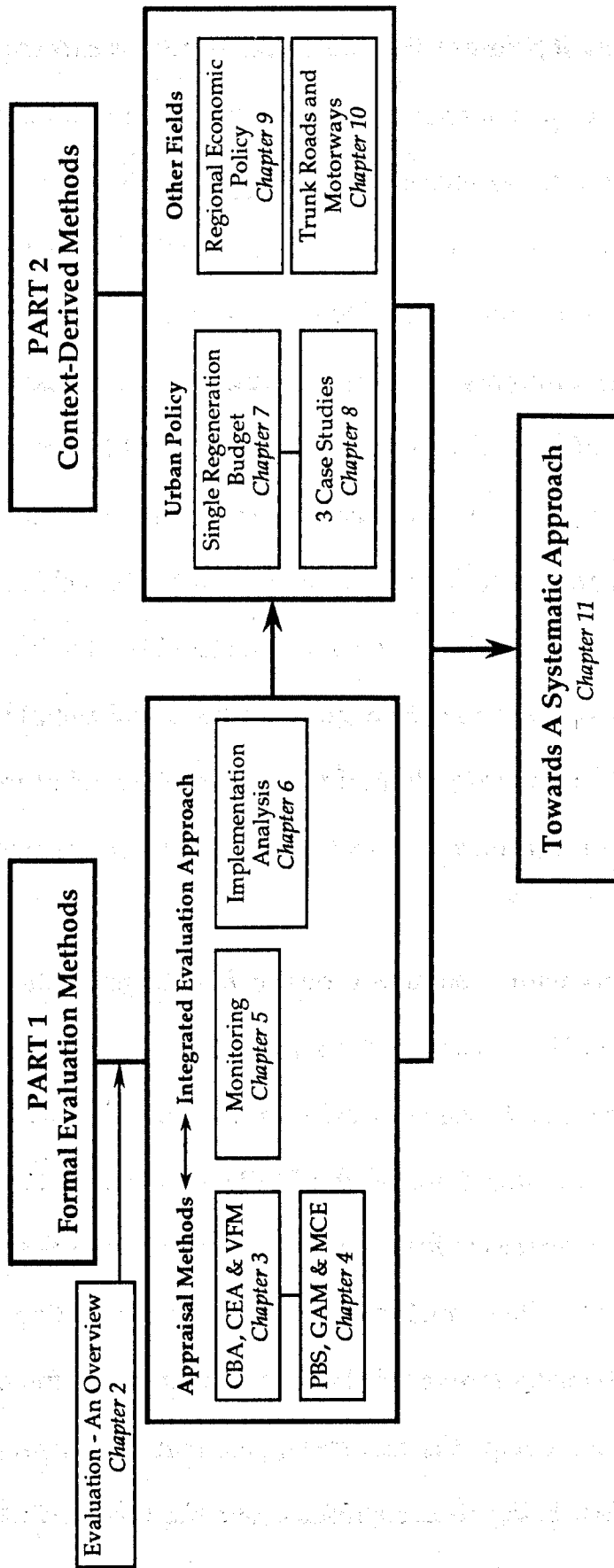


Fig 1.1: The Structure of the Research

Value for money (VFM) is the most recent notion in the field of economic evaluation. Apparently, it is an attempt to facilitate economic evaluation in public policy-making. The planning-balance sheet (PBS) aims both to broaden the scope of CBA and overcome its major limitations in the field of town and regional planning. Goals-achievement matrix (GAM) was introduced as a methodology to overcome the pitfalls of both CBA and PBS. GAM itself marked a new direction in evaluation research that came to be known as multi-criteria evaluation (MCE) methods.

CEA and GAM are two components of the 'Integrated Evaluation Approach'; the second starting point for the research. Part 1 then moves to examine the two remaining components of that approach: monitoring and implementation analysis. The underlying objective of chapters 5 and 6 is to determine the role both these tools can play within a comprehensive evaluation approach.

Part 2 critically examines evaluation methodologies adopted in a number of case studies in various fields. It starts with the first hand experience of the Single Regeneration Budget (SRB) which exhibits the progression from ex-ante to ex-post evaluation. The experience was also of great importance in exposing the political and administrative context and constraints within which ex-post evaluation will take place. In chapter 8, three of the most recent case studies in urban policy evaluation are reviewed. These are: the evaluation of the Programme for the Valleys (Victor Hausner & Associates, 1993);

'Assessing the Impact of Urban Policy' (Robson *et al.*, 1994) and, the interim national and local evaluation of City Challenge (Russell, 1994a and b).

In recognition of tradition in other disciplines, and in search for transferable lessons into the field of urban policy, evaluation experience in the fields of regional economic policy (in Britain) and the appraisal and evaluation of trunk road and motorway schemes (in UK, USA and the Netherlands) are also reviewed. The primary focus in chapters 8, 9 and 10 is on the methodology adopted, why it was adopted and the problems encountered and how they may have been overcome, if at all. The critical review of these methodologies centres around four issues: the measurement of impact, the treatment of the counter-factual problem, assessing the achievement of objectives and explanation of the results.

The final chapter represents the attempt to bring these different strands all together in the form of a systematic approach for comprehensive evaluation of urban policy. The choice of the 'components' of the approach is based, to a large extent, on the first part of the research. However, the organisation of the approach and the exact role each tool can play are greatly informed by the second part. The proposed approach is first presented in its 'full mode'. The flexibility of the approach is then demonstrated by the different modes it can adopt in accord with either resource constraints or political demands, or both.

PART 1:

FORMAL EVALUATION METHODS

INTRODUCTION TO PART 1

Part 1 is devoted to the analysis of 'formal evaluation methods'. It examines the applicability of existing appraisal methods, such as the planning balance sheet, in ex-post evaluation. It also examines the role both monitoring and implementation analysis can play in a comprehensive evaluation framework, the main objective of this research.

Chapter 2 summarises the results of the first step towards the achievement of this objective. That was a wide-ranging literature review of the state-of-the-art of evaluation research. It begins with a broad definition of evaluation and its role within a continuous planning process. It also defines the different types of evaluation and explains its prime function. It then examines the influence and consequences of the highly politicised planning and implementation process. One such consequence is that, to fulfil its function, evaluation will have to be multi-faceted and draw on several traditions at once. However, the literature review revealed a dearth of research in this direction. A notable exception is the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) which aims to synthesise various traditions so as to enable a comprehensive evaluation. This approach is adopted as the second starting point for the research. The first, and related point, is the existing appraisal methods. These two points together set the scene for the remainder of this Part.

Chapters 3 and 4 examine the applicability of existing appraisal methods in ex-post evaluation and the choice among them. Chapter 3 is devoted to the three related methods for economic evaluation: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and value for money (VFM). Chapter 4 examines the planning balance sheet (PBS), goals-achievement matrix (GAM) and multi-criteria evaluation (MCE) methods. The focus in each of the two chapters is on the underlying assumptions of each method and its strengths and weaknesses so as to enable an informed choice among them. These methods are examined in this particular order for one simple fact; each method seems to be a development over, or a response to the limitations of, its predecessor. CEA is an attempt to incorporate intangibles into

economic evaluation; a problematic issue in CBA. VFM, a cross between CBA and CEA, is apparently an attempt to facilitate economic evaluation in the public sector. The PBS attempts both to broaden traditional CBA and overcome its limitations. GAM starts from a severe criticism of both CBA and PBS and departs, substantially, from both. GAM, itself, marked a new direction in evaluation research that has come to be known as MCE.

Monitoring is the subject of chapter 5 which starts with the definition of monitoring and its role in the planning process. This is followed by a review of the two distinct views of monitoring: the control and the extended. The influence of the political arena on the approach to, and conduct of, monitoring is then examined. It also explains the need for both a diverse information base and a computerised information system and the problems associated with both. In concluding this chapter, it is explained why the control view is rejected in favour of the extended one and the vital role monitoring can play in evaluation is explained.

Chapter 6 is devoted to implementation analysis; the final component of the 'Integrated Evaluation Approach'. Implementation analysis is first defined and its scope and the questions it asks are explained. The two analytical approaches that dominate the field - the top-down and bottom-up - are then reviewed followed by one of the most notable attempts of synthesis. The strengths and weaknesses of top-down and bottom-up models are also examined in order to inform the choice on which model to be adopted. This choice depends largely on the dominant policy-making mode. However, the synthesis of approaches represents a good starting point for analysis.

CHAPTER 2: EVALUATION - AN OVERVIEW

2.1 INTRODUCTION

The main objective of this research is to develop a framework for comprehensive ex-post evaluation of urban strategies. 'Comprehensive' is defined as the attempt to answer *all* the questions stakeholders involved in the planning and implementation process are likely to raise at different stages of this process. The need for such a framework should be self-evident. There is widespread agreement that the main purpose of evaluation is to inform stakeholders so that better informed, more rational choices are made. If evaluation is to fulfil its prime function, it follows, timely answers to these questions will have to be provided. Since these questions are likely to vary in their nature and focus, evaluation will have to be multi-faceted and will have to employ several analytical tools. Traditional evaluation research, with its single-vision focus, is not equipped to meet these demands.

This chapter summarises the findings of the first step towards the achievement of this objective. This was a wide-ranging literature review of the state-of-the-art of evaluation research. To avoid repetition, however, several topics have been relegated to later chapters (eg evaluation methods). In section 2, evaluation is first defined in very broad terms and its role within a continuous planning process is explained. This leads to one of the most common classifications of evaluation: ex-ante and ex-post. Other types of evaluation are then defined and the section concludes with an answer to the

question: what is the purpose of evaluation? It is argued that the prime function of evaluation is to inform policy-makers and different stakeholders in order that more informed and rational decisions are made.

Section 3 tackles one of the most problematic dimensions of evaluation: its political and administrative context. It follows from the function of evaluation that it is intertwined with politics. The characteristics of the political arena have their significant repercussions on the methodology and focus of evaluation. One such characteristic is the multitude of stakeholders involved in the process. Another is the style and mode of the policy-making process itself. A third feature is the difference between political and evaluation timescales. These all lead to various constraints on evaluation. The neglect of these facts will have dire consequences for evaluation. Evaluation research will have to focus on, and explicitly incorporate, the political dimension.

Section 4 covers a closely related issue to the political arena: the utilization of evaluation results. It is argued (eg Rossi and Freeman, 1993: 443), quite rightly, that the worth of evaluation must be judged by its utilization. Yet, the issue is far from being straightforward and hinges, probably not surprisingly, on an understanding of the political conditions under which evaluation is conducted and will (may) be used.

One of the most important findings of the literature review was the severe lack of research on comprehensive evaluation. A remarkable attempt in this

field is the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984). This approach aims to create a synthesis of various evaluation traditions so as to enable a comprehensive evaluation with a multi-group perspective. Given the main objective of this research, the approach represents a valuable starting point that warrants closer examination. This is the subject matter of section 5. The questions the approach aims to answer, its components and the role of each and its advantages and pitfalls are all reviewed.

In concluding this chapter, section 6 explains why the 'Integrated Evaluation Approach' is adopted as one starting point for the research. The question, however, arises about the starting point of the evaluation *process*. The approach also raises a question about the focus of both monitoring and implementation analysis. Questions also arise about the applicability of existing evaluation (or rather, appraisal) methods, eg cost-benefit analysis, in ex-post evaluation.

These questions, between them, serve to set the scene for the remaining chapters of this Part. Chapters 3 and 4 assess both the applicability of appraisal methods in ex-post evaluation and the choice among them. Chapters 5 and 6 examine the scope and role of monitoring and implementation analysis, respectively, in a comprehensive evaluation approach.

2.2 DEFINITIONS, TYPES AND FUNCTION

To Evaluate "1. to ascertain or set the amount or value of. 2. to judge or assess the worth of." (Collins Concise English Dictionary, 3rd ed, 1992: 442)

The common-sense definition of evaluation is "to assess the worth or value of". Yet, there appears to be no single accepted definition of evaluation; any definition is likely to reflect either a particular methodological bias or a certain perspective on the political nature of evaluation, or both (Palumbo, 1987: 15 and Patton, 1987: 103). However, and in very broad terms, evaluation may be defined as

the application of systematic research methods to the assessment of program design, implementation and effectiveness (Chelimsky, 1985b: 488).

or,

a robust arena of activity directed at collecting, analyzing, and interpreting information on the need for, implementation of, and effectiveness and efficiency of intervention efforts (Rossi and Freeman, 1993: 3).

There is no doubt – as clearly indicated by the above two definitions – that evaluation "pervades the planning process" (McAllister, 1980: 5-6). The traditional concept of the planning process (survey - analysis - plan) has long given way to a more 'cyclical' process where each cycle builds upon the lessons of the previous one(s) (eg Friend and Jessop, 1977 and Bracken, 1981: 69, see Fig 2.1). In this cyclical process, evaluation "must take place at a number of different stages in the process in order to contribute to 'continuous refinement'" (Bracken, 1981: 69). For instance, the appraisal of alternatives may lead back to the 'design' cycle if many of them were judged to be of poor quality. Ex-post evaluation should feed into policy appraisal, if better

informed and more rational decisions were to be made.

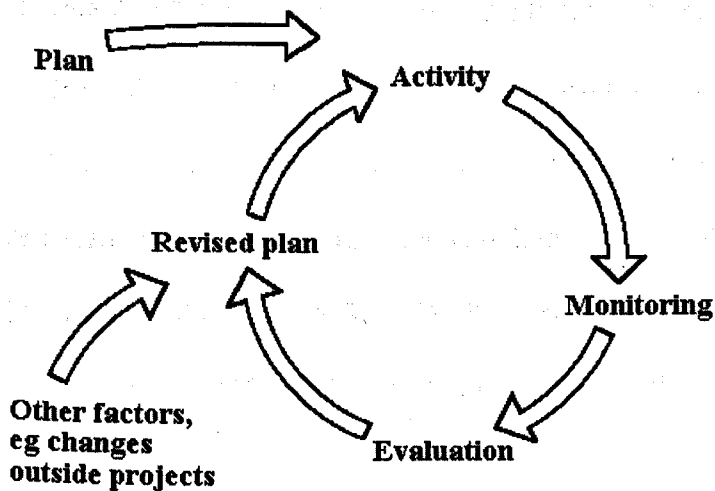


Fig 2.1: The Planning Cycle

Source: modified from Connor, 1993: 9.

It is not surprising then that the most common classification of evaluation types relates to the stage(s) of the planning process at which it takes place. Basically, the distinction is based on the fact whether evaluation is carried out before or after implementation of the plan had started; ie ex-ante and ex-post evaluation. Ex-ante evaluation – also known as policy appraisal (eg HM Treasury, 1988: 26) – aims at assisting policy-makers in making their choice(s) among the different alternatives. It is a "set of activities to classify and conveniently arrange the information needed for a choice in order that the participants in the choice process are enabled to make this choice as balanced as possible" (Nijkamp *et al.*, 1990a: 15). Hence, the focus in ex-ante evaluation is on anticipated effects of each alternative, and the overall worth of each, for evaluation takes place *before* implementation.

In contrast, ex-post evaluation attempts to estimate the impact of the chosen policy *during* or *after* implementation. It is "the process of examining a policy while it is in operation or after it has come to an end" (HM Treasury, 1988: 1). Ex-post evaluation can focus on one, or more, of several issues, eg the achievement of goals or the implementation process (see below). That is, whereas ex-ante evaluation has a 'forward looking' nature, ex-post evaluation has a 'backward looking' nature (Voogd, 1983a: 15-7 and Nijkamp *et al.*, 1990a: 15; see also Rossi *et al.*, 1979: 32-45 and Hill, 1985a: 27).

Ex-ante evaluation can be further broken down to *a priori* (discrete) and *a posteriori* (continuous) evaluation (Nijkamp *et al.*, 1990b: 150, see Fig 2.2a). In continuous evaluation the alternatives are not known explicitly and, in a sense, emerge during the process, whereas in discrete evaluation the alternatives are known beforehand

Other common classifications do exist. One such classification is based on the types of data and information used whereby three categories are identified (eg Voogd, 1983a: 75-7 and Nijkamp *et al.*, 1990a: 65-6):

1. Quantitative: where the data is measured on a cardinal scale (hard information);
2. Qualitative: where the data is measured on an ordinal, or other non-cardinal, scale (soft information);
3. Mixed: where a combination of both quantitative and qualitative data is used.

A closely related distinction is that of monetary and non-monetary evaluation (eg Nijkamp *et al.*, 1990a: 15-16, see Fig 2.2b). The former is characterised by the attempt to measure all policy impacts in their monetary values. Non-monetary evaluation, on the other hand, employs a wider range of measurement scales, one of which is money. Cost-benefit analysis is the best known example of monetary evaluation whereas goals-achievement matrix and multi-criteria evaluation methods are examples of non-monetary evaluation.

A distinction could also be made between explicit and implicit evaluation (Nijkamp *et al.*, 1990a: 16, see also Voogd, 1983a: 17). In explicit evaluation, "a distinct systematic analysis is pursued", whereby the focus is on the transparency and accountability of the final result. In contrast, implicit evaluation "focuses on the consensus of thought, whereby attention is directed toward the participation of—and negotiation between—all parties concerned."

Another classification distinguishes between internal and external evaluation. In some cases, evaluation is carried out by a group within the same organisation responsible for planning and/or implementing the programme. In others, a separate agency — a consultant or a research firm — is called upon to conduct the task. These two types are known as 'in-house', or 'internal', and 'consultant', or 'external', evaluation, respectively (eg Voogd, 1983a: 217; Posavac and Carey, 1989: 18 and Rossi and Freeman, 1993: 439).

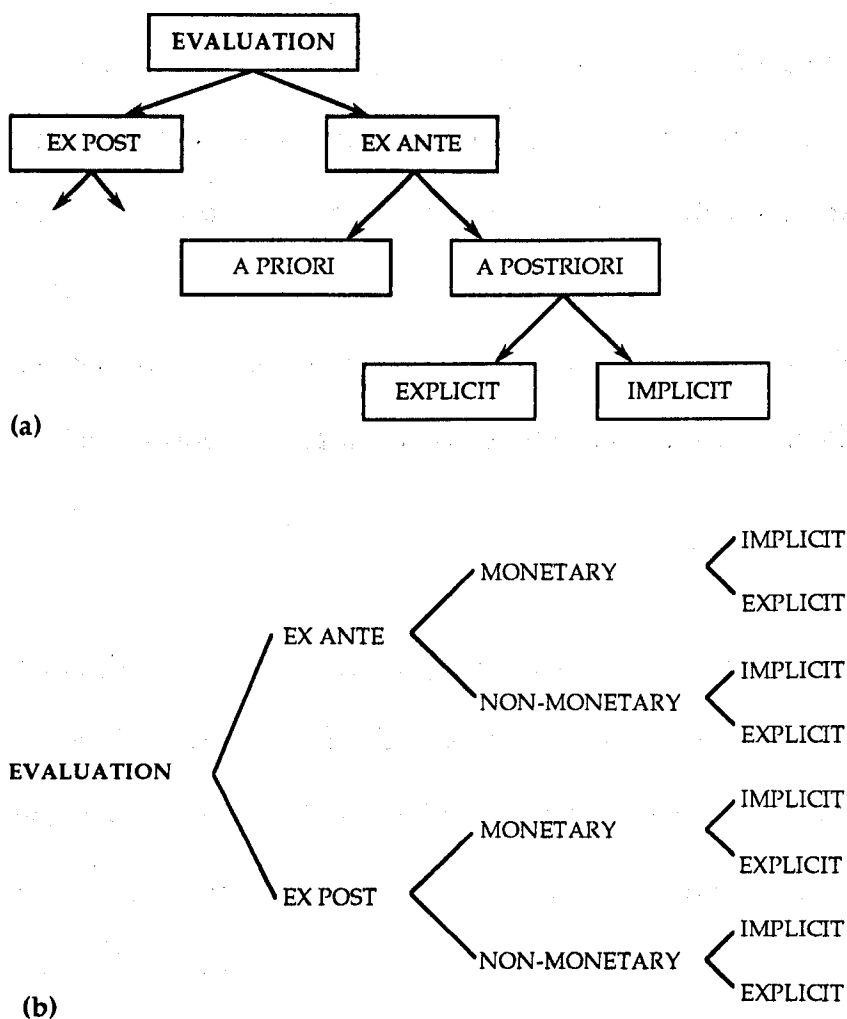


Fig 2.2 a & b: An Overview of Some Major Types of Evaluation

Source: (a) Voogd, 1983a: 16, and (b) Nijkamp *et al.*, 1990a: 15.

There is as well the distinction between formative and summative evaluation, also known as process and outcome evaluation, respectively. Formative (process) evaluation "concentrates upon assessing how a policy is put into practice, what happens on the ground, and relating this to how the policy is meant to work" (DoE, 1992: 6); ie implementation analysis (see chapter 6). Summative (outcome) evaluation, on the other hand, "is more concerned with the final impact of the project" (DoE, 1992: 6).

These classifications, however, are not disjunctive; they merely serve to highlight the various foci of and approaches to evaluation. For instance, an ex-post external evaluation can be carried out to examine explicitly the efficiency of a programme using (quantitative) monetary data.

The question may then be raised: 'Why evaluate? What is the purpose of evaluation?'. Hill (1985a: 31) summed up the answer as

If it does not sufficiently inform the decision-makers and the public so that they can use the information provided in order to arrive at more rational decisions, evaluation is an academic exercise

This view is widely accepted. Hart (1991: 260) asserted that the purpose of evaluation is to "inform policy-makers and to improve the policy-making process" (see also Barnekov *et al.*, 1990: 5 and Miller, 1990: 118). This means that measuring the success of evaluation should also assess the degree to which it was useful to different stakeholders, not only the extent to which it may have improved knowledge in a particular area (Chelimsky, 1987a: 72). At the 'operational' level, the DoE, for instance, considered it a "fundamental objective" of evaluation practice to "provide policy makers with information which will enable them to improve on existing policies" (DoE, 1992: 3; see also Rossi and Freeman, 1993: 34).

2.3 THE POLITICAL AND ADMINISTRATIVE CONTEXT

"Evaluation is inherently political, and cannot be otherwise" (Patton, 1987: 102). Evaluation and politics are entwined in a number of ways (eg Palumbo, 1987: 11; see also Chelimsky, 1987b and Weiss, 1987). Evaluation results can

be an input to a highly politicized process of bargaining, negotiation and often conflict among the various stakeholders. The choice of programmes to be evaluated, in the first place, and the questions evaluation will ask are both the outcome of a political process. Evaluation has its political stance as well. By asking how effective a programme is in achieving a given set of goals, evaluators have accepted the desirability of achieving those goals. Furthermore, by judging how well a programme is doing, evaluation is inherently and unavoidably political.

The political dimension has its significant repercussions on the methodology and conduct of evaluation. One particular characteristic of the political arena is the existence of a multitude of 'stakeholders' with multiple and often conflicting interests (eg Voogd, 1983a: 7; Chelimsky, 1987a: 76-80; Palumbo, 1987: 19 and Patton, 1987: 109) - the 'multi-organizational nature of public policy (see chapter 5, page 128). In broad terms, stakeholders are "individuals and groups who hold competing and sometimes combative views on the appropriateness of either the program or its evaluation, and whose interests will be affected by the outcome" (Rossi and Freeman, 1993: 110).

This phenomenon of multiple stakeholders leads to several strains that will invariably result from the conflicting interests of these stakeholders. First, analysts are faced with the question of whose perspective should be taken in designing an evaluation. Different stakeholders are very likely to be concerned with different issues, ask different questions and require different

information. This, in turn, entails a different focus and methodology of evaluation in each case. Second, evaluation sponsors may turn on the evaluators if the results do not support the programmes they (the sponsors) advocate. "There are legitimate grounds for concern: sponsors are a major source of referrals for additional work in the case of outside evaluators, and the providers of paychecks for inside ones" (Rossi and Freeman, 1993: 410). Third, there are difficulties of communicating with these different groups due both to different information requirements and different vocabulary.

Another crucial concern is the policy-making process itself which can take on one, or more, of a number of different modes, or styles, eg centralised, corporate or participatory⁽²⁾ (Hill, 1985a: 12-15; see also Hart, 1976: 182-200; Miller, 1990: 121-123 and Nijkamp *et al.*, 1990a: 11). The distinction between these modes is based mainly on the degree of centralization of power and responsibility of those making the decision. In the centralised (control) mode, authority is held by a single government agency. In the corporate mode, power is decentralised among a small number of organisations, in and out of government. In the participatory mode, power is dispersed among many actors. Each decision-making style has its own characteristics and entails a different mode of planning and implementation. It is argued, quite rightly, that evaluation strategy will have to vary in accordance with the pattern of decision-making (Hill, 1985a: 13). In addition, and related, the style of policy-

⁽²⁾ The attributes of these modes are explained, in more detail, in chapter 6; see Table 6.1, p. 172

making will impact on the participating stakeholders: who participates and what role each will play (see above).

A further problematic consequence of the political nature of evaluation stems from the difference between political and evaluation timescales (eg Chelimsky, 1987b: 25; Weiss, 1987: 55 and Rossi and Freeman, 1993: 420). Whereas evaluation may take years to complete and document, policy-makers have a much shorter time frame – sometimes months. This exerts pressures on evaluators both to complete the analysis more quickly than ‘best’ methods may allow and to release preliminary results.

It should be clear then that neglect of the political context will have dire consequences for evaluation. To put it rather briefly, if evaluation is to play a role in the policy-making process, it "must focus on and explicitly incorporate political dimensions" (Palumbo, 1987: 13, see also Chelimsky 1987a and b; Patton, 1987 and Weiss, 1987).

At the outset, a great deal of thought, care and attention has to be given to translating policy questions into evaluation questions which are more researchable (Chelimsky, 1987b: 27). This calls for a better understanding, on the part of evaluators, of their relationship with stakeholders, as well as the relationships among stakeholders (Rossi and Freeman, 1993: 407). Equally important, evaluators must consider carefully the needs of each of such a wide audience (Chelimsky, 1987a: 79). The relevance of evaluation and its use

depend largely on the extent to which it addresses and incorporates the interests of stakeholders in both its design and analysis (Palumbo, 1987: 37).

2.4 UTILIZATION OF EVALUATION RESULTS

The worth of evaluation must be judged by its utility (Rossi and Freeman, 1993: 443). This follows logically from the prime purpose of evaluation: to inform those participating in a decision-making process in order to make rational, more informed decisions (eg Hill, 1985a: 31, see page 27). More often than not, evaluators are disappointed about the extent to which the results of their research have been put to use by policy-makers and stakeholders (eg Chelimsky, 1987b: 24-25 and Rossi and Freeman, 1993: 444). The lack of utility can be attributed to one, or more, of several inter-related factors that emanate from the political nature and arena of evaluation.

First, evaluators and administrators — policy-makers in general — operate under different, competing imperatives which set the stage for conflicts (eg Palumbo, 1987: 27; see also Chelimsky, 1987b: 27). Second, and related, evaluation may turn out to be insensitive to various policy issues involved; it may alter policy questions asked by policy-makers and the results may have no policy significance (eg Chelimsky, 1987b: 27 and Rossi and Freeman, 1993: 425). Third, it is not unusual for a communication gap to exist between evaluators and users of evaluation results. One explanation for that is the different, contrasting attitudes both parties hold towards evaluation. For a long time, evaluators adopted a 'decisionistic' view that evaluation results are

to be fully utilized and be the key determinant in policy-making (Richardson, 1982: 177). Evaluation, however, is but one ingredient in the political process (eg Rossi and Freeman, 1993: 417). Moreover, policy-makers "believe it is their job to carry out their constituents' demands regardless of the results of evaluation – only one contending judgement among many desiderata" (Lipton, 1992: 175). Another contributing factor to this persistent impaired communication is "the differences in education and training, world view, expectations and frames of reference, constituencies, success criteria and language" between the two parties (Lipton, 1992: 178).

A closely related issue to, and determinant factor of, utilization is the presentation and reporting of evaluation results. It is argued, quite rightly, that "the first step in having findings utilized effectively..., [is] writing and presenting a report in a manner that will get it read" (Posavac and Carey, 1989: 268). However, it is not uncommon to find that conclusions are hidden in huge quantities of information probably with sophisticated mathematical procedures (Voogd, 1983a: 11). Although evaluation has grown as an industry, "the reports grew no clearer and the results no less equivocal" (Lipton, 1992: 176).

The resolution of these problematic concerns is in a better understanding of the political conditions under which evaluation is conducted and will be used (Chelimsky, 1987b: 31; Posavac and Carey, 1988: 27-30; Lipton, 1992: 184-188 and, Rossi and Freeman, 1993: 447-449). One such condition is that policy

questions posed must be those of fundamental interest to intended users. A second condition is that evaluation must answer these questions. A third political condition is credibility. "If the evaluation is assailable on grounds of poor methodological quality or ... partisanship, this will reduce the use made of its findings" (Chelimsky, 1987b: 31). The fourth condition is the timeliness of evaluation results. As noted above, politicians and evaluators have different timescales. Evaluators have to be concerned about the timing of their final product and how it relates to the policy cycle. Last, but not least, utilization and dissemination plans should be part of the evaluation design.

2.5 THE INTEGRATED EVALUATION APPROACH

One possible way of defining comprehensive evaluation is that it is the attempt to answer *all* the questions different stakeholders involved in the planning and implementation process are likely to raise; and to provide these answers at the time when they are most needed. That is, comprehensive evaluation will ask impact, process, efficiency, effectiveness and equity questions; it will have to provide information for ongoing and periodic decisions.

Many of the problems with traditional evaluation research arise from its single-vision focus on the assessment of outcomes without considering the process behind them. This approach has been criticised for: neglecting the needs of policy-makers, often remaining unused; ignoring the goals of participants other than high level officials; taking too long and costing too

much; assuming a fixed set of goals; and for using sophisticated and costly testing procedures (eg Alterman *et al.*, 1984: 381; Alterman, 1987: 348 and Hill *et al.*, 1990: 44). In methodological terms, the choice of a method seemed to depend "more on the professional training of the evaluator than on the nature of the problem or the questions to be decided upon" (Carmon *et al.*, 1980: 3). For instance, economists tend to be engaged only in cost-benefit analysis, or cost-effectiveness analysis. Although attention has been devoted recently to 'social' costs and benefits, these are generally translated into monetary terms.

It is in the integration of various traditions that evaluation can fulfil its function: to inform stakeholders, with the appropriate timing, so that they can make better informed, more rational choices and decisions. Yet, research in this direction is quite rare. Few, within the domain of social policy, have viewed the various approaches as complementary, rather than alternatives, to each other (eg Rossi *et al.*, 1979; Posavac and Carey, 1989 and Rossi and Freeman, 1993). Comprehensive evaluation was based on the sequential application of types of evaluation; the need for subsequent types is justified by the findings of the previous ones. For instance, it is argued that there is no reason for impact assessment unless monitoring indicated that outputs have actually been delivered on the ground (Rossi and Freeman, 1993: 167). However, this view ignores the fact that sometimes the mere intention of implementing a programme may lead to (significant) costs and outcomes.

In the light of these facts, the proposal for an 'Integrated Evaluation

Approach' (Carmon *et al.*, 1980 and Alterman *et al.*, 1984) is quite unique within evaluation literature. This approach aims to create a synthesis of the several evaluation traditions so as to enable a comprehensive, decision-oriented evaluation with a multi-group perspective (eg Hill *et al.*, 1990: 45). To achieve that, the approach draws on five distinct traditions: outcome evaluation, process evaluation and implementation analysis, monitoring, economic evaluation and, goals-achievement evaluation (see below). Although in this particular instance it was "particularly tailored to the evaluation of broad-aim social programs" (Alterman *et al.*, 1984: 381), there is no reason why the approach cannot be applied to other types of programme, including urban ones.

The approach was intended to answer the following main questions (Carmon *et al.*, 1980: 4-5 and Alterman *et al.*, 1984: 382):

1. Is the programme being implemented? To what extent does implementation conform with plans?
2. What are the costs? Who bears the costs? Is this distribution in accord with plans?
3. Is the implementation process effective?
4. What are the programme outcomes? Who are the beneficiaries?
5. What are the conditions (social, economic, political and administrative) that enhanced or hindered the achievement of goals?

It should be clear that answers to the first three questions are essential for ongoing decision-making; answers to all questions, in particular the latter two, are essential for periodic decisions concerning the continuation, modification

or termination of the programme (Carmon *et al.*, 1980: 5).

The initial proposal for the approach sought to answer the above questions by means of six components⁽³⁾ (see Table 2.1):

1. Identification of the relevant publics and their goals;
2. Description of the programme and the extent to which it is being implemented;
3. Monitoring (of outputs and costs);
4. Implementation process evaluation;
5. Measuring programme outcomes;
6. Evaluation of outcomes from a multi-group perspective.

The approach can be tailored to different situations (Carmon *et al.*, 1980: 16). For instance, a local authority may be interested in evaluation only as an aid to ongoing decision-making (continuation, or otherwise, of the programme is the responsibility of central government). In such a case, the local authority is likely to carry out intensively the third and fourth components of the approach. On the other hand, when the future of a programme is in question, "the performance of all six components of integrated evaluation is essential in order to make well-informed decisions" (p. 16). In addition, the level of detail of the analysis in each component is obviously dependent on several factors, eg the nature of the programme and the resources available for evaluation⁽⁴⁾.

⁽³⁾ The term 'components' was preferred to 'stages' for it was believed that the decision-making process is seldom sequential and thus a strict order in the evaluation process should be avoided (Carmon *et al.*, 1980: 5 and Alterman *et al.*, 1984: 382).

⁽⁴⁾ The same argument applies for the modified version of the approach (Alterman *et al.*, 1984).

Table 2.1: Initial Components of the Integrated Evaluation Approach

Component	Questions/Focus	Comment
1. Identification of the relevant publics and their goals	a) Who are the parties involved in, interested in or affected by the programme and what are their respective goals?	A logical starting point as the approach is "primarily concerned with the effects of the program on the multiple interest groups affected by or influencing the program" (Carmon <i>et al.</i> , 1980: 4).
2. Description of the programme and the extent to which it is being implemented	b) Translating programme goals into operational, measurable objectives (quantitatively, if possible) a) Who should do what, when and for whom? b) The social, economic, political and administrative context of the programme	These represent the "set of outcomes which will be measured in the evaluation process" (Carmon <i>et al.</i> , 1980: 7).
3. Monitoring (outputs and costs)	a) What is being delivered, to whom and when? b) What are the costs and opportunity costs? c) Who bears the costs? d) How do the findings compare with the plan?	The description is needed in order to determine whether the programme is indeed being implemented.
4. Implementation process evaluation	a) How does implementation take place? Who are the political and bureaucratic parties involved? How do they interact and what decisions do they make? b) How effective is the implementation process? Do the decisions being made enhance the chances that programme performance will be in accord with plan?	This information is necessary to provide the basis for the analysis of the conditions which encourage or prevent the implementation of the programme and the achievement of its goals.
5. Measuring programme outcomes	The list of outcomes is based on the operational expression of the goals (1 st component)	This information is necessary both for ongoing management (to keep implementation in line with plan) and for periodic decisions (programme continuation, modification or termination).
6. Evaluation of outcomes in light of the goal preferences of the relevant publics (Goals-Achievement Matrix).	This component is intended to provide policy-makers with a "clear statement of the achievements of the program from the points of view of the various publics who are involved in it" (Carmon <i>et al.</i> , 1980: 13).	The causal relationship is a primary problem. The experimental or quasi-experimental approaches were not recommended. Before-after, and when possible time series, analyses together with monitoring and the implementation process evaluation were believed to enable analysis to trace much of the actual relationship between programme activities, its outputs and its outcomes. The level of achievement of each goal is determined by aggregating the level of achievement of the various outcomes which relate to it. The information needed has already been generated through the previous components. In addition to the information contained in the matrix, an answer should be provided to the question: What factors have facilitated or hindered the achievement of the goals of the programme? The material needed is provided by the 2 nd and 3 rd components.

Source: compiled by the author drawing on material from Carmon *et al.*, 1980.

The approach was later presented in a slightly modified format (Alterman *et al.*, 1984). The number of components was then reduced to four:

1. Monitoring of outputs and costs
2. Implementation process evaluation
3. Economic evaluation (cost-effectiveness analysis and distributional equity).
4. Evaluating programme outcomes from a multi-group perspective.

However, a close examination of the tasks involved in each of the four components (Alterman *et al.*, 1984) reveals no fundamental differences between the two versions of the approach. The first, second and fifth components of the initial proposal (Carmon *et al.*, 1980) are no longer explicit elements of the modified approach; they, however, remain incorporated within the approach. 'Identification of the relevant publics and their goals' is a task that has to be accomplished both prior to goals-achievement analysis (the final component of both versions) and as part of the implementation process analysis. 'Description of the programme' has become a task of monitoring in the modified version, by adding the question: what has been planned? Both 'measuring programme outcomes' and 'translating goals into operational objectives' have become tasks within the fourth component. On the other hand, 'economic evaluation' has become a distinct component in the modified version rather than being a task within monitoring in the initial proposal.

In both the initial proposal (Carmon *et al.*, 1980) and the modified version (Alterman *et al.*, 1984), it was recognized that generalizations about the

methodology of implementation analysis were not possible. Considering that each implementation set-up and process is unique, "the exact questions posed and the methods suitable for tackling them will have to be tailored to each case; in fact, that is as it should be" (Carmon *et al.*, 1980: 10 and Alterman *et al.*, 1984: 384).

The only documented application of the approach was in the evaluation of Israel's Project Renewal (eg Hill, 1986; Alterman, 1987 and Carmon, 1987). The Project was a large-scale, comprehensive neighbourhood rehabilitation programme. Between 1979 (the year it started full operation) and 1985, a sum equivalent to \$600 million was spent on its programmes (Hill *et al.*, 1990: 43). By 1987, the Project covered 90 neighbourhoods of varying sizes throughout Israel, representing some 15% of the total population (Alterman, 1988: 455).

The evaluation of the Project was commissioned in 1982; it took a team of 20 researchers some four years (1982-1986) to study the Project through a sample of 10 neighbourhoods (eg Carmon, 1987: 363 and Hill *et al.*, 1990: 47).

The integrated approach, as applied in this case, was yet of another slightly different format having six components (Hill *et al.*, 1990: 46-47):

1. Evaluation of the implementation process (bottom-up approach),
2. Evaluation of citizen participation,
3. Monitoring of outputs,
4. Economic evaluation (cost-effectiveness analysis),
5. Evaluation of outcomes,
6. Goals-achievement evaluation.

Citizen participation is undoubtedly an aspect of the implementation process to be included in its analysis. It was, however, singled out because of its "special nature as both a substantive goal and a method of operation" (Hill *et al.*, 1990: 46). 'Evaluation of outcomes' was a component of the initial proposal (Carmon *et al.*, 1980). In all three versions, this component is intended to establish the causal relationship between observed changes and the policy being evaluated. In all three versions, this was accomplished by means of before-after and time series analyses, aided by monitoring and implementation analysis.

The approach, in any of its three forms, has its advantages over traditional evaluation (Alterman *et al.*, 1984: 387-388):

1. it takes into account the aims of the various parties involved in the programme, not only those of decision-makers;
2. it takes into consideration the distributional effects of the programme;
3. it recognizes the existence of several sets of goals that are likely to change over time;
4. economic costs are taken into account but without necessarily translating effects into monetary terms, unlike cost-benefit analysis;
5. at least some of the results can serve the decision-makers during the course of implementation;
6. connections between inputs and outcomes are traced by continuous monitoring not the almost impossible task of finding matched control groups for experimental designs;
7. monitoring and implementation analysis broaden the understanding of reasons of success or failure and thus the lessons that can be learned;

8. identifying relevant groups and their goals and measuring their preferences creates a rich source of evaluation criteria. This will, in turn, reduce the possibility of arriving at erroneous conclusions;
9. measurement criteria are not restricted to the goals of the decision-makers;
10. it applies social science theory at different points (identifying relevant groups, formulation of their goals in operational terms, ...).

Nevertheless, it was reckoned that, as with all methodologies, the components of the approach have their pitfalls (Alterman *et al.*, 1984: 386-387). Monitoring faces two major problems. First, on the input side, there is the difficulty of tracing budgetary displacement (substitution, additionality and dead-weight). Second, and on the output side, there is the great breadth of data that has to be collected from a wide range of sources. Implementation analysis is based on qualitative rather than quantitative information. It is likely, then, that different parties will make different judgements on the basis of the same information. It is also difficult to arrive at accurate assessment of influential factors due to the subjectivity of informants. Moreover, it may prove difficult to separate the effect of individual personalities from those of more structural factors. The quality of cost-effectiveness analysis is dependent on the quality of monitoring inputs and outputs and the measurement of outcomes. The information needed to assess distributional effects may prove difficult to assemble. Furthermore, there is the definitional problem - 'what is an equitable distribution?'. Finally, there are several potential problems with the final component of the approach (GAM); eg identifying the groups and their objectives and preferences (see chapter 4).

In addition, the approach as a whole has its drawbacks. It is very likely to require a considerable amount of resources: personnel, time and financial. Thus, it is not particularly suitable for the evaluation of limited-scope programmes. The broad scope of the approach, and the multi disciplinary nature of the team required to conduct it, both come at the expense of in-depth analysis of specific aspects. Finally, such an evaluation is unlikely to reach a conclusion "which states unequivocally whether the program was a success or a failure" (Alterman *et al.*, 1984: 387).

2.6 CONCLUDING REMARKS

The main objective of this research is to develop a framework for comprehensive ex-post evaluation of urban policies. 'Comprehensive' is defined as the attempt to answer *all* questions stakeholders are likely to pose. The need for such a wide-ranging framework should be self-evident. The multitude of parties who become involved in the planning and implementation process have different (conflicting) interests and will ask different questions at different times throughout the process. If evaluation is to fulfil its primary function, and if it is to inform stakeholders so that better informed, more rational decisions are to be made, timely answers will have to be provided to these questions. In order to do so, different analytical tools will have to be employed at different stages of the evaluation process. Moreover, evaluation will have to adopt a multi-group focus and encompass the different objectives and criteria for success each involved group holds. In other words, evaluation will have to draw on traditions of several disciplinary

fields. Traditional evaluation research, with its single-vision focus, is evidently ill equipped to meet such demands. It is in the integration of these different, yet complementary, traditions that evaluation can answer the various questions likely to be raised.

In the light of these facts, the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) represents a good starting point for the research. The approach recognises the limitations of traditional evaluation and attempts to provide a comprehensive framework by combining different traditions. It also recognises the need for different information by different stakeholders at various stages of the implementation process. And, it takes a multi-group perspective to evaluation.

Nevertheless, the approach is not without its pitfalls. In addition to those mentioned before, two more preliminary remarks can be made. Firstly, it was argued that the political process is seldom sequential and, hence, a strict order should be avoided in the evaluation process (Alterman *et al.*, 1984: 382). While this view is absolutely true, there remains a need to define a starting point for the whole evaluation *process*, without ignoring the possibilities of feedback loops. There are strong indications that different stakeholders share an interest in regular monitoring of the outputs of the policy being implemented. This is true within the context of the latest urban regeneration initiatives in England: City Challenge and the Single Regeneration Budget (see chapters 7 and 8). It is also true in other countries (eg The Netherlands) in the fields of

urban and regional planning and transportation (personal contacts, May 1995). It is quite plausible, then, that monitoring can be the starting point for the evaluation process. In addition to providing timely information to stakeholders, it also makes available data needed for further analysis, both economic and outcome evaluation.

Secondly, monitoring and implementation analysis within the approach appear both to have taken a narrow view focusing solely on the programme being evaluated with no regard to the wider environment within which it operates. It is not uncommon that decisions taken within the remit of one policy impact, positively or otherwise, on the policy being implemented. The uncertainty surrounding the wider environment (eg Friend and Hickling, 1988) is no doubt an influential factor on both the political and implementation processes. In this respect, attention has to be paid to this wider environment and to the decisions and actions taken by other groups. Also related, monitoring will have to adopt a more "futuristic" view (Floyd, 1978; see chapter 5). Plans and decisions under consideration, whether within the programme being evaluated or others, have also to be taken account of. Such decisions, if taken, may impact on the programme at hand and, hence, have to be incorporated in any judgment concerning its future.

With all of this in mind, the research will take the integrated approach as one of its main starting points. A second, and related, starting point is the existing evaluation (or rather, appraisal) techniques, eg cost-benefit analysis

(CBA), planning balance sheet (PBS) and multi-criteria evaluation methods (MCE). Although these methods have been conceived and practised mainly in ex-ante contexts, the integrated approach had shown that two of them can be transferred into ex-post settings: cost-effectiveness analysis (CEA) and goals-achievement matrix (GAM). Are any of the other methods also applicable in ex-post evaluation? One tends to answer positively, if only for one reason. Cost-effectiveness analysis is derived from cost-benefit analysis and goals-achievement matrix is one member of the multi-criteria evaluation methods family. If CEA is applicable in ex-post evaluation, why not CBA as well? Why use GAM rather than any other MCE method? Such questions have to be answered before conclusions about a framework for comprehensive ex-post evaluation can be ventured. In addition, and with regard to the above mentioned remarks, there appears to be a need for a closer examination of the theoretical basis and background of both monitoring and implementation analysis.

The above questions and remarks set the scene for the remaining chapters of this Part. Chapters 3 and 4 assess the applicability of existing appraisal methods in ex-post evaluation and the choice among them. Chapter 5 examines the focus and function of monitoring whereas chapter 6 investigates the role of implementation analysis and the choice of an analytical model.

CHAPTER 3: CBA, CEA & VFM

3.1 INTRODUCTION

There can be no doubt that the economic evaluation of public policies is of crucial importance. One reason is the scarcity of public resources combined with people's increasing demand for (better) services. Another reason is the growing trend towards a more accountable public sector. A third reason is the fact that any public policy entails public expenditure, sometimes substantial, which is incurred for the purpose of achieving policy objectives. Policy-makers are understandably concerned whether this policy and expenditure have had their 'hoped for' impacts. By the same token, evaluation of a public policy should assess the extent to which this expenditure has contributed to the achievement of its objectives. In other words, assessing the economy, efficiency and effectiveness should be among the issues addressed adequately and explicitly in the evaluation of public policies.

Cost-benefit analysis (CBA) is a long-established technique for economic appraisal (and evaluation) of projects. In simple terms, CBA attempts to quantify, in monetary units, and compare the costs of a project to its benefits. It is perhaps the most practised method in the private sector. It has also been, and is being, applied to numerous public projects. However, it has long been recognised that the application of CBA in public decision-making faces severe problems. Notable among these are the difficulty of measuring many of the project's effects in money units and the neglect of distributional effects.

One alternative is cost-effectiveness analysis (CEA) which is described as a 'truncated version' of CBA. CEA draws guidance from only one side of CBA; that is, only one side of the analysis (costs or benefits; mostly costs) is measured in money whereas the other side (mostly benefits) is measured in its original units (physical, social, ...). Another alternative that has been introduced some twenty years ago is 'value for money' (VFM). The term has come to be the 'catchphrase' underlying many of the British government's recent proposals for both services reforms (eg NHS) and urban policies (eg SRB). 'Value for Money' apparently means achieving the highest possible output per given (diminishing) resources.

One starting point for the research is the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984). Two components of this approach were appraisal methods: CEA and GAM. The application of the approach (eg Hill *et al.*, 1990) showed that both can be applied in ex-post evaluation. The question, therefore, was raised about the applicability of other methods (eg CBA and MCE), and the choice among them. This question is at the heart of this chapter, and the following one as well.

CEA is derived from CBA. If the former is applicable in ex-post evaluation, what about the latter? Why choose CEA not CBA? Is VFM any different from CBA and/or CEA? Which of the three methods, if any, is 'more appropriate' for ex-post economic evaluation of urban policies? In order to be able to answer these questions, the three techniques have to be examined in

detail, and in comparison with each other. This is the primary focus of this chapter. We start with CBA because it is the first systematic technique to be established for economic evaluation, being derived from economic theory. CEA has been derived directly from CBA. The term 'value for money', as mentioned above, is probably the most recent in economic appraisal literature. Its introduction may be traced back to mid- to late 1970s, in the UK, with the then central government growing emphasis on local authorities accountability.

Given these facts, section 2, covering the concept and technique of CBA, is the most extensive. The section first defines what is cost-benefit analysis and traces back its introduction. Following a brief account on the applicability of CBA in public decision-making, a distinction is made between financial and social CBA. The various elements of the analysis are then reviewed. Given the extensive literature on CBA, this review is kept to a minimum. The focus in this section, as in the following two, is on the underlying assumptions and the difficulties the technique faces so as to enable an informed choice between the three techniques.

Section 3 centres around those particular aspects of CEA that differ from CBA. CEA is first defined and compared to CBA. Except for the particular aspect of valuing intangibles, the process of the two techniques is virtually identical. However, it is argued that there is a gap in the literature as to how to discount intangibles. The issue, to say the least, is problematic and the alternative proposed is actually not to discount and compare streams of

effects instead of aggregate figures. The decision criterion of CEA is then identified and the difficulties associated with it are outlined.

Section 4 focuses on the recently introduced notion of 'value for money' (VFM). Similar to the previous two sections, VFM is first defined and the driving forces behind the search for it are also highlighted. It is then argued that the concept and scope of VFM have both shifted over time to become more of an objective and a framework for achieving that objective.

The final section of this chapter concludes on the difficulties each of the three techniques is likely to face in ex-post evaluation of urban policy. Perhaps not surprisingly, CBA, CEA and VFM face almost the same problems. CEA, however, has the relative advantage over CBA of allowing intangibles into the analysis in their original measurement units. VFM is not much different from CEA, except for the explicit emphasis on economy, efficiency and effectiveness - the three Es. The conclusion is, therefore, to choose CEA as the basis for economic evaluation - with an explicit emphasis on the three Es. However, the analysis has to adopt a multi-group perspective and the results should be presented in their disaggregated form if distributional impacts are to be taken into account.

3.2 COST-BENEFIT ANALYSIS

3.2.1 Definition

There is no argument that the purpose of evaluation research is to inform decision-makers in order to make more informed, rational decisions. Cost-Benefit Analysis (CBA, synonymous with benefit-cost analysis) "was developed as a technique to serve this very purpose" (Hill, 1968: 19). Broadly defined, CBA is a comparison between the costs of a project and its benefits (revenues). It "attempts to quantify in money terms as many of the costs and benefits as possible, including important items that are non-marketed or where the market does not reflect true economic values" (HM Treasury, 1991: 10). The basic idea in standard CBA is to choose a project only if the sum of its costs and benefits is positive (Copp, 1987: 68). Where several alternatives are considered, that with the highest positive sum will probably be chosen (HM Treasury, 1991: 13).

The introduction and application of CBA date back to the 1930s in the USA in conjunction with the 'Flood Control Act' of 1936 (eg Lichfield, 1964: 163; Hill, 1968: 19; Levin, 1975: 91-2; Nash *et al.*, 1975: 121; and Pearce and Nash, 1981: 1). This is not to say, however, that CBA has been confined to the evaluation of water resources projects. Again in the USA, with the advent of the 'Great Society' programmes in the mid-1960s and the emphasis on accountability and assessment, CBA was applied to social policies "with startling results" (Thompson, 1980: 2).

3.2.2 CBA & Public Decision-Making

The idea of CBA can be made more comprehensible by reference to the theory of the (profit-maximization) firm (eg Niskanen, 1967: 17; Hill, 1968: 19 and Mishan, 1988: xxix). A private firm, in pursuit of maximum profit in a perfectly competitive market, would allocate its resources where it is most profitable. In addition to that condition of a 'perfectly competitive market', the 'private model' of the analysis is also based on three more assumptions: 1) there are no externalities, 2) there are no barriers to the flow of funds and resources and, 3) prices are equal to marginal costs (Hill, 1968: 19 and Good, 1971: 39). Although these conditions are seldom evident in the private sector (Hill, 1968: 19), the application of CBA is, to a large extent, a straightforward task: there is one single objective – to maximize profit – and a limited range of, often easily measured, costs and benefits.

On the other hand, the application of CBA in public programmes is, to say the least, problematic. Public agencies are in pursuit of a wide range of (sometimes conflicting) objectives; many of their programmes' effects can hardly be measured in money; public agencies' freedom to borrow, lend or invest in the market is substantially limited, compared with private firms; and, there are also barriers to the flow of funds and resources (see Table 3.1).

3.2.3 Financial vs Social CBA

An evident consensus throughout the literature is the distinction between 'financial' and 'social' CBA (eg Sugden and Williams, 1978; Thompson, 1980;

Schofield, 1987 and, Mishan, 1988). Financial CBA, in brief, is primarily concerned with the 'monetary' measurement of costs and benefits, with the objective of maximizing profit. It is the 'original' form of CBA, being derived from economic theory.

Table 3.1: CBA in Private and Public Sectors

	Private Sector	Public Sector
Objectives	Single objective: Profit-maximization	A wide range of (often conflicting) societal objectives
Costs and Benefits (Programme effects)	A limited range of often measurable costs and benefits	A wide range of costs and benefits that are not always measurable
Application	Straightforward	Problematic

Source: modified from Lichfield (1966: 215).

However, the financial objective of maximizing net profit "is far too limited to represent the complexity of the public interest in the activities of public agencies" (Sugden and Williams, 1978: 6). More importantly, the assumptions and conditions upon which the 'private model' is based are rarely evident in the public sector (Hill, 1990: 4). The accountancy approach, hence, founders on several difficulties when attempted in public decision-making (Thompson, 1980: 38-9); it can, at best, be used in relation to a restricted class of decision problems (Sugden and Williams, 1978: 7).

Social CBA is, to a large extent, the result of adaptations and applications (of the original form) to social programmes over the last few decades

(Thompson, 1980: 2). Explicit incorporation of social considerations which are beyond the scope of financial appraisal widens the scope of CBA to acknowledge the existence and significance of external values in addition to those of the market (Doeleman, 1985: 150). Social CBA, in a sense, is an attempt to overcome the pitfalls of employing purely financial analysis to social policies. Hence, whereas financial CBA has been, and still is, primarily associated with the private sector, social CBA has come to be largely associated with public sector investment decisions.

Other differences stem from such a distinction. Whereas money is the measuring-rod for costs and benefits that enter the financial analysis, these are substituted with the broader, and less precise, notions of social costs, benefits and welfare (see Mishan, 1988: xxix). In addition, social CBA takes a wider point of view to include the economy of the society as a whole (Schofield, 1987: 2). Further, and related, the incidence of costs and benefits over various individuals (distributional effects) is a major issue in social CBA (though its treatment is debatable as will be explained later).

3.2.4 Elements of the Analysis

Identifying costs and benefits

In CBA, self-evidently, a programme's effects are classified into two categories: costs and benefits. In broad terms, a benefit is any gain in welfare or utility to any individual within the group in question whereas a cost is an impact that entails a loss in welfare or utility (Pearce, 1983: 12). Costs and

benefits to enter the calculation are those that would result solely from the programme at hand. When possible, costs should be estimated on the basis of 'opportunity costs' (HM Treasury, 1991: 12). The concept of opportunity costs refers to "what is forgone as a result of undertaking a project" (Sugden and Williams, 1978: 30). In other words, opportunity costs are "the benefits resources would achieve in alternative employments" (Thompson, 1980: 108).

Opportunity costs are most difficult to measure in the case of public projects. Unlike private firms, public agencies do not enjoy the freedom of allocating their resources where it is most profitable. They are under constant pressure from the public to fulfil demands for more services, most of which are not paid for – not fully at least. Public agencies are also made increasingly accountable to the public. This may encourage a 'risk-averse' culture within public agencies opting for the 'status-quo' alternative rather than taking any risks. The problem is therefore a practical one: to discover what these alternative benefits are worth (Sugden and Williams, 1978: 75). Unfortunately, there seems to be no best, single answer to this question.

Measuring Costs and Benefits

In financial cost-benefit analysis, it is most likely that both costs and benefits can be measured in money units, aggregated and compared with each other using one criterion or another. Even when monetary valuation is not possible, shadow prices (see below), for instance, may be used without provoking much controversy, if at all.

The situation, as may be expected, is almost entirely different for public agencies; ie in social CBA. The differences, and hence the difficulties, stem from two basic requirements of social CBA. First, to identify the effects of the project on the individual welfare of all members of the society. Second, to measure these effects in a common unit (usually money) so that costs and benefits can each be aggregated and then compared with each other (see, for instance, Sugden and Williams, 1978: 82, 89; Thompson, 1980: 15, 147 and Copp, 1987: 66).

Whereas many of the costs and benefits of any public project can be measured in terms of their market values, many others cannot. The most obvious example is environmental effects. For instance, how can we value noise, pollution or loss of open space? Such effects have come to be known generally as 'intangible' effects (eg Pearce and Nash, 1981: 119-20 and Mishan, 1988: 203-4). Their basic characteristic is that they are not traded in the market. When prices are attached to any of these effects, they do not truly reflect their social value (eg imposing fines on pollution-emitting industries may reflect the cost of treatment but not the social value of living in a polluted environment).

However, given the requirement to aggregate costs and benefits, the need arose for an approach to bring intangibles within the framework of the analysis and, whenever possible, within the measuring-rod of money. There are four major approaches for dealing with intangibles: *non-quantification* (eg

Thompson, 1980: 40 and Schofield, 1987: 65; see also HM Treasury, 1991: 52-53); *shadow prices* (eg Sugden and Williams, 1978: 179-180; Thompson, 1980: 40; Pearce and Nash, 1981: 117; Pearce, 1983: 32; Schofield, 1987: 63-64 and Mishan, 1988: 83-84); *cost-effectiveness analysis* (see Section 3.3) and, *compensating variations - CVs* (willingness to pay and willingness to accept) (eg Sugden and Williams, 1978: 8; Thompson, 1980: 40-5, 148; Pearce, 1983: 11; Copp, 1987: 66-67; Schofield, 1987: 37-38; Mishan, 1988: 8, 22 and, Adamowicz *et al.*, 1993: 416).

Externalities

External effects (also known as externalities, side effects and "more suggestively as 'spillover effects'") are an abbreviation for external economies and diseconomies (Mishan, 1988: 116). An external effect is the "incidental impact of an action in the private sector on persons with no decision control over it" (Thompson, 1980: 70). The common characteristic of externalities is their "incidental, or unintentional, nature" (Mishan, 1988: 117; see also Pearce and Nash, 1981: 120). Public decisions are prone to create spillover effects, though the concept becomes "murkier"; "For even the most direct consequences, those affected (...) may not be consulted and may thus feel that they have no control over the decision" (Thompson, 1980: 70).

Externalities are another problematic issue in CBA. First, there is the difficulty of identifying those effects (it may be difficult to identify the *direct* effects of a project, in the first place). Second, the potential number of those

effects is unpredictable. A little reflection is enough to convince us that "the number of external effects in the real world is virtually unlimited" (Mishan, 1988: 122). Third, many of those effects are 'intangibles'. Mishan (1988: 127) urged economists to resign themselves "to the prospect of never being able to internalize ... important environmental spillovers within the market economy".

For less problematic external effects, measurement is a straightforward task since each is nothing but an effect of the project on the individual welfare. Thus, compensating variations (CVs) still holds as the best measurement concept, though only the *difference* in individual's welfare attributable to the project is to be taken into account (see Mishan, 1988: chapter 19).

Discounting

Benefits and costs of any project occur at different times. In public and private sector alike, decisions have very often to be made about whether to incur present costs for future benefits, or vice versa (HM Treasury, 1991: 13). In such cases, it is of little help to state only the aggregated figures of costs and benefits. To compare, or trade-off, between both, they have to be brought together into a specific point in time.

The process of translating values at one time to valuations at a different time is called *discounting* (eg Thompson, 1980: 28). There are two forms of discounting: present and future. In present discounting, values that occur in

the future are translated into their equivalent valuations in an earlier time. In contrast, future discounting is to translate the value of something at one time to valuation at a later time. Present discounting is the form most commonly applied. Present values of costs and benefits are then aggregated to obtain the 'net present value' - NPV - of the project. The *net* present value is the present value of benefits *less* the present value of costs (eg Mishan, 1988: 225 and HM Treasury, 1991: 25).

The Discount Rate

A discount rate is the figure which (through mathematical procedures) relates the value at one time to the value at another time (eg Thompson, 1980: 28). It is a commonplace that the NPV of a project depends to a large extent on the discount rate used. In situations where several alternatives are being considered, the choice, in the sense of highest returns, will also depend on the discount rate (Sugden and Williams, 1978: 20 and Mishan, 1988: 227). In addition, when comparing between two alternatives, there is "a *critical value* of the interest [discount] rate which would lead to the result that the two alternatives were equally preferred" (Sugden and Williams, 1978: 21, stress in original). Thus, careful consideration should be given to the identification of the discount rate to be used in the analysis.

In the private sector, the rate at which a firm can borrow or lend (often known as the cost of capital) is usually taken as its discount rate for project appraisal (Sugden and Williams, 1978: 46 and, Thompson, 1980: 157). In an

analogy to this approach, public projects may be discounted at the *social opportunity cost of capital* (Pearce, 1983: 43). The idea is to calculate the rate of return of alternative employments of the resources, ie through the opportunity cost of resources. The approach is appealing, but only when funds come from borrowing. Furthermore, the difficulty is that public agencies are not free to borrow or lend, and the choice of a discount rate becomes a problem.

There are two major approaches to determine the discount rate for public projects, each has its problems (eg Sugden and Williams, 1978: 43; Pearce and Nash, 1981: 143 and, Mishan, 1988: 292-4). The first approach is the social discount rate which is an average of individuals' 'marginal time preference rates' (MTPRs). The second is the 'internal rate of return'; the discount rate at which the NPV of a project is zero. The decision rule, in this case, is to choose a project if its internal rate of return is greater than the average MTPR (which makes it a somewhat circular approach and of little use - eg HM Treasury, 1991: 26-27).

It is argued (eg Pearce, 1983: 48-50) that when funds - for a public project - come from taxation (people's forgone consumption) the social discount rate (people's time preference) is to be used. Where borrowing (forgone private benefits) is the source of funding, the cost of capital - the interest rate - is what should be used in the analysis.

Discounting and Inflation

"Inflation complicates discounting" (Thompson, 1980: 162). In the presence of inflation, values are bound to change over time. Therefore, adjustments have to be made in order to discover the values of costs and benefits in *real terms*, ie "net of any general movements in price levels" (Pearce, 1983: 40). Such adjustments will give rise to more practical difficulties: determining which effects are susceptible to inflation and which are not; deciding at which rate should inflation enter the calculation, and dealing with several discount rates given that different goods may have different inflation rates.

Discounting and Future Generations

A basic problem with CBA is that "the effect of discounting is to discriminate against the future" (Pearce, 1983: 53). The problem is two-fold, particularly for those projects that yield their benefits (or incur their costs) further into the future, imparting on other generations. First, there is the question whether those alive at the time of making the decision are the proper 'electorate'. Second, no matter what discount rate is employed, near-certain benefits (or costs) that occur, say, 20 years after enacting the project, will have a present discounted value of a fraction. Compared with their capital outlays, many worthwhile projects may never be launched. Unfortunately, "there is no consensus at all on what to do about this aspect of CBA" (Pearce, 1983: 53; see also Sugden and Williams, 1978: 219-21 and, Mishan, 1988: chapter 41).

Uncertainty and Risk

Uncertainty and risk are two further problems with CBA. There is always bound to be "some guesswork about the magnitude of future costs and benefits" (Mishan, 1988: 375). The point is, future "can never be known with certainty" (Schofield, 1987: 78). In the private sector, three common, though rather crude, investment criteria are employed in uncertain, risky situations (Mishan, 1988: 221-4): the cut-off period, the pay-off period and the net average rate of return. Of these three methods only the first one, cut-off period, was seen to be applicable in public decision-making, though it is still "the crudest way of dealing with uncertainty" (Mishan, 1988: 409). The idea is to choose a date in the future "beyond which all costs and benefits are ignored" (Sugden and Williams, 1978: 63). In other words, "to truncate the time horizon of the analysis" (Mishan, 1988: 79)⁽⁵⁾.

Several other methods for dealing with uncertainty and risk have been primarily linked with 'social' CBA. The most common of these are: risk premium (Sugden and Williams, 1978: 60; Schofield, 1987: 79 and, Mishan, 1988: 410), certainty equivalent (Pearce and Nash, 1981: 67-9 and Mishan, 1988: 376-7), probability adjustment of risks (Mishan, 1988: 390), use of the market to yield a discount rate (Mishan, 1988: 414) and, sensitivity analysis (Thompson, 1980: 25, 90; Pearce, 1983: 89; Schofield, 1987: 85 and, HM Treasury, 1991: 60).

⁽⁵⁾ This is the method used by the DoT for appraising trunk road and motorway schemes (see chapter 10).

Net Benefit and Benefit Cost ratio

The most common means to relate aggregate costs and benefits (at present or future values) together is either 'net benefits': $B-C$ or 'benefit-cost ratio': $B \div C$ ⁽⁶⁾. "Sometimes the differences, sometimes the quotients are good decision guides; sometimes neither are" (Thompson, 1980: 71). In general, the more or less standard rules for using these two methods are (see Table 3.2):

- (1) for choices between spending resources on a program and not doing so (the "go-no-go" decisions), the program should be enacted if and only if net benefits are positive, which occurs precisely when the benefit-cost ratio exceeds 1.0; and
- (2) for choices among mutually exclusive, competing programs, that with the highest net benefits is to be preferred. (Thompson, 1980:73)

Table 3.2: Appropriate Decision Criteria

<i>Decision Condition</i>	<i>Appropriate Decision Criterion: $B - C$ or $B \div C$</i>
Go-no-go decision presuming that potential Pareto improvements (other things being equal) are good	either
Decision on mutually exclusive alternate project versions presuming that potential Pareto improvements are good	$B - C$
Decision on projects to complete a roster of fixed total budget size	$B \div C$
Decision on alternate project versions presuming that all unspent monies can go to other projects (otherwise unfunded) with net benefits	neither

Source: Thompson, 1980: 79.

However, each method has its disadvantages. A project with a positive net benefit may still involve considerable costs or risks. Benefit-cost ratios, and indeed net benefit, are deceptive in that they effectively compare alternatives

⁽⁶⁾ There are other, less common, alternatives, eg the gain-loss ratio and the public-private ratio; see Thompson, 1980: 82-86.

with the status-quo, null option of doing nothing (Thompson, 1980: 78), obscuring the 'relative-merit' of alternatives when compared together. Clearly, a better and reliable technique would be to compare best alternatives directly⁽⁷⁾. "Unfortunately, this is unwieldy: To select one project from n possible alternatives, it is necessary to make n-1 comparisons" (Thompson, 1980: 75). Further and related, net benefits and benefit-cost ratios both fail to provide a rational basis for decisions when funds have opportunity costs. Opportunity costs are reflected in neither of the two calculations (Thompson, 1980: 76-7).

3.2.5 Public Decision-Making Criteria and Distributional Effects

Is a positive net present value, or a benefit-cost ratio that is greater than 1.0, a sufficient criterion for the decision-maker to decide whether or not to enact a project? In conventional CBA, the answer is 'yes'. The argument is based on a concept that lies at the heart of cost-benefit analysis; namely, the Pareto improvement (eg Pearce and Nash, 1981: 28; Pearce, 1983: 16 and, Schofield, 1987: 11). A Pareto improvement is

any change such that at least one person is made better off and no one is worse off (Thompson, 1980: 43)

The criterion is apparently almost impossible to satisfy. The standard choice rule thus becomes what was described as a "diluted version" of the Pareto improvement: the *potential* Pareto improvement criterion (eg Sugden

⁽⁷⁾ In contrast, the 'Green Book' argued that a "do minimum" or "do nothing" option should normally be identified as a base case" (HM Treasury, 1991: 17).

and Williams, 1978: 89; Pearce, 1983: 16 and, Copp, 1987: 66). That is

any change such that, *with suitable hypothetical redistributions*, at least one person would be made better off and no one is worse off (Thompson, 1980: 43, stress in original)

The potential Pareto improvement criterion (pPi) is sometimes referred to as the Kaldor-Hicks test, both terms are used interchangeably (eg Nash *et al.*, 1975: 126; Pearce, 1983: 16, and Schofield, 1987: 21). The approach is simply to add up individual compensating variations - CVs. The result is thus the net benefit of the project. The Kaldor-Hicks test argues basically that a positive sum of CVs is a necessary and sufficient condition for enacting a program (eg Thompson, 1980: 42)

The potential Pareto improvement, or the Kaldor-Hicks, criterion clearly ignores the resulting change in income (or rather, welfare) distribution (Mishan, 1988: 169). The rule effectively ensures that the efficiency of a proposal is dependent on the *sum* of individuals' CVs not on their distribution among individuals (Copp, 1987: 67). These hypothetical redistributions never happen in the real world. Nash *et al.* (1975: 126) asserted that they "know of no government decision that has ever brought a Pareto improvement in welfare in its strict form."

The 'orthodox' reply of the exponents of the technique is that it is not the job of cost-benefit analysis to evaluate such effects; CBA is not intended to address such issues (Sugden and Williams, 1978: 199; see also Copp, 1987: 65).

Nevertheless, social justice and equity is a primary concern for public decision-makers and careful consideration always has to be given to the distributional effects of any project.

A common approach to incorporate such effects is by means of distributional weighting systems (Nash *et al.*, 1975: 127; Sugden and Williams, 1978: 201-2; Pearce and Nash, 1981: 31-3, and Mishan, 1988: 200 -1). One particular form of this approach is to express benefits and costs in terms of utility rather than money. The basic assumption is that, for every pound or dollar of gain or loss, there is a corresponding marginal utility that is related to the income of the individual (or group). In effect, all CVs are to be transformed into utility terms and "the cost-benefit criterion is met when the gains in terms of total utility exceed the losses" (Mishan, 1988: 200; see also Pearce and Nash, 1981: 31-3).

In principle, the conventional CBA, as based on the notions of Pareto improvement or potential Pareto improvement, incorporates a set of distributional weights. That is the so-called 'unitary weights': the marginal utility of, say, £1 gain or loss is the same for every individual irrespective of who gains or loses (Mishan, 1988: 201). In effect, conventional CBA assumes an optimal distribution of income (Pearce, 1983: 60). Only then, when the distribution of income is optimal, can unadjusted market prices and unitary weights be used in project appraisal, otherwise neither can be used (Pearce and Nash, 1981: 35).

Another form of the same approach is to account for distributional effects by means of 'social values' (see Sugden and Williams, 1978: 201-2 and, Thompson, 1980: 177-8). The idea is to discover the value that society attaches to the increase, or otherwise, in different individuals' welfare (An alternative is to discover the values underlying previous government decisions, eg taxation systems (Pearce, 1983: 68-9)). In this case, the net present value of a project will be the "weighted sum of ... individuals gains and losses" (Sugden and Williams, 1978: 201).

The approach of distributional weights is controversial and has been severely criticised to the extent that Mishan (1988: 211) urged economists to resist it. First, there is the difficulty of identifying that set of weights or values that will secure wide-spread acceptance (Mishan, 1988: 201). Second, any set of weights will of necessity be somewhat arbitrary (Mishan, 1988: 200). On the other hand, Pearce (1983: 62) maintained, in clear contrast to Mishan, that 'politically-determined' weights are derived by reference to some social objective function, which, even if laid by political authority, is not arbitrary. Yet, he admits that there are 'ethical' arguments against placing values explicitly on the 'deservingness', or otherwise, of different individuals or groups (Pearce, 1983: 66-7).

One particular problem of the 'utility' approach is that, in some cases, differences in marginal utility may not be measurable, if observable (Pearce, 1983: 68). The approach also involves problems of comparability and

consistency. Different weighting sets are bound to be employed in different studies. This raises questions about the comparability of individual studies. However, such arguments about comparability were considered "an attempt to impose a *specific* set of rules ..., namely the adoption of the so-called potential Pareto improvement criterion" (Nash *et al.*, 1975: 124, stress in original). The challenge of consistency remains yet to be resolved. Different results may be obtained for the same project when evaluated at different dates. The reason is the (potential) change in the weighting system (since its based on subjective judgments) (Pearce and Nash, 1981: 34).

There are two alternatives to the approach of distributional weights which attempt to avoid making value judgments (without much success). The first is "to cast the entire cost-benefit analysis in terms of conditional statements" (Nash *et al.*, 1975: 123). The difficulty, nevertheless, is that the analyst will have to choose which objective to assume. Hence, the approach returns immediately to the necessity of making value judgements. The second alternative, which is a more modest one, is to describe but not value distributional effects. Given the difficulties in formulating any set of value judgements, the analyst can just "keep track of groups and individuals who bear costs and realize gains and ... provide information (...) affecting the way that society would value benefits and costs received by different people" (Thompson, 1980: 181). "Until a satisfactory methodology for capturing distributional concerns is developed, this more limited approach seems best" (Thompson, 1980: 181).

3.3 COST-EFFECTIVENESS ANALYSIS

3.3.1 Definition

Narrowly defined, cost-effectiveness analysis (CEA) is a "comparison of alternative courses of action in terms of their costs and effectiveness in attaining some specific objective" (Quade, 1967: 1). The main purpose of the analysis is "to provide the decision-maker with a logical and [as far as possible] quantitative framework for a judgment decision" (Berg, 1967: 91). Cost-effectiveness analysis has been derived by direct analogy from the more general approach of CBA (Grosse, 1967: v and, Levin, 1975: 91). It has been described as a limited version of CBA (Shefer and Kaess, 1990: 110); a "half way house on the road to cost-benefit analysis" (Sugden and Williams, 1978: 190) and, a "truncated form of cost-benefit analysis" (Mishan, 1988: 110).

Cost-effectiveness analysis shares the same objective of CBA: to inform decision-makers in order to make more informed, rational decisions. However, CEA adopts a different approach to achieving this objective. In comparing a project's costs and benefits, CBA focuses (solely) on money as the measuring-rod for all types of effects. Cost-effectiveness analysis, on the other hand, draws guidance from only one side of CBA: either costs or benefits (Mishan, 1988: 110). That is, only one side of the analysis (costs or benefits, but mostly costs) is measured in money, whereas the other side (mostly benefits) is expressed in its 'original' units (physical, social, ...) (see Levin, 1975: 91; Sugden and Williams, 1978: 190 and, Schofield, 1987: 64). Thus, it is commonly held that CEA is most useful and appealing in problems

where programme inputs can be measured in their market valuations (prices) while outputs cannot (eg Niskanen, 1967: 18; Levin, 1975: 92-3; Thompson, 1980: 248 and, Mishan, 1988: 110).

Another fundamental difference between the two techniques is in the political constraint imposed in each of them (see Levin, 1975: 89; Sugden and Williams, 1978: 190; Schofield, 1987: 64 and, Mishan, 1988: 110-1). In CBA, benefits of the project are compared to its costs with the underlying objective of maximizing the net benefits. Although the same comparison is held in CEA, there are two other alternative constraints; either to maximize output for a given resource or budget or, to minimize costs for a given level of output.

Broadly speaking, the only difference between cost-effectiveness analysis and cost-benefit analysis is the fact that 'intangibles' will have not to be quantified in the former, unlike the latter. However, writings on CEA have apparently ignored the question of discounting, taking for granted the assumption that it is the same as in CBA.

The crucial aspect that seems to have escaped the literature (both on CBA and CEA) is how to discount 'intangibles'? The 'Green Book' recognized that it may not be possible to discount benefits and the role of discounting "may be simply to calculate a present value of the expenditure or costs associated with the option" (HM Treasury, 1991: 14). Yet, it was maintained that those

effects that can be measured in physical units may be discounted over time. The only rule is that the discount rate should be consistent with that used to discount monetary values (HM Treasury, 1991: 26). Nevertheless, it remains unclear how such effects as "better air quality" can be discounted.

3.3.2 Decision-Making Criteria

Assuming that a decision has been made about how to discount various effects (or even not to discount at all), the next question will be how to relate costs and benefits to each other? As the name of the technique suggests, the "most common finding of a cost effectiveness analysis is that net program benefits are gained at net monetary costs" (Thompson; 1980: 230, see also Levin, 1975: 92-3). This is known as the 'cost-effectiveness ratio'⁽⁸⁾. The general rule thus becomes: "The less the costs of any means to achieve a given goal, the more cost-effective is that means" (Thompson, 1980: 224).

Nonetheless, this seemingly simple rule involves several practical difficulties. First, there is no way to ascertain whether or not an alternative is 'worth it' in the sense that benefits exceed the (least) costs (Levin, 1975: 93). The two sides of the analysis are evidently incommensurable. Thus, an alternative "that appears to yield better results in terms of comparative effectiveness may have costs that outweigh its superiority in results" (Levin,

⁽⁸⁾ Less commonly used are 'the average cost per unit of effectiveness' and 'marginal cost per unit of effectiveness' (see Levin, 1975: 108-10). The two measures, however, are prone to the same practical difficulties as the cost-effectiveness ratio. Niskanen (1967: 18) argued that it is as much a mistake to use that criterion as it is to use the benefit-cost ratio in CBA.

1975: 90). Second, the rule is of little practical value when comparing between alternatives with different outputs. For instance, how can 'one acre of open-space per £10,000' be weighed against '1% reduction in noise levels for £10,000'? (Lichfield, 1964: 162-163).

A third problem, for programmes with multiple outputs, is to integrate these various outputs into the analysis (Levin, 1975: 112). There is no way of measuring 'net benefits' unless different benefits can be aggregated. One way of doing so is "to express the value of the outputs in common units" (Levin, 1975: 112). But, what common units can be found among different types of output? One alternative, as Levin (1975: 113) suggested, is to "assign *arbitrary* weights to the outputs so that they can be aggregated into an effectiveness index" (stress added). Ignoring practical difficulties, the crucial question here is: Can a 'rational' decision be based on arbitrary criteria? Finally, there is the problem of 'non-monetary' effects at both sides of the analysis; when some of the costs, as well as the benefits, cannot be valued in money.

3.4 VALUE FOR MONEY

3.4.1 Definition

Value for money (VFM) may be defined as an "attempt to maximise the ratio between a) the benefits enjoyed by the community ... and b) the ... resources needed to achieve these benefits" (Long, 1980a). More broadly defined, it is "an expression of the economy, the efficiency and the effectiveness with which all institutions, large and small, operate in the public

sector" (Butt and Palmer, 1985: 9). These three 'Es': economy, efficiency and effectiveness, are the basic inter-related elements of VFM (eg Butt and Palmer, 1985: 10-11; Bovaird *et al.*, 1988: 17; Jackson, 1988: 12-13 and, Palmer, 1993: 31). They are defined as follows (Troman, 1984: 25-6; Butt and Palmer, 1985: 10-11 and Palmer, 1993: 31):

- **Economy:** the terms and conditions under which resources are acquired. An economical organization or operation acquires resources at the appropriate quantity and quality at the lowest cost.
- **Efficiency:** the relation between outputs and inputs (resources) used to produce them. An efficient operation would produce the maximum output for given resources, or alternatively, uses minimum resources to achieve a given level of output.
- **Effectiveness:** the extent to which outputs are being achieved in relation to overall aims.

The search for value for money has been stimulated by three main factors (Butt and Palmer, 1985: 3-6):

1. **The economic and social climate:** The growing demand for better services by the public, on the one hand, and the limited public resources on the other.
2. **The trend towards greater accountability and disclosure:** The demands for more accountable public sector. In the UK, for instance, this trend has become quite evident by the early 1980s with the Central Government's growing emphasis on local authorities accountability to the public (see also Brook, 1980: 15 and, Long, 1980b: 42).
3. **The influence of legislation:** In many cases (eg the UK), government's emphasis on accountability resulted in legislation that provided auditors, public and private alike, with enough powers to review the activities of public bodies.

3.4.2 The Practice

It seems that the concept and practice of 'Value for Money' have been confined to the public sector. It seems that value for money has become to be

a 'substitute' for 'profitability measurement' in the private sector, where such techniques as cost-benefit analysis have long been practised with satisfactory results. The British Government's Green Paper on the Role of the Comptroller and Auditor General (C&AG) (1980) provides perhaps the earliest 'statutory' definition and mandate for value for money. It summarized the spectrum of activities covered by the C&AG as follows:

Value for Money Audit

- (c) an examination of economy and efficiency, to bring to light examples of wasteful, extravagant or unrewarding expenditure, failure to maximise receipts or financial arrangements detrimental to the Exchequer, and weakness leading to them.

Effectiveness Audit

- (d) an examination to assess whether programmes or projects undertaken to meet established policy goals or objectives may have met those aims. (Cmnd 7845: 7 stress in original)

However, it seems that over time both the concept and scope of 'value for money' have evolved, away from the traditions of accounting and auditing, into something different from that given by the Green Paper. As mentioned earlier, it is now widely accepted that the three basic elements of VFM are: economy, efficiency and effectiveness. A 'value for money' audit has, thus, a wider scope than conventional audits that only assess accuracy of financial statements and compliance with laws (Jackson, 1988: 12-3). In addition to judging the effectiveness of an agency/operation, the auditor of public agencies has the further function of examining possibilities of loss due to lack of economy or efficiency in the use of resources (Troman, 1984: 22).

Furthermore, and in many cases, VFM is being taken to be more than an auditing task. In the UK, for instance, many of the current reform proposals

for several public services (eg the NHS) and urban regeneration initiatives (eg the Single Regeneration Budget) appear to be underlined with the 'objective' of achieving greater value for money. VFM has apparently come to be both an objective and a way, or a framework, of planning and implementing programmes and projects to achieve it.

Such a framework, or process for achieving value for money, as Butt and Palmer (1985: 24-6) conceived it, comprises six key elements (see Fig 3.1):

1. Clearly define the strategy and objectives: otherwise there is nothing against which to measure either VFM or particularly the effectiveness of delivering policy objectives.
2. Introduce a comprehensive budgeting process: that not only controls expenditure but also questions the underlying assumptions in the main spending areas.
3. Establish a continuous 'rolling' cost-based review process: which would imply a critical questioning of all major spending areas and concentrate specifically on key resource costs.
4. Provide an effective monitoring process: to ensure that the economy, efficiency, and effectiveness of each main function is maintained or improved following the cost-based review process. An important feature of the monitoring system would be the use of key performance measures.
5. Ensure effective use of the organisation's resources.
6. Develop a robust management structure: which provides powerful leadership at the top to ensure that VFM improvements are followed through.

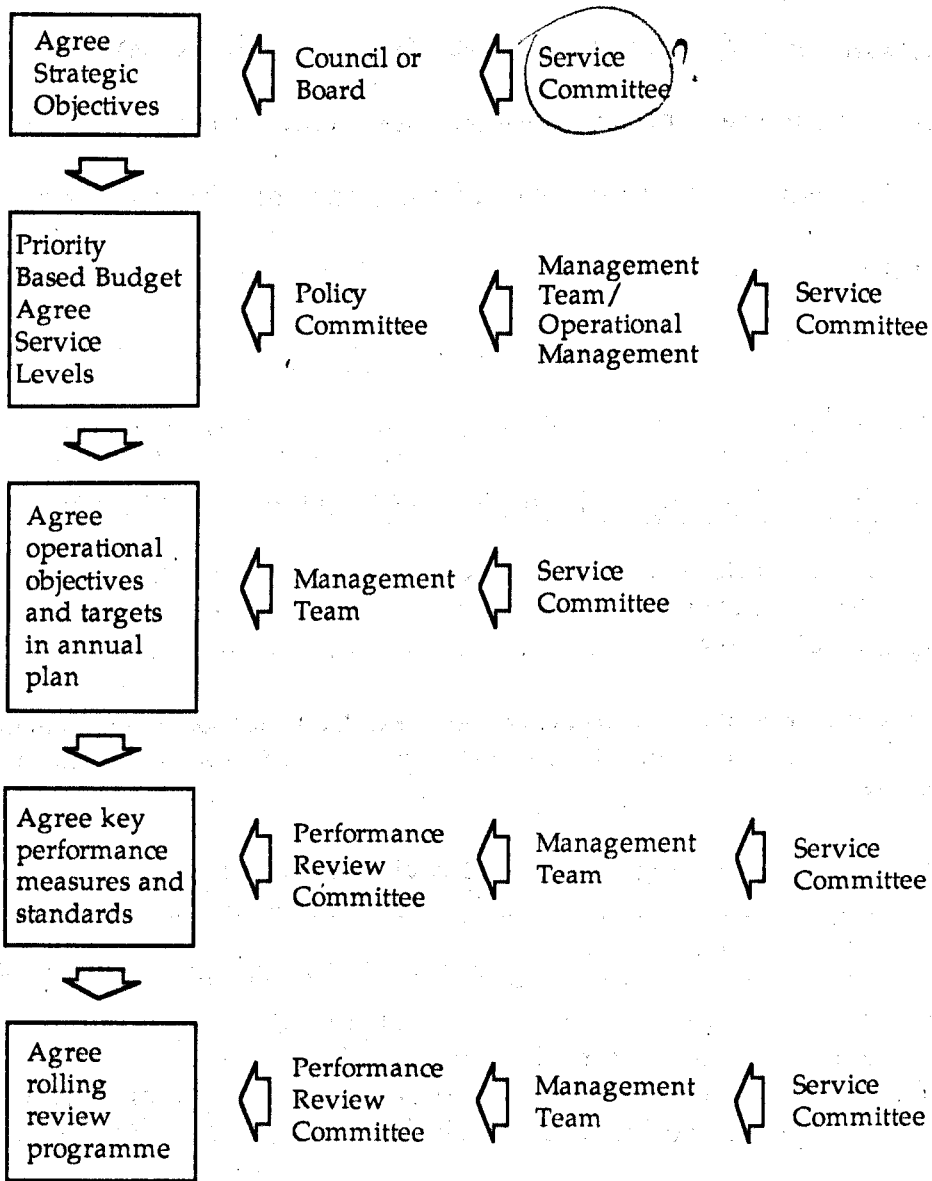


Fig 3.1: An Approach to Organizing for VFM in Local Authorities

Source: Butt and Palmer, 1985: 25.

3.5 CBA, CEA, VFM AND EX-POST EVALUATION

As stated in the introduction to this chapter, there can be no doubt on the importance of economic evaluation as a component of a comprehensive approach to the evaluation of urban policy. The question is which, if any, of the three techniques, CBA, CEA and VFM, is 'more appropriate' for such a task?

To start with CBA, and in addition to the difficulties detailed throughout section 2 of this chapter, the technique faces several other problems (though admittedly not unique to it). First, there is the counter-factual problem: what would have happened in the absence of the programme? Closely related is the issue of additionality. Which of the observed impacts are attributable solely to the project and which are the result of other trends/projects?

CBA also fails to account for displacement effects, and in broader terms equity considerations. It will remain unclear whether benefits to one group/area are at the expense of costs incurred by another. The focus on aggregate figures (NPV) conceals the distributional effects of the project: who gains the benefits and who bears the costs. To take into account displacement effects, a wider coverage (population/area) is required; to account for distributional effects, costs and benefits have to be measured for each group/area separately. It is somewhat unclear on which basis would the boundaries of the analysis be drawn as it is a problem to identify, in a

meaningful way, various groups within the society and then draw up an overall picture of the net impact of the policy. At the outset, it is a problem to define what is an 'equitable' distribution.

The concept of opportunity costs raises another particular problem within the context of public decision-making. It is very likely that different groups of society would conceive of different 'alternative uses' for the resources drawn from them through taxation. Can the analysis be broken down in a way that allows for such a consideration?

The emphasis on using money as the measuring-rod for all programme effects is one of the most criticised features of CBA. It has long been recognised that not all costs or benefits of a project can be measured in money units. Cost-effectiveness analysis (CEA) is, in a sense, an attempt to overcome this particular problem. Nevertheless, CEA faces almost the same difficulties mentioned above. In addition, to arrive at an overall effectiveness index the various effects have to be measured in a common unit, or weighed against each other. Neither task is an easy one. Perhaps more important is the fact that intangibles can not always be discounted, if ever. That makes it more difficult to achieve this effectiveness index since costs (and benefits) occurring over time can hardly be brought together through discounting.

Value for money, although apparently different, is nothing but a 'cross' between CBA and CEA, probably closer to CEA than CBA. The difference, it

appears, is nothing but the *explicit* emphasis on economic efficiency in both acquiring and allocating resources. It can be argued that such an emphasis is in fact implicit in CBA, and hence CEA, being derived from the theory of the profit maximizing firm. It is difficult to envisage a profit maximizing firm that does not strive to economically acquire and allocate its resources. Given the growing trend of public sector accountability, it is equally difficult to envisage a government that does not at least attempt to adhere to the three Es: economy, efficiency and effectiveness. The notion of value for money does not, however, offer any new break through the difficulties of both CBA and CEA; it is prone to the same difficulties.

In brief, the three techniques suffer from almost the same problems. That should not be surprising given the fact that CEA and VFM are derived directly from CBA. The only comparative advantage of CEA over CBA is the fact that intangibles have neither to be measured in money units nor in quantified manner. The advantage of VFM over CEA is the explicit emphasis on the three Es.

Therefore, with all the above mentioned problems in mind, the conclusion appears to be the use of CEA for the economic evaluation of urban policy with an explicit emphasis on the three Es. The analysis has, however, to be conducted in a multi-group fashion if equity considerations are to be taken into account. Furthermore, given both the lack of a common measurement unit among *all* policy effects (costs and benefits) and the difficulty of

discounting intangibles, the results of the analysis should be presented in their disaggregated form: the streams of benefits and costs and their incidence on different groups/areas.

CHAPTER 4: PBS, GAM & MCE

4.1 INTRODUCTION

As has already been established in the previous chapter, cost-benefit analysis (CBA) is perhaps the first, and most commonly applied, systematic method for project appraisal and evaluation. The method, however, faces a multitude of difficulties and raises several issues of concern when applied in the context of public policy-making. Notable among these is the difficulty of including both intangibles and distributional effects.

Attempts have been made, in the field of urban and regional planning, to overcome these problems. The earliest attempt, and the most direct descendant of CBA, is the Planning Balance Sheet (PBS). Lichfield first introduced the methodology in the mid-1950s and applied it in several studies in the 1960s and early 1970s. In the PBS, community groups are defined on the basis of their economic activity (those who produce services and those who consume them). Costs and benefits are defined in a similar way to CBA: the value of inputs and outputs, respectively. The PBS, however, incorporates intangibles explicitly into the analysis.

The PBS itself was severely criticized by Hill who introduced the Goals-Achievement Matrix (GAM) as a methodology to overcome the pitfalls of both CBA and PBS. Community groups, however, are not defined in advance. Benefits and costs are defined in terms of goals achievement: progress

towards objectives and retrogression from objectives, respectively. GAM makes explicit the weights the community as a whole and different groups attach to different objectives.

The GAM marked a new direction in evaluation research that has come to be known as Multi-Criteria Evaluation (MCE) methods. In these methods, however, neither community groups nor evaluation criteria are defined beforehand; both are defined in accord with the situation at hand. Moreover, unlike GAM, no attempt is made to arrive at a composite index of the relative 'worth' of each alternative.

The three methods, in other words, share the same objective: to broaden the analysis to incorporate intangibles and equity considerations. Furthermore, each method seems to have been developed as a response to the limitations of its predecessor. Yet, each method adopts a distinct approach towards the achievement of that objective.

CEA and the GAM are two components of the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984); in itself, a starting point for this research. In chapter 2, the question was raised about the applicability of other 'appraisal' methods (eg CBA, PBS ...) in ex-post evaluation. Chapter 3 dealt with this question in regard to the three related techniques of CBA, CEA and VFM. This chapter aims to answer the question with respect to the PBS, GAM and MCE methods. If GAM is applicable in ex-post evaluation (eg Hill *et al.*,

1990), what about the PBS and MCE? If they were applicable, why choose GAM and not any of them; which of the three methods is 'more suitable' for ex-post evaluation? If not, why?

These are the questions this chapter aims to answer. In order to do so, the assumptions and strengths and weaknesses of each method have to be examined in detail. The main objective is to make an informed choice whether any of the three methods could be a component of a comprehensive framework for ex-post evaluation of urban policy; the main objective of this research.

Sections 2, 3 and 4 of this chapter tackle the PBS, GAM and MCE, respectively. The structure of the three sections is quite similar. Following a brief introduction, the procedures and assumptions of the method are reviewed. Then, an account is given of the strengths and weaknesses of the method. By highlighting the advantages and disadvantages of each method, the aim is to inform the choice on which method, if any, should be applied in ex-post evaluation.

This choice is the central focus of section 5. This section starts with a comparison of the three methods relative to each other. It is noted that writings on the methodology of the PBS and GAM ceased around the mid-1970s. MCE methods, on the other hand, were the subject of some recent publications (eg Nijkamp *et al.*, 1990a). Moreover, they represent a prominent

feature of public policy-making in the Netherlands. Yet, GAM is the only one of the three methods that has been applied in ex-post evaluation, albeit once (eg Hill, 1986).

In principle, the three methods are applicable in ex-post evaluation. The question then becomes: which of the three methods to choose? It is argued that MCE methods represent the 'most appropriate' choice. They provide a considerable degree of (much needed) freedom in defining community groups and impact categories. It is also argued that a composite index should be avoided for it conceals a great deal of information and, thus, violates the primary function of evaluation. The section finally concludes with a remark on the treatment of the counterfactual problem.

4.2 THE PLANNING BALANCE SHEET

Cost-benefit analysis (CBA) is a long-established appraisal technique that is well founded in economic theories. When applied in public decision-making contexts, however, the technique faces a host of difficulties and raises several crucial concerns (see chapter 3). Notable among these are (1) the difficulty of measuring many of the policy effects in monetary units and, (2) the considerable neglect of distributional effects.

The Planning Balance Sheet (PBS) is a methodology of (ex-ante) cost-benefit analysis for city and regional planning (eg Lichfield, 1960, 1966b and 1968). It is a particular application of the so-called 'social' cost-benefit analysis

to evaluation which attempts to overcome the pitfalls of conventional CBA (Lichfield *et al.*, 1975: 60). The PBS was first introduced in the mid-1950s (Lichfield, 1956) and was applied in several case studies throughout the 1960s in Britain.

Since it was first conceived, hardly any modifications have been introduced to the PBS, though it has been argued that the method was modified and improved in the 1970s and 1980s (Lichfield, 1990: 87). The only "change" that can be traced is in the 'label' of the method. Lichfield (1985: 59) maintained that within the PBS, it is "the whole array of impacts on the whole community which are under consideration" (stress in original). Given this, and the argument that PBS incorporates impact *assessment* as much as evaluation, the method was then labelled CIE: Community Impact Evaluation (Lichfield, 1985: 59).

4.2.1 The Methodology

The analysis is always presented in a tabulated format (known as Table A; see Figures 4.1 and 4.2). The first step of the analysis is to enumerate as sectors the various 'producers/operators' and 'consumers'. Producers or operators are individuals or groups who take part in creating and operating the project whereas consumers are individuals or groups who consume the services produced. The number of individuals in each sector should be identified since both total costs and benefits and averages (cost and benefit per individual) are of interest.

Fig 4.1: An Illustration of The Standard Format of the Planning Balance Sheet

Item No.	Sector	Number		Instrumental Objectives	Scheme A				Scheme B				Balance	Net Advantage To
		A	B		Benefit		Cost		Benefit		Cost			
					C	A	C	A	C	A	C	A		
Producers														
1.0.	D.C. as Developer				272,000		287,000		319,000		298,000		21,000 > -15,000	B
1.0.	Reduction	n	$n+$		£	£	£	£	21,000		£	£	≡	N
3.0.	Displaced													
3.1.	Reduction													
3.3.	Not Displaced				M			0					0 < M	A
3.3.1.														
3.3.2.														
3.3.	Reduction				m_1, m_2, m_3				m_2+				$m_2+ < m_1, m_2, m_3$	A
5.0.														
Consumers														
2.0.	New Users													
2.2.	Occupiers of New Private Buildings	Sq. Ft. 91,000	Sq. ft. 125,000				m_6		i_1+				$i_1+ > i_1, i_2 > m_6, i_1+ > m_7$	B
2.2.1.	Shops				i_1									
2.2.2.														
2.2.	Reduction				i_1, i_2, i_3		m_6, m_7, m_{10}		$i_1+ i_2+ i_3+$				m_2, m_7, m_{11}	B
2.4.	Occupiers of New Public Buildings													
2.4.1.														
2.4.	Reduction	n	n		i_7, i_8		m_{14}		$i_7+ i_8+$				m_{14}	N/C
2.6.														

Source: modified from Lichfield, 1966c: 134-138.

As far as it is possible, each producer or operator is paired with the appropriate consumer. Each linked pair is considered to be engaged in a 'transaction' where producers produce services 'for sale' to the consumers. Transactions are not confined to market goods; they also include 'indirect' transactions such as 'traffic noise' imparted on residents near a newly constructed road. The table thus represents a set of 'social accounts' summarising all transactions in the project. No transaction is to be omitted no matter how 'intangible' the service produced might be.

Costs and benefits are defined in a similar fashion to that of CBA. Costs are the value of resources used to produce or operate the project (the inputs) and benefits are the value of goods or services produced (the outputs). Costs and benefits that enter the analysis might be direct or indirect, or real or transfer. It follows that the costs and benefits of producers/operators and consumers are not linked in the same way in all pairs.

Fig 4.2: An Abstract Form of the PBS

Producer	Plan A		Plan B		Consumer	Plan A		Plan B	
	Benefit	Cost	Benefit	Cost		Benefit	Cost	Benefit	Cost
1					2				
3					4				
5					6				
7					8				

Source: McAllister, 1980: 150.

The common denominator for those goods and services that are exchanged at the market is money. However, many of the consumers' costs and benefits, unlike those of producers, are intangibles: utilities or disutilities to the consumers. These are measured by means of 'instrumental objectives': "to specify the hoped-for "dimensions of utility or welfare," enumerating several categories as necessary for any particular good or service" (Lichfield, 1966c: 140). Subjective judgment is then applied to forecast the extent to which the plan will achieve these objectives. In the first place, these 'instrumental objectives' are expressions of the goals of the plan or the problems it aimed to tackle.

The analysis is primarily a 'with and without' comparison: what would happen had the project been implemented compared to what would happen without implementation. The next step, thus, is to forecast the difference in all costs and benefits that would accrue to each producer/operator or consumer if the project were implemented. Comparisons can be made in terms either of absolutes or differences. Although it is the differences which are of interest, their significance cannot be gauged accurately without absolutes.

Allowance must then be made for time through discounting - present discounting to the time when the analysis is made. Although the impracticability of discounting intangibles was recognized (Lichfield, 1966c: 141), no approach was suggested to overcome that difficulty.

One column in the table (Remarks/Balance) shows the algebraic reduction of entries for lines containing costs and benefits which are not in figures and therefore cannot be arithmetically reduced. This will indicate, for instance, which costs and benefits are greater or lesser ($<$ or $>$), where they are equal (\equiv), where difference is unknown (N/C) and where probabilities can be seen (Prob.). The next column shows the conclusion by indicating which scheme has the net advantage for that particular entry, and where there is uncertainty (N/C). The important rule to follow in all this is that one scheme is compared with another so that differences always show in favour of, or against, a particular scheme. Consequently, there is not much hesitation in making interpersonal comparisons.

The implications of the table, however, will remain complex if it contains many unmeasured entries. To simplify, a summation table is prepared (known as Table B; see Fig 4.3), to which the reduced entries are transferred. This table itemises all producers/operators and consumers as in table A but with a regrouping of items in order to simplify conclusions. Not only are the reduced entries of table A shown but also the differences between them. The differences are then summed algebraically, showing the 'net' situation. This net column in table B represents the ultimate reduction of social accounts. It facilitates comparisons of projects in terms of their costs and benefits and the incidence of these costs and benefits.

Fig 4.3: An Illustration of the Summary Table of PBS

Item No.	Sector Reduction	Scheme B Minus Scheme A						Net Advantage To
		Number		Benefit	Cost	Net		
		A	B					
Producers/Operators								
1.0.	D.C. as Developer			£21,000	£15,000	£36,000		B
3.0.	Current Landowners							
3.3.1.	Real Changes			m_2+				A
5.0.	D.C. as L.A.			m_5+ m_5	m_4- m_4	m_5+m_4- m_5m_4		B
Consumers								
2.0.	New Users							
2.2.	Occupiers of New Private Buildings			$i_1+i_2+i_3+i_1i_2i_3$	$m_7m_9m_{11}m_6m_8m_{10}$	$i_1+i_2+i_3+m_7m_9m_{11}$ $i_1i_2i_3m_6m_8m_{10}$		B
2.6.	Motor Vehicle Users							
2.8.	Shopping Public	n	n		$i_5-i_{14}-i_5i_{14}$	$i_5-i_{14}-i_5i_{14}$		B
2.1.0.	Public at Large	n	n	$i_{15}-i_{16}+i_{15}i_{16}$		$i_{15}-i_{16}+i_{15}i_{16}$		N/C
4.0.	Current Occupiers							
4.2.	Displaces	n	$n+$		$p_1+p_2+p_3+$	$p_1+p_2+p_3+$		A
4.6.	Ratepayers			m_4-m_4	m_5+m_5	$m_4-m_5+m_4m_5$		A

Source: modified from Lichfield, 1966c: 142-143.

The summation in table B, supported by table A, provides the analysis on the basis of which the decision-makers can act. In some cases, it may point clearly to conclusions; in others, it may yet leave considerable room for value judgment. The decision criteria are quite similar to those common in typical CBA: either a positive net benefit or, if possible, a benefit-cost ratio that is greater than 1.0. However, PBS in itself "cannot, and does not, aim to provide a conclusion in terms of a rate of return or net profit measured by money values" as the case with conventional CBA (Lichfield, 1969: 128).

Fig 4.4: An Abstract Form of the Summation Table of PBS

Producer	Plan A minus Plan B			Net Benefit to	Consumer	Plan A minus Plan B			Net Benefit to
	Benefit	Cost	Net			Benefit	Cost	Net	
1					2				
3					4				
5					6				
7					8				

Source: McAllister, 1980: 150.

4.2.2 The Pros and Cons of PBS

Although it is a direct descendant of cost-benefit analysis, the planning balance sheet attempts to overcome the pitfalls of CBA especially when applied in public decision-making (see chapter 3). However, the PBS itself has its shortfalls. In many cases, criticism of the PBS has taken the form of a

comparison with CBA. The following is an account of the major critiques mounted against PBS together with the counter arguments. By highlighting both the advantages and shortcomings of this method, as with others, this sub-section aims to inform any future choice of the method(s) to be employed in ex-post evaluation.

Perhaps surprisingly, considerable criticism centred around the inclusion of unquantifiables into the analysis, in particular, and the presentation of information in general. Unquantified effects are, of necessity, rated in a subjective manner (Alexander, 1978: 65). The way they are recorded still leaves them at something of a disadvantage (McAllister, 1980: 155). Alphabetical symbols of unquantified impacts require reference to the written descriptions provided elsewhere. This, in turn, requires time which may not be available to the decision-makers who may, thus, base the decision on quantified impacts only. It was also recognized that such explicit statements of intangibles may make decisions more difficult as they tend to force decision-takers to make explicit their subjective judgments (Lichfield *et al.*, 1975: 60). Moreover, the recording of impact information in the form of transactions may fail to reveal the most important equity effects. That is, the adverse impacts on disadvantaged groups which are, usually, not a unified producer or consumer group (McAllister, 1980: 155).

Nevertheless, it is argued that the statement of intangibles in the same table as quantified impacts means that decision-takers will have a convenient

summary of all the information produced, which is more likely to give these intangibles due consideration (see McAllister, 1980: 148, 155, 158, and Lichfield *et al.*, 1975: 60-1).

In more general terms, the PBS was charged for the "somewhat confusing array" of monetary and quantitative measures and qualitative judgements "superimposed on each other"; "one is bombarded with information and invited to judge the relative merits of alternative schemes" (Peters, 1973: 45). Moreover, the analysis does not provide clear cut conclusions; in cases "there was no obvious balance of advantages ... and consequently there was much room for value judgements" (Peters, 1973: 45). Lichfield (1990: 87) argued that considerable effort was always made in search for meaning of the analysis in terms of the issues raised and that the complexity of the analysis arose from the diverse impacts of alternatives on different sectors. He also maintained that the different experiences of conclusions were very much a matter of chance. The nature of alternatives, for instance, is an important factor. "If they themselves are not clear cut, how can the choice between them be so?" (Lichfield, 1990: 86).

In comparison to CBA, the PBS has two further disadvantages (McAllister, 1980: 155-8). Firstly, selecting the impact categories in the form of 'transactions' in contrast to the analyst's free hand, in CBA, to choose the important categories in accord with the context of the analysis. If financial transactions dominate the selection, which is most likely, certain types of

environmental, social and political impacts may be neglected. In addition, breaking down impacts by sectors may fail to capture important aggregate effects on the whole society. For instance, emissions of air pollutants may be small from each sector and thus treated as unimportant. Even though the aggregate air pollution impact may indeed be substantial, this is not reported in the PBS.

In recognition of the latter problem, Lichfield (1966a: 28) added the 'Public at large' as one consumer group. However, this was considered a 'definitional departure' – creating a new category on a different basis than the rest of community groups – which makes the impact categories of PBS appear arbitrary and unimportant (McAllister, 1980: 159) .

The second disadvantage of PBS relative to CBA is that the major value judgments allowed into the analysis violate the criterion of objectivity. Indeed, there is no reason why planners should regard their judgments as an indicator of their clients values. CBA, on the other hand, does not, formally at least, permit the insertion of evaluators' values into the analysis.

Forecasting is another aspect of PBS, as indeed any ex-ante evaluation method, that raises concern. Predicting the future with any reliable degree of accuracy is almost impossible, particularly in town planning where projects are likely to have long lifespans. The further into the future an impact occurs, the more difficult it becomes to forecast it or its magnitude.

As with CBA, discounting is another problematic issue in PBS. Although there appears to be no other alternative but to treat the different streams of costs and benefits, PBS by adopting the same approach lends itself to the same difficulties and criticism of discounting (see chapter 3). More importantly, the element of time is not taken fully into account in the analysis (Alexander, 1978: 51). There is no way to discount the many intangibles a standard PBS contains. In the case of the Peterborough study (Lichfield, 1969), for instance, only twelve of the sixty-eight items included in the Sheet were measured in monetary units.

Nevertheless, Lichfield (1970: 154) - having reviewed 24 appraisal methodologies then available, including the GAM - concluded that the PBS is the most comprehensive of them all. Moreover, he later argued that the method is no different from multi-criteria evaluation methods (as advocated by Voogd, 1983a) (see Lichfield, 1990).

4.3 GOALS-ACHIEVEMENT MATRIX

The goals-achievement matrix (GAM) was first introduced by Hill as a methodology for the evaluation of transportation plans (eg Shefer and Tsubari, 1990: 32). The major thrust behind the methodology was, apparently, the contention that neither CBA nor PBS was a satisfactory method for the evaluation of alternative courses of action (Hill, 1968). CBA was severely criticized for its focus on the single objective of economic efficiency which may lead to overlooking other objectives. PBS was charged for its failure to

recognise that goals and benefits can only be compared if they related to a common objective. GAM is thus an attempt to broaden the scope of evaluation and relate its criteria to the objectives of the plan being evaluated.

4.3.1 The Methodology

The basic procedure of goals-achievement matrix (GAM) is quite simple (Hill, 1968). The set of goals is identified and the relative weight the community as a whole attaches to each goal is established. These goals are to be defined operationally, ie objectives, rather than in abstract form. For each objective, the consequences of each alternative are then determined. The incidence of consequences is also measured and the relative weight each community group attaches to each objective is established. The final product of the analysis for every alternative is then presented in a table (see Fig 4.5).

Fig 4.5: Goals-Achievement Matrix

Goal description	α			β			τ			δ		
Relative weight	2			3			5			4		
Incidence	Relative Weight	Costs	Ben.	Relative weight	Costs	Ben.	Relative weight	Costs	Ben.	Relative weight	Costs	Ben.
group a	1	A	D	5	E	-	1		N	1	Q	R
group b	3	H		4	-	R	2		-	2	S	T
group c	1	L	J	3	-	S	3	M	-	1	V	W
group d	2	-		2	-	-	4		-	2	-	-
group e	1	-	K	1	T	U	5		P	1	-	-
		Σ	Σ					Σ	Σ			

Source: Hill, 1990: 14.

"A *consequence* is a change in a given situation caused by a policy" (Hill, 1968: 22, stress in original). A 'positively valued' change is a benefit and a 'negatively valued' one is a cost. That is, consequences are the costs and benefits of a plan. However, both costs and benefits are defined in a different manner from CBA and PBS (Hill, 1968: 23). In GAM, both costs and benefits are defined in terms of goals achievement; progress towards objectives is a benefit whereas retrogression from objectives is a cost. This contrasts with CBA and PBS where costs are defined in terms of the value of resources used to produce or operate the project (inputs) and benefits are the value of goods or services produced (outputs).

The weighting of objectives, activities and groups is the "key to plan evaluation by means of goals-achievement analysis" (Hill, 1968: 27). The determination of both community objectives and their relative weights "is no easy task and requires considerable research" (Hill, 1968: 27). Methods differ from one community to another, from one sector to the next and from one governmental context to the other. Direct methods include consultations with elected officials and community and interest groups, sampling and public hearings. Indirect approaches include inference of relative values underlying either patterns of groups' behaviour or previous allocations of public resources and investments.

There are several approaches to comparing the effectiveness of different alternatives in achieving the set of objectives. "The simplest strategy is to

present the decisionmaker with the entire goals-achievement account without attempting to synthesize the extent of goals achievement" (Hill, 1968: 25). That is, the decision-maker is presented with as many tables as the alternatives under consideration. The onus of trading-off between the alternatives would then be on the decision-maker.

Another approach is the weighted index of goals-achievement (Hill, 1968: 25-26). One alternative is to treat all objectives as if they were all measured on an ordinal scale. That is, each plan is evaluated to determine whether it increases, decreases or leaves goals-achievement at about the same level. Arbitrary values are then assigned and summed up to provide an overall index of each alternative. Another alternative is to introduce the weights of individual objectives and their incidence into the calculation to obtain a weighted index of goals-achievement.

A third approach, which "is difficult to put into operation", is the goals-achievement transformation function (Hill, 1968: 26). For quantitative objectives, quantities on one scale are equated with quantities on another scale and then the relative values on the first scale are determined on the second. "The problem is considerably complicated if there are more than two quantitative objectives, all measured in different units"; it is also "much more tenuous" to transform qualitative objectives.

4.3.2 The Pros and Cons of GAM

GAM was a direct response to the limitations of both CBA and PBS (see Hill, 1968). When compared with CBA, the goals-achievement matrix has several advantages (McAllister, 1980: 168-169). Firstly, GAM establishes a formal procedure for assessing equity effects, though this will raise the cost of conducting the evaluation. Secondly, it includes all quantified impacts in its composite index, not only monetized impacts. Thirdly, GAM emphasizes the significance of organising impacts in categories which relate directly to community goals. Similarly, GAM has its advantages over PBS (McAllister, 1980: 169):

1. In GAM the evaluator has a freer hand in selecting the community groups to be used in assessing equity effects. ...
2. The evaluator also has a freer hand in selecting the impact categories.

However, GAM itself has its limitations and faces some severe criticism. Firstly, GAM seems to exaggerate the importance of goals statements to the point that costs and benefits have meaning only in relation to well-defined objectives (see Hill, 1968: 21). This apparently overlooks the fact that the "most common problem encountered by evaluators is 'fuzziness' of goals and objectives" (Nietied and Schevz, 1988: 398). Perhaps more importantly, "formally prepared goals statements should not be used as a rigid set of criteria for evaluation" (McAllister, 1980: 165-166). Throughout the planning process, goals can, and perhaps even should, be reformulated and augmented in the light of new knowledge. Furthermore, formally stated goals may not

cover all community objectives. Should the analyst ignore such community objectives simply because they are not formally stated?

Secondly, the conclusion that GAM "is not very useful if weights cannot be objectively determined or assumed" (Hill, 1968: 27) "seems to place GAM on very shaky grounds"; valid measures for establishing value weights are not likely to be developed (McAllister, 1980: 168). Indeed, the major criticism of GAM centres around the numerical weights attached to the various objectives and community groups (Shefer and Tsubari, 1990: 33). Weighting "is a process which is not only unlikely, but theoretically impossible, unless some kind of process of registering choice and agreement to the emendation of choice is available. And how might interest groups agree to a weighting which places their own weight lower than others in a ranking?" (Chadwick, 1971: 269 as quoted in Shefer and Tsubari, 1990: 33). This criticism should not, however, be seen as unique to GAM. It applies, with the same strength, to any method that attempts explicitly to incorporate equity effects.

Thirdly, among the criticism mounted against GAM was "the reduction *ad absurdum* of the quality of a given alternative into a single aggregated number" (Shefer and Tsubari, 1990: 33, stress in original). Also related is the complexity of the arithmetical procedures which makes the results of the analysis difficult to digest (McAllister, 1980: 167 and Shefer and Tsubari, 1990: 33). However, a distinction has to be made between the initial conception of the method and later developments. In the first instance, Hill (1968: 27)

recognised the fact that the method does not lend itself to a single number outcome. Approaches to arrive at a composite index were alternatives to a simpler strategy (see above) and their limitations were acknowledged. The criticism applies, nonetheless, to later developments and applications of the method. Attempts were made to incorporate GAM into mathematical programming where complex procedures were employed to obtain a composite index of each alternative (see, for instance, Hill and Shechter, 1971 and 1978 and Hill and Werczberger, 1978).

Fourthly, a major disadvantage of GAM "is that interaction and interdependence between objectives is not registered" (Hill, 1968: 28). Until such interaction is accounted for, the method is "recommended only for the evaluation of plans in a single sector" (p. 28). Fifthly, GAM fails to define what are community goals - as distinct from groups' objectives - or procedures for preparing them (McAllister, 1980: 166). Sixthly, GAM "calls for an extremely complex, time-consuming, and expensive task" (Hill, 1968: 27).

Finally, a major shortfall of GAM that appears to have gone unnoticed is its partiality. The primary emphasis of CBA on the single objective of economic efficiency was criticised to the effect that GAM fails to provide any knowledge on such a crucial concern. This follows, logically, from the manner in which costs and benefits are defined in GAM. These are defined in terms of goals-achievement; there is no account of the value of either the inputs or the outputs of alternatives. Economic efficiency is no doubt a crucial concern

for decision-makers and many others of the stakeholders and, thus, has to be accounted for in evaluation. A strong support to this conclusion comes from both the proposal for 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) and the only documented application of GAM in ex-post evaluation (eg Hill *et al.*, 1990). In both cases, the evaluation approach incorporated a distinct element of economic evaluation - in the form of cost-effectiveness analysis - in addition to the other components, including GAM.

4.4 MULTI-CRITERIA EVALUATION METHODS

GAM is one of the "better-known examples of the multi-criteria evaluation technique in urban and regional planning" (Shefer and Tsubari, 1990: 32). Yet, the names mostly associated with the approach of multi-criteria evaluation (MCE) are those of P. Nijkamp, P. Rietveld and H. Voogd. Indeed, as will be explained in the next section, the Dutch context provides a different experience in the application of MCE.

The crucial distinction between GAM and what has come to be known as MCE relates to the definition of evaluation criteria. In GAM, as mentioned above, alternatives are evaluated in terms of goals achievement, and costs and benefits are defined accordingly. MCE methods, on the other hand, adopt a different approach; evaluation criteria are not prescribed beforehand. The analyst has a free hand to define the criteria in accord with the context at hand. These may then include the costs and benefits as defined in conventional CBA and PBS (value of inputs and outputs), as defined in GAM

(progress towards goals), or any other form as deemed appropriate or necessary. This particular distinction is the reason why MCE methods are singled out and distinguished from GAM.

4.4.1 The Methodology

The basic principle of multi-criteria evaluation methods is fairly simple (Voogd, 1983a: 28-30). MCE methods start from a number of explicitly formulated criteria which are not measured in one single unit. Then, a matrix is constructed which reflects the characteristics of the given set of choice possibilities by means of these criteria. This is known as the evaluation matrix (see Fig 4.6). Choice-possibilities can be the alternatives of a plan, groups of the community, regions and so forth.

CRITERIA	CHOICE-POSSIBILITIES				
	A	B	C	D	E
1	Criterion Scores				
2					
3					
4					
.					
.					

Fig 4.6: The Evaluation Matrix

Source: modified from Voogd, 1983a: 29.

In some cases the evaluation matrix provides a clear picture of the general differences between the alternatives and no further analysis is needed. In others, however, a straightforward interpretation of the matrix is almost impossible due to the conflicting criteria. Information is thus needed about the

relative importance attached to the criteria. This is summarized in a matrix; the so-called priority matrix (see Fig 4.7). These two matrices are then linked (by means of mathematical procedures) to produce the appraisal matrix (see Fig 4.8) which provides an indication of the general quality of the alternatives.

VIEWS	CRITERIA				
	1	2	3	4	5
I	Priorities				
II					
III					
IV					
.					
.					

Fig 4.7: The Priority Matrix

Source: modified from Voogd, 1983a: 30.

VIEWS	CHOICE-POSSIBILITIES				
	A	B	C	D	E
I	Appraisals				
II					
III					
IV					
.					
.					

Fig 4.8: The Appraisal Matrix

Source: modified from Voogd, 1983a: 31.

The fact that evaluation criteria are measured on different scales and in different units raises two crucial issues that have to be resolved. Firstly, quantitative criterion scores are mutually incompatible. Consequently, there is a need to transform these scores into one common measurement unit. This transformation is known as *standardization* and can be achieved by means of

one of several formulae (see Voogd, 1983a: 77-80, 86-91; Massam: 1988: 32-36 and Nijkamp *et al.*, 1990a: 42-43, 67-68).

The second problematic issue is qualitative data. There are several approaches to treating qualitative data (Voogd, 1983a: 45-47). The simplest approach is to transform qualitative data into rankings and use these as if they were metric quantities (eg Hill, 1968 see above). Another is the lexicographic method where the best alternative is that with the best score for the most important criterion, regardless of other criterion scores (eg Holmes, 1972; Voogd, 1983a and Nijkamp *et al.*, 1990a). A third approach makes use of a permutation method which appraises all possible (final) rankings of the alternatives in order to find the best 'final' ranking (eg Nijkamp, 1979 and Voogd, 1983a). A fourth approach is the geometric scaling models which provide not only a final ranking of the alternatives but also additional information about both the quality of the outcomes and the characteristics of the judgment problem at hand (eg Nijkamp, 1979; Nijkamp and Voogd, 1980 and Voogd, 1983a).

The drive behind standardization and the treatment of qualitative data is the attempt to simplify the evaluation matrix in order that it can be linked to the priority matrix (which includes numerical weights) to arrive at the appraisal matrix. This linkage can be done by means of one of a variety of mathematical procedures, eg weighted summation, concordance analysis and regime analysis (eg Nijkamp, 1979; Voogd, 1983a and Massam, 1988).

The important point to be always borne in mind is that every standardization method is based on assumptions "which cannot often be properly explained nor justified" (Voogd, 1983a: 40). This applies, with the same strength, both to methods for treating qualitative data and for compiling the appraisal matrix. The examples given of applying MCE methods (eg Voogd, 1983a and Nijkamp *et al.*, 1990a) have thus employed several methods each to elicit the effect different methods have on the outcome of the analysis. Yet, in none of these examples has an attempt been made to derive a composite index for each alternative.

4.4.2 The Pros and Cons of MCE Methods

MCE methods, as such, have their advantages. The most detailed account of these was that of Voogd (1983a: 32-34) who saw the methods as means to:

- arrive at a surveyable classification of factual information;
- get a better insight into the various value judgments;
- incorporate difference in interests and/or political views;
- give more substance to the notion of openness in the planning process;
- arrive at a reduction of the available information;
- arrive at substantially better considered decisions;
- arrive at a more controllable position of the expert in the planning process;
- account for or justify policy decisions, and
- structure research contributions in the planning process.

Yet, Voogd (1983a: 34-35) recognized that the use of MCE methods may be associated with some disadvantages:

- they may lead to an eventually premature disclosure of policy intentions. The disclosure of certain 'weaknesses' may be detrimental to the negotiations between stakeholders;
- they can be technically too complex to be understandable by non-experts, and
- they offer the possibility of manipulating the political opinion making.

MCE methods also share some of the criticism levelled against GAM (see pp. 99-102). The methods are "not very useful if weights cannot be objectively determined or assumed" (Hill, 1968: 27). This issue of establishing and using explicit weights in MCE was indeed severely criticised (Buckley, 1988). However, and as mentioned before, a distinction has to be made between a method that attempts to incorporate societal and political values within the analysis and the difficulties associated with measuring these values. Voogd (1983a: 95) cautioned against these difficulties and the sensitivity of the analysis and its outcome to the hypotheses made at arriving at these weights. The methods also call for a (very) complex, time-consuming and costly task.

4.5 PBS, GAM, MCE AND EX-POST EVALUATION

The three techniques of PBS, GAM and MCE share the same objective. They all aim to broaden the scope of evaluation to include the various types of impact as well as equity considerations. Yet, the approach each adopts to achieve this goal is quite distinct from the others. The PBS is highly influenced by the economic nature of its predecessor - CBA. This is evident in the definition of community groups (on the basis of their economic activity) and costs and benefits (value of inputs and outputs). The PBS, however, incorporates intangibles explicitly and does not aim to provide a single figure of net benefit or rate of return.

GAM departs entirely from this approach and starts from a severe criticism of both CBA and PBS. Community groups are not defined in advance. Costs and benefits are measured in terms of goals achievement rather than (economic) values of inputs and outputs. This particular feature of GAM, however, may be its very undoing. Even if goals were defined operationally, it is fairly common that objectives are not quantified. Although the direction of change can then be assessed, there is no means of judging its significance. Moreover, the crucial concern of economic efficiency or effectiveness is simply ignored, unless it was one of the explicitly stated goals. In more general terms, GAM is in danger of confining the analysis to formally stated goals and overlooking others, particularly those of community groups.

MCE methods, on the other hand, offer the analyst considerably more freedom in both respects. The evaluator, in dialogue with decision-makers and stakeholders, can define community groups and impact categories (evaluation criteria) in accord with the characteristics of the situation at hand. None is prescribed beforehand. Unlike GAM, these methods do not aim at a composite index of each alternative, though considerable effort is expended in simplifying the results.

Time is a crucial dimension that has to be accounted for, in ex-post as well as ex-ante evaluation. Costs and benefits that occur at different times do not have the same weight or value; they have to be brought together to a single point in time through discounting (see chapter 3). The three methods fail to address discounting adequately. Lichfield (eg 1966c: 144) recognised the difficulty of discounting intangibles but left the issue unresolved. Hill (1968: 23) contended that the issue had already received considerable attention and, thus, did not discuss it. MCE methods have apparently overlooked the issue altogether.

GAM and MCE share the problematic feature of making preferences explicit in the analysis. The measurement of these preferences, in the first place, is fraught with difficulties. Moreover, making preferences explicit may simply increase tensions between interested groups. However, they can be a significant aid to (democratic, objective) decision-makers to base their decisions on community's needs and aspirations. The PBS avoids this

problem. The task of trading-off between the different alternatives (and implicitly, different groups) is left for the decision-makers. Only in one case study (Edgware, see Lichfield and Chapman, 1968) were community preferences made explicit in the analysis.

Writings on the methodology, and applications, of PBS and GAM in ex-ante appraisal appear to have ceased around the mid-1970s. MCE methods, on the other hand, seem to have enjoyed a strikingly different fortune, at least in the Netherlands. MCE methods have been the subject of some recent publications (eg Nijkamp *et al.*, 1990a). Moreover, there are strong indications of a different governmental attitude, in the Netherlands, to evaluation research (Field Trip, 1995). In meetings with senior officers at the national and provincial levels, a clear commitment to explicit policy appraisal and evaluation was expressed. In 1991, the Lower House of the Dutch Parliament published guidelines on policy evaluation, in the civil service, making it almost mandatory. The Dutch Ministry of Finance, in collaboration with Free University, Amsterdam, have recently developed a computer package for policy appraisal using MCE. The approach of MCE has been, and is being, applied to real-world case studies. Why is the situation so? There is, unfortunately, no answer.

The experience of ex-post evaluation is quite different. No record could be traced for applications of either the PBS or MCE in an ex-post context. Of the three methods, only GAM has been applied in an ex-post real-world study.

That was the evaluation of Israel's Project Renewal (eg Hill *et al.*, 1990). In this case, the first difficulty encountered was the fact that there was no formal statement of the Project's goals. The research team had to go through a long process of meetings with policy-makers and residents in order to compose a set of goals to be evaluated. The impact of the Project was disentangled from the effects of other trends on the basis of before-after and time-series analyses (aided by monitoring and implementation analysis in line with the 'Integrated Evaluation' approach to the task; eg Hill *et al.*(1990), see chapter 2).

This leads to the question of ex-post applicability. Are the three methods applicable in ex-post evaluation as in ex-ante appraisal? For GAM, the answer is evidently 'Yes' and, in principle, there is no reason why both the PBS and MCE cannot be applied in ex-post evaluation as well. The point to bear in mind is that the difficulties each encounters in ex-ante appraisal are all likely to surface in ex-post evaluation.

The question then becomes: which of the three methods is most suitable as a component of a comprehensive framework for ex-post evaluation of urban policy? The answer is MCE. For one reason, the approach has a particular advantage that is believed to outweigh both its disadvantages and the advantages of both the PBS and GAM. That is, the considerable freedom it offers in defining both community groups and impact categories. This provides flexibility which is much needed for circumstances differ, sometimes dramatically, from one case to the other. Another related reason is that MCE

avoids the inherent danger of GAM of confining the analysis to formally stated goals and overlooking both community objectives and unintended impacts.

However, if ex-post evaluation was to assess the achievement of policy goals, GAM may seem a better choice, if only for its definition of costs and benefits. Nevertheless, it is not uncommon that goals are stated in a very vague manner (if at all). In such cases, 'proxies' have to be developed which reflect, as closely as feasible, the intentions of the policy (which was the case in the ex-post application; see Hill *et al.*, 1990). For instance, increasing economic welfare may be defined in terms of jobs created, increase in disposable income, decrease in unemployment and so forth. The impact of the policy is then measured in terms of these 'indicators' and judgment has then to be exercised as to how these contribute to the higher-level goals. In other words, the achievement of goals has to be measured *indirectly* and the analysis is no longer a GAM; rather it has become a MCE.

Should a composite index be established for the policy being evaluated? It could be argued that such an index would give an indication of the overall achievement of the policy. However, the fact that a wide range of stakeholders are involved in the process means that there is no such single indicator. Each interested group will judge the achievement, or otherwise, of the policy from a different perspective. This is due, clearly, to the different interests and priorities each holds. The composite index conceals a great deal

of information and, consequently, violates the primary function of evaluation: to inform stakeholders. A composite index also involves the almost impossible task of weighting different groups explicitly. Furthermore, and related, if distributional effects were to be taken into account, policy impact on different groups will have to be reported separately. That is not the case with a composite index.

It is thus believed that the results should be presented to the decision-makers and stakeholders in their disaggregated form. If it proved feasible, groups' preferences are to be measured and reported as well. It is believed that these can be a significant aid to decision-makers (since it is unrealistic to expect them to make their value judgments explicit).

The remaining issue is the treatment of the counter-factual problem: what would have happened in the absence of the policy? That is, to disentangle the effects of the policy from those of others. There are two approaches to dealing with this problem. The simplest approach is before-after and/or time-series analysis (eg Alterman *et al.*, 1984). The major drawback of this approach is that it only indicates a policy effect but does not establish its magnitude. The second approach is statistical analysis, eg regression analysis. The credibility of this analysis, however, hinges on the definition of the problem and the measurement of all the relevant variables accurately. The approach, thus, requires a huge amount of data that may not always be available. Furthermore, statistical analysis cannot incorporate qualitative data, unlike a

'descriptive' before-after analysis. The choice, however, between the two approaches is not a clear-cut one. The matter depends on several issues such as the resources available and the availability and types of data.

CHAPTER 5: MONITORING

5.1 INTRODUCTION

Monitoring, in a very broad sense, is the regular, deliberate and systematic collection and analysis of information (McLoughlin, 1973: 157). It is a crucial feedback loop within a cyclical process of policy-making providing information on both the performance of the policy at hand and its surrounding environment. Currently, there are strong indications of growing government commitment to regular monitoring of urban policies driven to a large extent by both the trend towards more public sector accountability and the search for greater value for money (as evident, for instance, in the SRB, see chapter 7; see also chapters 3 and 8).

Being an ex-post exercise, monitoring has a vital role to play in evaluation. As it focuses, among other things, on expenditure and outputs, monitoring is a necessary step towards assessing the effectiveness of a plan (Rossi and Freeman, 1989: 167). It can also guide further evaluation of a plan's indirect effects (Nientied and Schevz, 1988: 400). In itself, monitoring is an indispensable tool to respond to policy-makers' (continuous) demands for information on the outputs and performance of policies. Such information, in turn, is a crucial input to further research on the *impact* of policy.

Monitoring is one of the four components of the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) - one of the starting points for this research.

However, within that approach, monitoring had quite a narrow focus. It concentrated on the costs and outputs of the policy and their incidence (Alterman *et al.*, 1984: 383). These, no doubt, are crucial issues in assessing the effects of any policy. However, there seems to be a need for a broader focus of the analysis, rather than merely counting costs and benefits. It may suffice to recall that the uncertainty surrounding the wider environment (eg Friend and Hickling, 1988) is a crucial concern that has to be accounted for.

The question is, can monitoring play a role in broadening the scope of the analysis, or is it confined to the tasks it was assigned in the Integrated Approach? In other words, what role can monitoring play within a comprehensive evaluation framework - the main objective of this research? This chapter is the attempt to answer this question.

The following section first defines monitoring and its role and function within the planning and implementation process. It then explains the reasons why monitoring may be called for. It also distinguishes between monitoring and the related, yet distinct, concept of review. The section concludes with a brief review of the three, oft-cited categories of monitoring: strategic, impact and implementation.

Section 3 examines the two distinct views of monitoring: the control and extended views. Monitoring, for quite some time, was seen as a control exercise aimed at bringing the plan back on track. This view then gave way

to an extended view of monitoring which is more future oriented and wider in scope. Section 3 first delineates these two distinct views of monitoring and concludes with brief review of the definition and role of performance measures and indicators.

Section 4 looks at the influence of the political and administrative context of monitoring. This influence stems from three particular features of that context. First, its 'multi-organizational' nature. Second, the different policy-making modes, or styles. Third, and finally, the different levels at which decisions are made. Each of these features and their implications on the approach to, and conduct of, monitoring are all examined.

Although monitoring is a source of viable information, it requires a diverse range of data and information to start with. Information is both a requirement and a major problem in its own. Section 5 first explains the reasons why such a diverse information base is needed. It then looks at the different sources for this information and the problems associated with each and all of them. Finally, the concept of a computerised information system, its role and design criteria and the problems it faces are all reviewed.

The chapter concludes with remarks on the role and function of monitoring within both strategic planning and comprehensive evaluation and the theoretical basis of monitoring systems. The traditional, control view is rejected in favour of an extended view that can aid further evaluation of both

policy impact and the process behind this impact. It is, however, noted that there is no such thing as a 'universally applicable' monitoring system.

5.2 MONITORING: A BACKGROUND

5.2.1 Definition and Role

The classical, oft-cited definition states that monitoring is "the regular, deliberate and systematic collection and analysis of information (in a very broad sense)" (McLoughlin, 1973: 157). This brief definition acknowledges most, if not all, of the characteristics of an efficient monitoring activity. It is a continuous activity to be carried out on regular intervals, not a 'snap-shot' task. It is based on both a clear understanding of its role and function and on a sound systematic, perhaps theoretical, model or process, not a haphazard endeavour. Most important, monitoring is not only a data-collection task. It is the *analysis* of this data that makes monitoring an indispensable cycle of the planning process.

Other definitions attempt to delineate the role and function of monitoring. Within the context of a sub-regional policy, monitoring was assigned the task of attempting to "anticipate future developments, respond to current situations by a process of policy adjustments and consider past policies to establish their impact on current and future situations" (Perry and Chamberlain, 1977: 138) In broad terms, monitoring "establishes what is happening now and may happen in the future. It then compares these trends against existing policies and hence determines what needs to be done"

(Francis, 1981: 181). This involves answering three questions: "Are policies effective in achieving objectives?', 'Have policies resulted in unintended consequences?' and 'Are the assumptions and objectives of current policies still relevant?'" (Francis, 1981: 181).

In evaluating social policy, monitoring adopts a somewhat different, though complementary, focus. It attempts to assess programme coverage and delivery (Rossi and Freeman, 1989: 170). Programme coverage is the extent to which a programme is reaching its targeted population. Programme delivery, on the other hand, is the degree to which services are delivered in accord with the plan. In addition, monitoring is also directed at estimating the costs of the programme. This is the same view adopted in the Integrated Approach (Alterman *et al.*, 1984). That should come as no surprise since the approach was initially aimed at broad-aim social programmes.

5.2.2 Why Monitor?

There is a number of reasons why monitoring may be called for. It has long been recognised that the concern with monitoring is a consequence of plans' inability to cope with change and uncertainty as much as it is of theoretical understanding (Rose, 1979: 24). One way of facing uncertainty is through a continuous planning process that incorporates continuous feedback, about both the performance of the plan and the changes in its environment, through regular, continuous monitoring (eg Brown, 1984: 87; see also Floyd, 1978: 477 and Nijkamp and Rietveld, 1989: 234).

Monitoring may be the only systematic tool available to assess the viability of a programme in a time when rigorous evaluations are threatened, or even terminated, by budget shortfalls (Reid, 1990: 253-254). As it focuses, among other things, on expenditure and outputs, monitoring is a necessary step towards estimating the effectiveness of a plan relative to its costs (Rossi and Freeman, 1989: 167). Information generated by monitoring can guide further evaluation of a plan's indirect, and in the first place direct, impacts (Nientied and Schevz, 1988: 400); it is even argued that monitoring is a prerequisite to impact evaluation (eg Rossi *et al.*, 1979: 121). Within the context of social programmes, monitoring at the early phases is of vital importance. "Indeed, failure to monitor adequately and thoroughly during the piloting and early phases of programs may account for the failure of innovative programs to prosper when they are put in place on a large-scale basis" (Rossi and Freeman, 1989: 179-80).

5.2.3 Monitoring and Review

Monitoring should not be confused with the distinct, yet related, concept of 'review'. Monitoring is more concerned with the continuous re-assessment of different aspects of a plan or policy rather than with the periodic, comprehensive re-evaluation that is implicit in the review process (Scheurwater and Masser, 1981: 193 and Brown, 1984: 92). That is, "review is the process by which a planning authority re-appraises [re-assess the worth of] its adopted policy" (Lavery, 1977: 168). The timing of monitoring is dependent on the timing of events within the planning agency. It is also

dependent on the availability of information and the ability to process it. Review, on the other hand, is not a continuous activity (Kingston, 1981: 35-36). It can be seen as an 'independent' activity involving its procedures of data collection and analysis (see Scheurwater and Masser, 1981: 193 and Brown, 1984: 82).

5.2.4 Types of Monitoring

The classical, oft-cited categorization of monitoring has been highly influenced by "models of management control" (Floyd, 1978: 477; see also Brown, 1984: 85). This classification sees monitoring in terms of three levels corresponding to both the extent of uncertainty about the planning environment and the level of control that can be exercised over this environment (see Table 5.1).

Table 5.1: Types of Monitoring

Monitoring Type	Environment (Uncertainty)	Level of Control	Planning Level
Strategic	Imperfectly known (High)	Low	Strategic
Impact	Slightly unstable (Moderate)	Moderate	Operational/managerial
Implementation	Well understood (Low)	High	Implementation

Source: compiled by the author.

Strategic monitoring is much broader in its scope (Brown, 1984: 85) and is concerned with the "anticipation of possible future developments and the initiation of new policy in response to current situations" (Scheurwater and Masser, 1981: 193; see also Reid, 1990: 260). Impact monitoring is concerned with the assessment of two issues: 1) whether implementation is achieving the planned aims and, 2) are the forecasts upon which the plan was based still relevant and reliable (eg McLoughlin, 1975: 161; Wedgwood-Oppenheim *et al.*, 1975: 9 and, Reid, 1990: 261). Implementation (output) monitoring is intended to check whether implementation is in accord with the plan or not (eg McLoughlin, 1975: 161; Wedgwood-Oppenheim *et al.*, 1975: 9 and, Brown, 1984: 85).

5.3 MONITORING: TWO DIFFERENT VIEWS

5.3.1 The Traditional View

The 1970s witnessed a British intensive activity in monitoring research prompted, to a large extent, by the 1968 Town and Country Planning Act, followed by the 1971 Act and its amendments and subsequent Central Government Circulars (see, for instance, DoE 1974 and 1975). Most of this research focused on devising 'Monitoring and Advisory' (M/A), systems and/or units as part of the then newly prepared, mostly sub-regional, Structure Plans. Most notable among these are the Notts./Derbys. Monitoring and Advisory System (Riera and Jackson, 1971) and Unit (Gillis *et al.*, 1974). The underlying concept of the monitoring process upon which this research was based was

to provide an early warning system giving information on any deviations from the intended course of the plan and suggesting possible remedial action which would form the basis of a systematic review of the ... plan (Lavery, 1977: 166).

This came to be known as the 'traditional, control' view of monitoring (eg Floyd, 1978: 476; Bracken, 1981: 84 and, Brown 1984: 83). It was based on the rational-comprehensive model of planning which, in turn, was based on the rationalistic model of decision-making. This model has been the subject of severe criticism (eg Etzioni, 1967: 385-6; Haynes, 1974: 6 and, Drummond, 1991a: 25-40) and, in effect, the traditional view of monitoring as a control activity, sustained mounting criticism (eg Bennett, 1978: 311-313 and Bracken, 1981: 86-89) This criticism centred around the impractical assumptions of "complete information, singular objective, exhaustive consideration of alternatives, and rationality of decision making" (Brown, 1984: 84).

5.3.2 The Extended View

The recognition of the limitations of the traditional view of monitoring and the then emerging approach to planning as a process of strategic choice (eg Friend and Jessop, 1977) together with the 'mixed-scanning' approach (Etzioni, 1967)⁽⁹⁾ all led to a move away from that 'control' view of monitoring towards a broader one; the so-called 'extended view'. Haynes (1974: 18-20) was perhaps the first to write about a 'new' view of monitoring. He saw monitoring at the "interface between the *information field* and the

⁽⁹⁾ In simple terms, 'mixed-scanning' is a two-step model: a 'coarse-grained' scanning of the entire area of interest "so that no major option will be left uncovered" followed by an in-depth "highly detailed" investigation of certain aspects "so that the option selected can be explored as fully as feasible" (Etzioni, 1967: 389).

problem identification function" (see Fig 5.1). The latter is being "sustained by an inflow of information" from the former which is maintained by the monitoring function. "The monitoring function is required to perform the task of selecting relevant information from the information field, arranging and organizing that information, and disseminating it to the appropriate user."

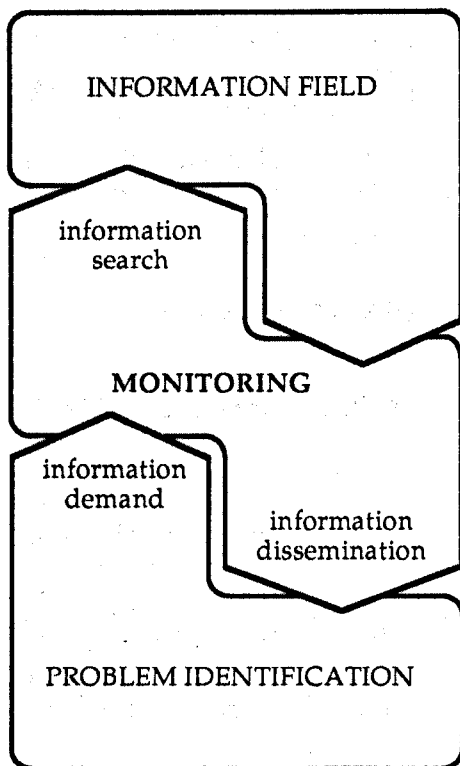


Fig 5.1: The Context of Monitoring Activities

Source: Haynes, 1974: 18.

The argument for the 'extended view' of monitoring beyond that of control, in line with the concept of strategic choice, was best elaborated by Floyd (1978: 477):

an examination of the major choices of action facing an authority will reveal that there are some that are not urgent; they can be deferred and exploratory activities set in train to reduce some of the uncertainties surrounding the choice. ... It follows that a central feature of monitoring will be the resolution of the uncertainties surrounding these deferred choices and hence its role in providing a basis for making more informed decisions in the future. It will also be concerned with anticipating new problems and choices. Equally, monitoring should help to identify critical choices to be faced by other agencies so that sufficient time is available for the authority to try to influence the decisions that are eventually taken. Monitoring then should be more concerned with the *future* rather than with past changes — except in so far as an understanding of them informs future decisions (stress in original).

This view apparently moves beyond the limited concern with policy to examine "new issues, changing values, unforeseen problems, and new opportunities" that may entail a need to modify existing policies or introduce new ones (Brown, 1984: 88).

Nevertheless, Floyd (1978: 479-480) recognized that this view has also its implications, and may actually face difficulties. At the outset, "there is the problem of identifying important areas of future choice facing the authority or other agencies"; the so-called "key issues". Once these key issues have been identified, "monitoring clearly has a 'follow-up' function of keeping these issues under observation — of monitoring their development." Forecasting will play an important role in both identifying new issues and in assessing whether they will be of central importance over time. That is, the job of monitoring is to stay "a few steps ahead of the game". Forecasting is not, however, an error-free, ever-accurate activity. Finally, there is the issue of effective communication and utilization of monitoring findings. This "depends largely on the way in which monitoring is organized."

5.3.3 Performance Indicators

Performance measures are a link between the inputs of an activity and its outputs; a link that can be extended to include not only dimensions of economy and efficiency but effectiveness as well (Butt and Palmer, 1985: 32). The two terms, indicators and measures, may be used interchangeably and, the two concepts may even merge in practice. However, a conceptual distinction, which stems mainly from the measurability of performance, was drawn between them

Where performance can be measured precisely, reference is made to a performance measure. ... If, however, as is frequently the case, there is no precise interpretation of the data, then it is referred to as a performance indicator (Jackson, 1988: 11).

Performance indicators can play a number of important roles, eg provide information for monitoring and ex-post evaluation and increase management accountability (Jackson, 1988: 12). They are an important tool in the search for value for money in the public sector (Butt and Palmer, 1985: 32). Yet, their design and application are faced with a multitude of problems, eg the vague statements of policy objectives, information problems (availability, timeliness, ...), the focus on quantity rather than quality, the tendency to use a limited number of indicators and the danger of inducing 'short-termism' (Carley, 1988: 23; Flynn *et al.*, 1988: 35; Holtham, 1988: 31; Jackson, 1988: 14 and Palmer, 1993: 35).

Although performance indicators were a criticised aspect of the M/A systems (eg Bennett, 1978: 312; Floyd, 1978: 476 and Rose, 1979: 31), the call

for an extended view of monitoring should not rule out their use. There will always be a need to measure the performance of the plan. For one reason, an understanding of past and current trends will be an aid to future decisions (eg Floyd, 1978: 477). Another reason is the growing emphasis on both public sector accountability and achieving greater value for money which has given rise to a culture of output measurement⁽¹⁰⁾.

However, and in order to maximise their usefulness, there are several considerations to be taken account of in the design and application of performance measures/indicators (see Jackson, 1988: 12 and Likierman, 1993: 15-21). Perhaps the most important of these considerations is to choose those indicators which reflect the nature of the policy at hand and its objectives as closely as possible (or feasible).

5.4 MONITORING AND THE POLITICAL CONTEXT

Whether it is conducted as part of a 'comprehensive' evaluation exercise, or in isolation, monitoring - as the case with evaluation - can hardly be divorced from its political context. Given that "planning is coloured with politics" (Voogd, 1983a: 7), and since monitoring is but one cycle of a continuous planning process, it follows that it will also be coloured with politics. For one thing, monitoring, like evaluation, will turn out to be an

⁽¹⁰⁾ This is particularly clear in governmental guidelines for the monitoring and evaluation of urban regeneration initiatives, eg City Challenge and the Single Regeneration Budget. Although the distinction between outputs and outcomes was recognised, the focus is primarily on measuring outputs, eg number of jobs created (see chapters 7 and 8).

academic exercise "if it does not sufficiently inform the decision-makers and the public so that they can use the information to arrive at more rational decisions" (Hill, 1985a: 31). The operational setting is, therefore, a fundamental contextual element influencing the "choice, development and utilization of monitoring techniques" (Rose, 1979: 23-4; see also Wedgwood-Oppenheim *et al.*, 1975: 7 and Connor, 1993: 10).

The situation is further complicated by three factors. Firstly, in almost any policy-making exercise, a multitude of agencies, of various levels, interests and powers, become involved in the planning and implementation process. It has been shown that even the most agreed-upon plans may fail to achieve their goals simply because a (relatively small) number of agencies were involved in implementation (Pressman and Wildavsky, 1984). There are two distinct forms of organizational connections. In the first form, different agencies may be connected through "allegiance to a common *authority* as in the case of the departments of a local authority" whereas in the second form different agencies may be connected "through the nature of their operations even though no common authority exists" (Friend and Jessop, 1977: 121, stress in original). These two forms are known as 'organizational' and 'multi-organizational', respectively.

It should go without saying that planning will almost always fall within the second category. It was argued that "in important matters of public policy, the evidence was that the 'multi-organizational' context of decision-making

would have to be regarded not so much the exception as the rule" (Friend *et al.*, 1974: xi).

This multi-organizational nature of policy-making has a significant bearing on monitoring. The extent to which pressures for commitment "are responded to must then be monitored, together with the other agencies' adherence to advocated policy guidelines and their fulfilment of commitment they have made" (Floyd, 1978: 478). In other words, "continuing monitoring of these pressures will represent an equally important aspect of the monitoring activity" (Floyd, 1978: 477). It should go without saying that such information is a vital input into any attempt to explain why things happened the way they did, ie implementation analysis (see chapter 6).

The second fact that complicates monitoring is the existence of different modes, or styles, of policy-making. There are several modes of planning and decision-making each entailing a slightly different mode of implementation (see chapter 2, p. 29 and chapter 6, p. 172). It can be argued that each political/planning, and in turn implementation, mode will require a different approach and focus of monitoring. For instance, in the 'central' mode, where compliance with the advocated policy is the principal variable, monitoring is very likely to veer towards the traditional view of 'control'.

The third and final factor is the different levels at which decisions are made (strategic, managerial, ...). At each of these levels, a different type of

monitoring will be more appropriate than others (Rose, 1979: 24; see Table 5.1 above). This is due in large to the different information needed at each level. Apparently, each of these three facts will have its implications on the approach and conduct of any monitoring process. "Different political consideration will clearly permit or inhibit different approaches to monitoring" (Bracken, 1981: 85).

5.5 INFORMATION: A REQUIREMENT AND A PROBLEM

5.5.1 The Diverse Information Base

A diverse information base is one of the most crucial requirements for monitoring. The demand for such a base is initiated by several factors. First, the expansion in our conception of the urban system to include a multitude of social, economic, environmental and political variables. This means that the information system "*must* expand correspondingly if anything like effective understanding ... is to be achieved" (McLoughlin, 1973: 227, stress in original). Second, the move towards a cyclical process of strategic choice. Planners preparing a 'strategic' policy were confronted with an enlarged demand for information far beyond the simple 'land-use' domain into a much wider area (Howells and Smith, 1977: 150). In particular, 'intelligence' (qualitative, verbal information) became more and more of central importance to monitoring (Brown, 1984: 93).

Third, and related, the extended view of monitoring at the strategic level. "To reduce uncertainty, the information base for strategic monitoring must be

far more than a passive exercise in the collection and neat arrangement of numbers easily available on these current trends which are vaguely thought to be relevant. It includes selection, analysis and a systematic search for verbal information" (Francis, 1981: 182; see also Wedgwood-Oppenheim *et al.*, 1975: 15 and Brown, 1984: 91).

Finally, the 'multi-organizational' nature of planning (see above). The "different but overlapping aspects of the relationship" between the various agencies themselves, and between them and the public "require information of many different kinds organized in as many different ways" (McLoughlin, 1973: 235-6). The matter is further complicated by the fact that the "public" is not a unitary group. It comprises various interest groups with different values and, hence, different information requirements. Furthermore, each level of the decision-making process has its different information requirements (Haynes, 1974: 13 and, Scheurwater and Masser, 1981: 195) which depend on the "powers, functions and responsibilities of the policymakers" (McLoughlin, 1975: 161).

5.5.2 Information Sources

Monitoring will draw upon a wide range of information sources (Francis, 1981: 181). However, since monitoring is not a mere data collection task, special attention "must be given to the development of an efficient means of information selection in terms of particular tasks" (Scheurwater and Masser, 1981: 193).

Basically, the source of information "for the urban management and planning system is the environment of that system, the urban systems-of-interest themselves" (McLoughlin, 1973: 243; see also Haynes, 1974: 19-20). There are several sources of information: surveys, observations, questionnaires, etc. Surveys are, however, "ruled out in all but the most exceptional cases, not only by questions of cost but also because of the long lead time needed to prepare a survey, carry it out and make its findings available" (Masser, 1984: 9, see also Floyd, 1978: 481).

Table 5.2 lists some of the most common information sources. Clearly, there can be no clear-cut choice as to which source is 'more appropriate', or 'more relevant', than the others. Each source has its advantages and drawbacks. The choice also depends on the characteristic of the case at hand (particularly the resources and time available). Perhaps it would be more appropriate to regard the different sources as supplementary, rather than alternatives, to each other. Indeed, it is not uncommon for monitoring and evaluation to draw upon several sources at once.

It should be noted, however, that data collection is not solely a task of accumulating statistical material. 'Intelligence' has come to be of central importance to monitoring. There is a near-consensus in the literature that intelligence gathering is "an essential part of the data collection process" (Masser, 1984: 11; see also Francis, 1981: 183 and Brown, 1984: 93). Intelligence is further divided into two types: "hard (e.g. policies approved, commitments

made, ...) and soft (e.g. information on emerging policies, proposals under consideration, ...)" (Brown, 1984: 93; see also Francis, 1981: 182-183).

Sources for such information are "virtually infinite and the procedures used to sense and scan such material will therefore need careful attention and regular review to ensure that relevant information is channelled into the monitoring/decision-making process" (Brown, 1984: 93). These sources include published and unpublished reports, committee papers, the press, personal contacts and consultations with other agencies and interested groups. "A key task for those involved in strategic monitoring, therefore, is to establish formal and informal networks that facilitate intelligence gathering" (Masser, 1984: 11; see also Francis, 1981: 183 and Brown, 1984: 93).

5.5.3 Information Problems

In addition to the problems associated with each source of information, monitoring faces "a number of fundamental problems that are associated with the use of information drawn from such a wide range of different sources" (Brown, 1984: 93). One such problem is the availability of data and information, or rather lack of both (Steeley, 1976: 11 and personal contacts, 1995). Accessibility is another related problem. Information may exist in the exact form required, but it may be inaccessible for a variety of reasons (eg confidentiality and security).

Table 5.2: Some Common Sources of Information

Source	Advantages	Disadvantages
1. Programme Records	Most of the programmes keep their records. This means they can be a ready source of data without additional costs or time. They also have the advantage of setting down events at the time of their occurrence not in retrospect.	Records differ widely in their quality, extensiveness, accuracy and reliability from one programme to another. "It should be noted that all record systems are subject to some degree of unreliability." In addition, programme records are designed "primarily to serve the administrative and management needs of program staff" and they may, thus, be of little use for monitoring purposes (Rossi and Freeman, 1989: 209). In social, individually-based programmes, examining records may be faced with ethical or legal constraints.
2. Observations	"Observations can be highly credible when seen as the report of what actually took place presented by disinterested outsider(s)" (King <i>et al.</i> , 1987: 47).	To observe is to disturb, this is the major criticism of observation methods. The presence of observer(s) 'may change the behavior of program personnel and participants' (Rossi and Freeman, 1989: 209). They are not easily taught, they are highly time-consuming and they can produce data that are difficult to structure or analyze.
3. Questionnaires	Answers to a wide variety of questions can be obtained anonymously. Questionnaires allow respondent time to think before answering. They can also cover distant sites easier and cheaper than other methods. Further, uniformity of the data supplied is imposed easily.	Compared to interviews, questionnaires are more rigid and constraining to respondents since it is usually easier for people to express themselves orally than in writing. There is also no guarantee as to response rates.
4. Interviews	Interviews permit flexibility. They also allow the interviewer the chance to pursue unanticipated lines of inquiry.	They are time-consuming, and in cases can be costly as well. There is always the difficulty of covering distant sites. In some cases, the interviewer, unintentionally, can influence the interviewee's response.

Source: modified from King *et al.*, 1987 and drawing on material from Rossi and Freeman, 1989.

Perhaps the most difficult problem is the sheer bulk of information an extended-view monitoring is very likely to require. It is not a problem of assembling enough information "but rather of filtering out the overwhelming abundance of surplus irrelevant material" (Brown, 1984: 94; see also Wedgwood-Oppenheim *et al.*, 1975: 23 and Haynes, 1974: 21). Consistency and comparability of information drawn from such diverse sources is another major difficulty (Masser, 1984: 10). There are bound to be differences in the time and spatial scales, types, ... etc of such information. For instance, a very common problem is the "incompatibility of the geographic areas for which statistical series are compiled" (Brown, 1984: 93).

5.5.4 Information Systems

The sheer bulk and diversity of data, information and intelligence; the constant inflow of information from the environment at different types and time and spatial scales; the policy-makers and interest groups continuous demands for different types of information and, the number of important decisions that have to be made at various stages and levels of the planning and implementation process are all reasons why information systems have become a crucial component of any monitoring process. The intricacy of urban problems and the considerable social and economic costs of wrong decisions "certainly cost-justify the development of better information systems to support public policymaking" (Worrall, 1990: 461).

The concept underlying information systems is yet very simple. In a very

broad sense, an information system is a systematic procedure for capturing, sorting, storing, retrieving, manipulating and presenting information (eg Haynes, 1974: 13). The primary components of an information system are "a dynamic data base and a data processing system consisting of both computer hardware and software" (Haynes, 1974: 13). A principal aim of these systems is to "supply the relevant data in a usable form to the right user at the right time in the decision making process" (Haynes, 1974: 13). In other words,

The fundamental test of the usefulness of these systems must lie in the contribution they have made to the development of more responsive policy and to the more effective planning of urban areas and regions (Worrall, 1990: 451)

Nevertheless, such a seemingly straightforward concept involves major difficulties and at times may be a far from easy task. First, and foremost, an information system "must be designed in the context of the planning process" it serves (Worrall, 1990: 451). Careful a priori planning is perhaps the single most significant variable in setting-up an information system (Semke and Nurius, 1991: 354). At the outset, "it is essential to properly conceptualize the information that will be required by various users of the system" (Rossi and Freeman, 1989: 214). Another critical consideration is that "all the persons who provide and enter data must understand the utility of the system, its rules and definitions, and their responsibility to collaborate in its implementation" (Rossi and Freeman, 1989: 214).

There are several considerations that have to be taken into account in setting-up an information system; the so-called 'design criteria'. These criteria,

in themselves, highlight the numerous problems such a task is bound to face. Periodicity, timeliness, integrability, reliability, definitional rigour, topicality, ability to forewarn and flexibility are all among these considerations (Worrall, 1990: 452-453). Consistency and comparability of data are two more requirements that need to be guaranteed "as far as it is possible" (Scheurwater and Masser, 1981: 200).

In a broader sense, the organizational context will represent another consideration since it will determine the extent of overlap "with other activities in the organization" (Gillis *et al.*, 1974: 30). It also has a bearing on the relation with other agencies in terms of data collection and reporting procedures. The organizational setting of information systems dictates two more considerations. First, an information system should support decision-making at the various levels of the organization. Second, care should be taken in the design process to minimize the burden of "clerical minutiae and dissonance with direct practitioner values" (Semke and Nurius, 1991: 354).

Finally, when setting-up an information system there are two fundamental considerations to be taken into account. First, the development of an information system should not be seen as a 'one-off' task. It "should evolve through experimentation and operational experience, with successive stages of the system being designed to supplement or improve upon its predecessors" (Haynes, 1974: 13). Second, in all phases of developing an information system the issues of "information use, relevance and the costs and

benefits of different modes of data gathering must be addressed" (Semke and Nurius, 1991: 355).

5.6 CONCLUDING REMARKS

5.6.1 Monitoring and Strategic Planning

There can be no argument about the proposition that monitoring is a vital component of the continuous planning process. It represents one feedback loop within a cyclical process where each cycle builds upon the lessons of the previous one(s). The question here is 'What kind of feedback?'. Put another way, what function is monitoring to fulfil within the planning process and how does it relate to other cycles?

There are two possible types of feedback (Haynes, 1974: 10-11):

1. "Negative, or deviation-controlling feedback, ... an error-correcting mechanism ... which tends to preserve or maintain a system's given form, organization or state."
2. "Positive, or deviation-amplifying feedback ... the elaboration of the system's form, organization or state".

These two types correspond almost perfectly to the two views of monitoring discussed above: the control and the extended views, respectively. The idea behind negative feedback is almost identical to that underlying many of the M/A systems of the 1970s in Britain. Deviations from what was initially planned are detected by means of performance indicators. The significance of these deviations is then assessed and the results are feedback forming the basis of a review of the plan to bring it back on track. On the

other hand, the extended view, as elaborated by Floyd (1978), is more 'future-oriented' and concerned with anticipating future important issues, decisions, actions, potentials and problems. Since the two views are, in a sense, complementary to each other, Floyd (1978) recognized that past and current changes are also to be examined given that an understanding of them will aid future decisions.

Although the extended view, with its positive feedback, may lead to a plan being reviewed and, ultimately modified, the main thrust behind this is entirely different from the control view. Where in the latter a plan is modified in response to past and present changes, the extended view aims at 'rolling a plan forward' to cope with the predicted future — and, to varying degrees, to avoid past mistakes and learn from achievements.

Apparently, in a strategic planning context where uncertainties are a major concern, the extended view becomes of mounting importance. However, it has to be stressed that current and past changes are not to be overlooked. It also follows from the concept of continuous planning that concern with the future should not lead to the neglect of lessons of the past. Indeed, monitoring appears to be the only viable, systematic tool to provide information on a regular basis about the performance of plans; the best way to overcome the oft-cited criticism of delayed evaluation results. Furthermore, concern with incidents of failure should not be at the expense of achievements; these should be granted as much attention as failures (eg Haynes, 1974: 22).

However, there seems to be a (wide) gap between theoretical developments and practical applications of monitoring. On the one hand, the latest urban regeneration initiatives in Britain (City Challenge and the Single Regeneration Budget) exhibit a very clear commitment to monitoring (see chapters 7 and 8). In the SRB, for instance, continuation of funding was even made subject to the setting up of monitoring arrangements, regardless of the actual outcomes of the scheme. On the other hand, government guidelines on monitoring (and the actual exercise) are primarily concerned with measuring the outputs (and to a lesser degree, the outcomes) of various schemes (ie implementation monitoring). Partnerships are urged (with a threat of withholding grants) to re-consider their plans in the event of under- or poor performance. In other words, it seems that the government's conception of monitoring still adheres to the traditional, control view. It is largely unknown whether partnerships adopt a different view, though there are no apparent incentives for them to go beyond the scope of the guidelines.

5.6.2 Monitoring and Comprehensive Evaluation

The 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) seems to have adopted a view of monitoring that is closer to the traditional view than the extended one. Monitoring was confined to the measurement of policy costs and outputs and their incidence (Alterman *et al.*, 1984: 383). In neither the Approach nor its application (eg Hill *et al.*, 1990) has any attention been paid to either the wider environment or emerging issues, policies, potentials and

problems. The above discussion on the role and scope of monitoring would serve only to reject such a circumscribed view.

Having asserted the need for an extended view of monitoring, the difficult question then arises: 'What monitoring system is to be adopted?'. Unfortunately, there is no single approach that is applicable in all situations (eg Haynes, 1974: 17; Brown, 1984: 90 and Masser, 1984: 2). Differences "in the scale, scope and internal configuration of planning agencies mitigate against the complete design of a universally applicable systems" (Haynes, 1974: 17). Differences in the nature and scope of a plan, the political and administrative context, the socio-economic conditions and the environment within which a plan is to operate are all reasons why any attempt to forward a 'universally applicable' monitoring system is bound to fail.

However, recommendations can be advanced on the theoretical basis of monitoring. The 'mixed-scanning' approach is probably the most suitable to adopt. The extended conception of the urban system, and consequently of urban policies; the multitude of policy- and decision-makers, agencies and interest groups that eventually become involved in planning with their different, often conflicting, information requirements, and the scarce resources are all reasons that militate against a holistic approach of monitoring every single aspect of the plan and its environment. Nevertheless, it is as much a mistake to adopt a narrow scope of monitoring right from the outset; important issues are almost destined to escape unnoticed. Besides, to focus

analysis on some particular issues rather than others is more of a political decision than a technical one. For all of that the 'mixed-scanning' approach seems to provide the best possible answer to such a dilemma.

Within a comprehensive evaluation framework (the main objective of this research) monitoring has clearly a crucial role to play. It is an indispensable source of information on policy outputs (and possibly outcomes). The clear emphasis on regular monitoring built into the latest urban policies (City Challenge and the Single Regeneration Budget) serves only to confirm policy-makers' growing interest in such information (regardless of their reasons). This information, in itself, represents a crucial input to any further evaluation of the policy at hand, either in tackling the counter-factual problem or in explaining why things happened the way they did (implementation analysis - see chapter 6).

However, to fulfil this role, a much wider scope of monitoring must be adopted. First, it is insufficient merely to measure the outputs of policy; it is far more important to measure its outcomes, if we were to ascertain the *impact* of that policy. To do so, performance measures/indicators must reflect as far as feasible the nature of the policy and its objectives.

To tackle the counter-factual problem and disentangle the impact of the policy from other trends both these trends and their impacts must equally be under constant assessment. That is, monitoring must have a wider scope than

merely the policy at hand. The extended view of monitoring, by definition, launches monitoring into the political context of the policy at hand, and even beyond. The multi-organizational nature of policy-making, which should be regarded as the norm rather than the exception, has a significant bearing on the outputs and outcomes of the process. An understanding of these outputs/outcomes requires knowledge of the workings of these dynamic structures. In other words, this information will be an important input into implementation analysis.

In addition, the monitoring system should not be the result of a 'one-shot' design task. It should evolve over time through experimentation and experience of previous attempts. The system itself should be the subject of regular, objective monitoring and evaluation to examine its efficiency. Improvements "should not be based on a subjective wish to do better" (Riera and Jackson, 1971: 3.7). They should be based on an as objective appraisal of the system as possible.

Any monitoring system will draw upon a diverse range of information of various types and sources. This entails, first, the need to establish a dialogue right from the beginning with the various groups and agencies involved in the planning and implementation process. This facilitates both the flow of information and the communication of results. Second, there is a need for an information system that is capable of both handling such a diverse data-base and presenting information at the different required forms at the right time

in the decision-making process. No doubt this information system will be an automated, computerised one. However, technical specifications should not be the only factor determining the choice of technology to be used. It has to be borne in mind that the principal aim of these systems is to provide the right information at the right time for potential users.

CHAPTER 6: IMPLEMENTATION ANALYSIS

6.1 INTRODUCTION

Implementation is the *process* through which a policy/project and its objectives are put into effect. Implementation analysis, in turn, is the attempt to identify and understand the factors and forces influencing this process. In other words, implementation analysis aims to answer the crucial question: why did things happen the way they did? In so doing, implementation analysis is a 'learning' process and should, therefore, be an integral component of any evaluation exercise. The sheer cost of wrong public decisions should cost-justify the endeavour to understand why policies succeed or fail in achieving their objectives.

Implementation analysis is the fourth, and final, component of the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984). Two major questions were asked: 1) how does implementation take place? and, 2) how effective is the implementation process? However, the then rudimentary state-of-the-art of the field provided no set of tested methods to answer these questions. Which analytical model should be adopted was an unanswered question.

The current situation is remarkably different. Implementation analysis models have long been established and applied in numerous studies. The strengths and weaknesses of each of these models have also long been recognised. We seem to be in a better position to search for an answer to that

question - which model should be adopted as an element of a comprehensive evaluation approach? This chapter is devoted to the attempt to answer this question.

Section 2 first defines the implementation process. This is followed by the definition of implementation analysis and its scope and the questions it aims to answer. These, in turn, indicate the crucial role implementation analysis can play in evaluation. However, the relationship between implementation analysis and impact evaluation has been the subject of some contrasting arguments which are also reviewed in brief.

There are four modes of implementation (top-down, bottom-up, adaptive and evolutionary) and two principal approaches to analysis (top-down and bottom-up). The primary focus of the top-down approach is on the higher-level of policy-making process (the Centre). In contrast, the bottom-up approach focuses on the lower-level of the process; the implementing agencies and target groups (the Periphery). Thus far, few attempts have been made to synthesize the two approaches. Section 3 reviews two examples of the top-down approach (Van Meter and Van Horn, 1975 and, Sabatier and Mazmanian, 1981 and 1983), one example of the bottom-up approach (Hanf *et al.*, 1978) and the most notable attempt of synthesis (Sabatier, 1986). It also examines the strengths and weaknesses of top-down and bottom-up models in order to inform the choice on which model could be adopted in comprehensive evaluation.

This question of choice is the central focus of the final section. Since circumstances differ dramatically from one case to the other, there can be no one single answer to that question. The choice, moreover, is largely dependent on the political and administrative context within which planning, implementation and evaluation all take place. An examination of the attributes of different policy-making modes, however, provides no clear-cut choice. It is therefore argued that the synthesis of approaches represents a better starting point. Whether emphasis should be on the Centre or the Periphery is still dependent on the dominant policy-making mode. However, emphasis on one side should not be at the expense of neglecting the other.

6.2 DEFINITION AND SCOPE

The term 'implementation' should, to a large extent, be a self-explanatory one. Basically, it is the *process* through which a policy/project and its objectives are put into effect (eg Barrett and Fudge, 1981: 11; Sabatier and Mazmanian, 1981: 5 and 1983: 143 and, Pressman and Wildavsky, 1984: xxiii).

In more detailed terms, implementation is

the process by which decisions taken by various actors enhance or weaken the chances that intervention will be undertaken in accordance with the policy-in-reference (Alterman, 1981: 5)

This definition explicitly emphasises four key facts about implementation. First, it takes into account not only decisions in favour of the policy but also those that are not. It is not uncommon that decisions are taken which run counter to the goals and objectives of the policy at hand (as was the case, for instance, with the Programme for the Valleys, see chapter 8). Second, it takes

a probabilistic point of view; talking of *chances* as there is often no guarantee of success or failure. Third, it is relative; it takes the point of view of the particular policy of interest. Again, the same decision may be in favour of one policy and run counter to another. Finally, it recognises explicitly the fact that there is a multitude of actors involved in both the policy at hand and other policies and projects operating within the same environment.

Implementation analysis is, then, an endeavour that seeks to identify and understand the factors and forces influencing the process; the purpose is to enhance the prospects of more effective implementation (Alterman, 1981: 3). Implementation analysis offers "a new understanding of how the system succeeds or fails in translating general policy objectives into concrete and meaningful public services" (Van Meter and Van Horn, 1975: 405).

In other words, the primary concern in implementation analysis is to explain why things happened on the ground the way they did; what were the influential factors behind the outputs/outcomes of the policy (eg Judd, 1987: 24). It follows that implementation analysis will facilitate the modification of a current programme to make it more efficient (construct validity); it can also facilitate recommendations on the applicability of a programme, or parts of it, into other locations (external validity) (eg Palumbo and Oliverio, 1989: 340; see also Judd, 1987: 26). It is also argued that implementation analysis enables a closer examination of a policy's unintended and indirect impacts (Judd, 1987: 26). Outcome evaluation, by focusing on the variables a programme is

assumed to influence, is unlikely to detect such effects. Implementation analysis, on the other hand, includes a wider set of variables that might not be affected by the programme.

Failure to examine how a programme is being/has been implemented is also a threat to other dimensions of validity: internal and statistical conclusion (Palumbo and Oliverio, 1989: 337-9). Internal validity is concerned with the extent to which the effects of a programme can be disentangled from extraneous factors. Statistical conclusion validity is concerned with the possibility of making errors type I and II. A Type I error, or false positive, is to make a positive decision when the correct decision is a negative one; that is, to conclude that a policy has an effect when it actually does not. A Type II error, or false negative, is to make a negative decision when the correct decision is a positive one; that is, failing to detect a real policy effect (see Rossi and Freeman, 1993: 228).

Implementation analysis is, therefore, an indispensable analytical tool in policy evaluation. Nevertheless, the relationship between implementation analysis and impact evaluation has been the subject of debate. On the one hand, it was acknowledged that "analysis of social, psychological or environmental impacts is clearly a different ballgame than analysis of the decision-making process in implementation, requiring different theory and different tools" (Alterman, 1981: 7). On the other hand, "such a circumscribed approach" was rejected because it "often precludes some of the most

interesting and important aspects of implementation analysis, namely, the adequacy of the underlying causal theory and, in regulatory programs, the degree of (private) group compliance with agency decisions"; "it is precisely in revealing the inadequacy of the underlying causal theory or the limited ability of regulatory agencies to bring target groups into compliance that implementation studies can often make their greatest contribution" (Sabatier and Mazmanian, 1983: 148-9).

Clearly, Sabatier and Mazmanian are primarily concerned with regulatory programmes, which need not always be the case. Effectively, however, what they are arguing in favour of is a 'comprehensive' view of evaluation; a view that combines the two questions: "What happened?" and, "Why did it happen this way?", together. It is, nevertheless, to be borne in mind that outcome and process evaluation are each underlined with a different goal (Judd, 1987: 23). In outcome evaluation, the goal is demonstration; to demonstrate the policy's effects. The goal in process evaluation, on the other hand, is of explanation; why there is, or there is not, an effect. The two concepts - implementation analysis and impact evaluation - are "distinct though not unrelated" (Van Meter and Van Horn, 1975: 448):

The study of impact searches for the consequences of a policy decision. ... By focusing on those activities that affect the rendering of public services (i.e., performance), the study of policy implementation highlights one of the forces that determines policy impact. ..., impact studies typically ask "What happened?" whereas implementation studies ask "Why did it happen this way?" (Van Meter and Van Horn, 1975: 448)

Each requires different analytical tools (Alterman, 1981: 7). Yet, it is by combining the two together that a better understanding of a policy's

performance, and perhaps more rational decisions, can be achieved.

6.3 IMPLEMENTATION: THEORY AND ANALYSIS

There are four models of implementation: top-down, bottom-up, adaptive and evolutionary (Palumbo and Oliverio, 1989: 341-2). The top-down model maintains that ideal implementation takes place in a centralised organisation with control emanating from the top (eg Sabatier and Mazmanian, 1981 and 1983). It assumes that policy is driven by prospective rationality (the rational model of policy-making; eg Drummond, 1991a⁽¹¹⁾). The bottom-up model focuses on the considerable discretion of 'street-level' implementors and their influence on the direction of programme implementation (eg Hanf *et al.*, 1978). This approach is grounded in the theory of retrospective rationality; organisations do not plan beforehand. Adaptive implementation believes that programmes change and adapt in response to local conditions (in situations of ambiguous goals and lack of technology) (eg Majone and Wildavsky, 1979). Evolutionary implementation is similar to adaptive in its belief. However, it sees progressive improvements during implementation rather than the simple adjustments in adaptive implementation (eg Majone and Wildavsky, 1979).

The first two of these models correspond directly to the two analytical frameworks that dominate the field of implementation analysis: the top-down and bottom-up, respectively. There are no analytical models that explicitly

⁽¹¹⁾ This is the model upon which the traditional, control view of monitoring was based; see chapter 5, p. 122

address either adaptive or evolutionary implementation. However, it seems plausible that both veer more towards the bottom-up approach rather than the top-down one if only because of the former's focus on the operational and local levels.

6.3.1 Top-Down Approaches

These models start "with a policy decision (usually a statute) and examine[d] the extent to which its legally-mandated objectives were achieved over time and why" (Sabatier, 1986: 22). In other words, they take the policy as given and attempt to explain its success, or otherwise, by examining what was right, or wrong, in the implementation process and the implementing agency (Hambleton, 1983: 405). Analysis of the implementation process is carried out through the examination of a 'list' of variables these models usually postulate. The two, most common frameworks within this category are those of Van Meter and Van Horn (1975) and Sabatier and Mazmanian (1981 and 1983).

The first model (Van Meter and Van Horn, 1975: 462) "posits six variables which shape the linkage between policy and performance" :

1. Policy Standards and Objectives.
2. Policy Resources.
3. Interorganizational Communication and Enforcement Activities.
4. Characteristics of the Implementing Agencies.
5. Economic, Social and Political Conditions.
6. The Disposition of Implementers.

These are known as the independent variables. The model, as depicted in Fig

6.1, shows the relationships between them and the dependent variable - policy performance. It also makes explicit the relationships among the independent variables. It was believed that examination of these linkages would "lead to more systematic explanation of policy performance" (p. 478).

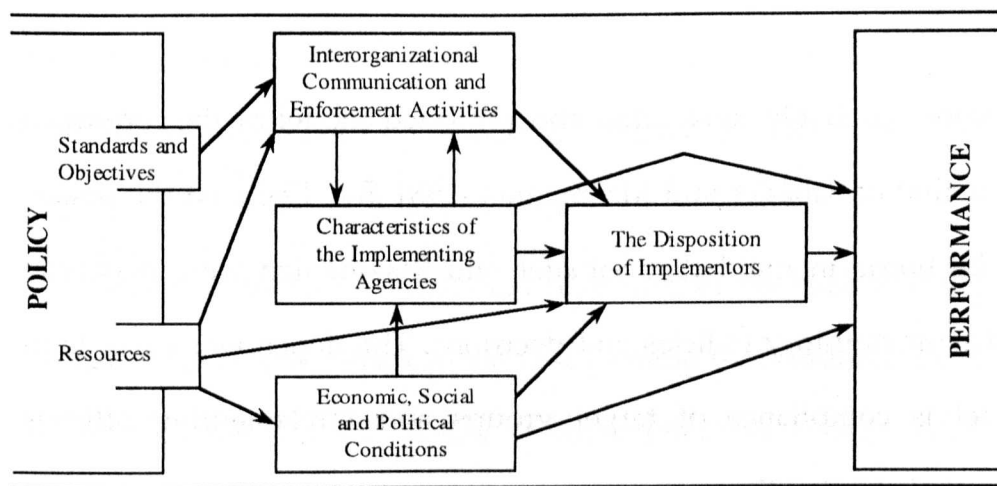


Fig 6.1: A Model of the Policy Implementation Process

Source: Van Meter and Van Horn, 1975: 463.

Although Van Meter and Van Horn (1975: 478-483) recognized that the model was "relatively complex", they argued it was capable of addressing three of the most common implementation problems: Communications, Capability (of implementing agencies) and Dispositional Conflicts. They concluded (p. 483) that the model "offers a blueprint for the description and analysis of the policy implementation process and that it proposes explanations for program achievements and failures."

This model was described as "the most comprehensive framework to date"

(Sabatier and Mazmanian, 1981: 5). Nevertheless, it suffered from some of the traditional defects of abstract system models:

1. Many of the factors are "essentially amorphous categories rather than variables that can easily be operationalized";
2. It does not identify which factors are controlled by different actors and is, "therefore, unlikely to be of much use to policy practitioners" (Sabatier and Mazmanian, 1981: 5; see also Gurnack and Harty, 1987: 368-9)

The second, probably most cited and criticized model of the top-down approach is that of Sabatier and Mazmanian (1981 and 1983). At the outset, it should be borne in mind that Sabatier and Mazmanian were primarily concerned with statutory policies and decisions. The dependent variable in their model is compliance of target groups and implementing officials (Browning *et al.*, 1981: 127).

The model divides "the generic factors affecting the implementation process" into three broad categories of independent variables:

- A) The Tractability of the Problem(s):
 1. Availability of valid technical theory and technology.
 2. Diversity of target-group behaviour.
 3. Target group as a percentage of the population.
 4. Extent of behavioral change required.
- B) Ability of Statute to Structure Implementation:
 1. Clear and consistent objectives.
 2. Incorporation of adequate causal theory.
 3. Financial resources.
 4. Hierarchical integration with and among implementation institutions.
 5. Decision-rules of implementing agencies.
 6. Recruitment of implementing officials.
 7. Formal access by outsiders.

C) Nonstatutory Variables Affecting Implementation:

1. Socio-economic conditions and technology.
2. Media attention to the problem.
3. Public support.
4. Attitudes and resources of constituency groups.
5. Support from sovereigns.
6. Commitment and leadership skill of implementing officials (Sabatier and Mazmanian, 1981).

Within this framework, and following the passage of the 'basic statute', implementation process was viewed in terms of five stages (see Fig 6.2):

1. Policy outputs of implementing agencies (decisions).
2. Compliance of target groups with those decisions.
3. Actual impacts of policy outputs.
4. Perceived impacts of policy outputs.
5. Major revisions (or attempted revisions) of the basic statute.

The last stage is to feedback to the first forming the "feedback loop" (Sabatier and Mazmanian, 1981: 7, 20-24). In an application of the framework, this delineation of the implementation phases was considered "particularly useful"; it "allows the policy analyst to flesh out realistic phases in the development of ... [the] case in greater detail" (Gurnack and Harty, 1987: 393). Conceptualization of implementation as a process over time provides a framework which permits further precision in diagnosing the success or failure of implementation effort.

Those variables were then synthesized into a list of "six sufficient and generally necessary conditions for the effective implementation of *legal objectives*" (Sabatier and Mazmanian, 1981: 23, stress added):

- (1) Clear and consistent objectives...
- (2) Adequate causal theory...

- (3) Implementation process legally structured to enhance compliance by implementing officials and target groups...
- (4) Committed and skilful implementing officials...
- (5) Support of interest groups and sovereigns...
- (6) Changes in socio-economic conditions which do not substantially undermine political support of causal theory. (Sabatier and Mazmanian, 1981: 23-5)

Several of these conditions have been the subject of criticism, especially the first one. Thrasher and Dunkerley (1982: 351) recalled that "a frequent criticism of urban policies has been the failure to specify exactly what these measures seek to achieve." Hambleton (1983: 407) argued that "there are limits to the degree of clarity we can expect in the policy message - that there are, in fact, sound reasons for policy ambiguity" (see also Barrett and Fudge, 1981: 18-9). Sabatier himself later admitted that "the emphasis ... placed on 'clear and consistent objectives' was a mistake. Experience has confirmed the critics' charge that very few programs meet this criterion, either initially or after a decade" (Sabatier, 1986: 29).

As for the third condition, the analysis of the EDA (Economic Development Administration) project in Oakland - California, revealed that: "Given that the span of influences on implementation can never be entirely preconceived, the actual implementation process will always be less structured than the expected process" (Browne and Wildavsky, 1984: 229).

In general terms, "the sort of 'conditions' ... prescribed are precisely those which empirical evidence suggests are not met in the real world" (Barrett and Fudge, 1981: 18). Although they found "none better in the literature" than this

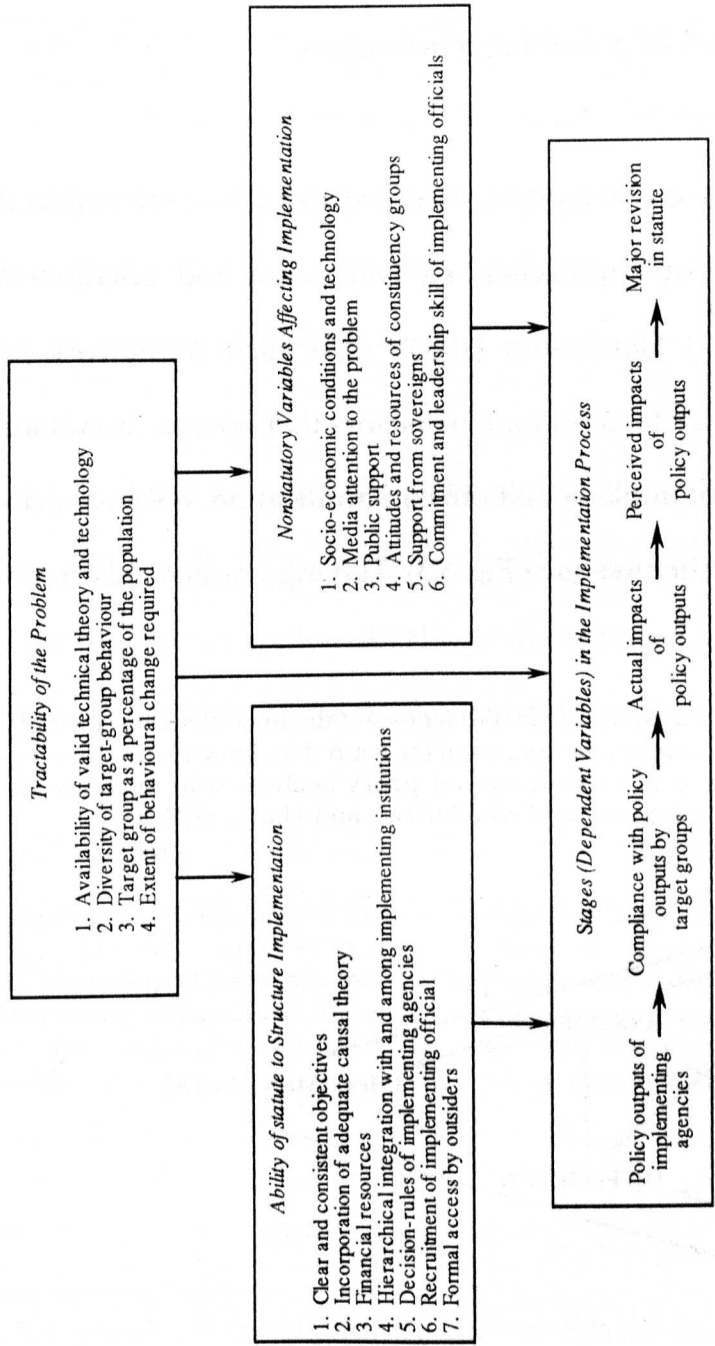


Fig 6.2: Skeletal Flow Diagram of the Variables Involved in the Implementation Process
 Source: Sabatier and Mazmanian, 1981: 7.

model, Browne and Wildavsky (1984: 231-2) maintained that "like similar statements, [it] turns the subject matter around." The objectives of implementation literature "have become its preconditions. Instead of ending up with more effective implementation, it is apparently deemed necessary to begin with it. Implementation analysts, therefore, work toward influencing policy design instead of policy implementation."

In addition to these two models, there are others that fall within the broad category of top-down approaches, eg Nakamura and Smallwood (1980), Edwards (1980) and Hambleton (1983). One such framework is that of Edwards (1980) which focused on four factors that operate simultaneously to impact the process of implementation: communication, resources, disposition and bureaucratic structures (see Fig 6.3). Having applied this framework, it was concluded:

This model provides a rapid, succinct diagnosis of the ingredients for the policy success. ... One could have predicted the success of the court decision with this model [which was the case]. Therefore, it is a useful tool for policy analysts with regard to the ease and facility with which it may be applied (Gurnack and Harty, 1987: 386)

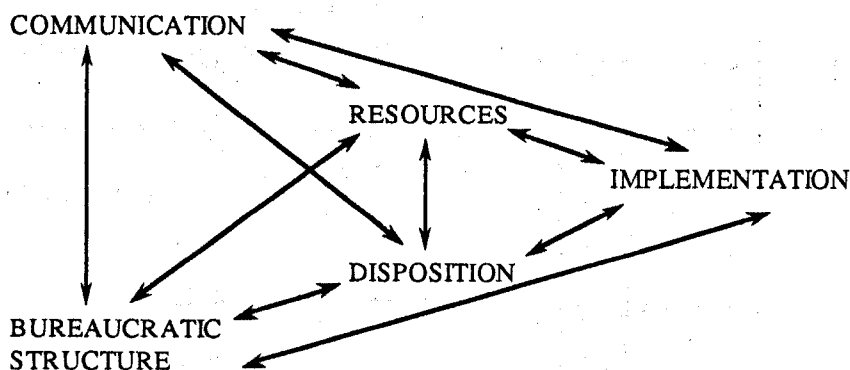


Fig 6.3: Direct and Indirect Impacts on Implementation

Source: Gurnack and Harty, 1987: 366.

6.3.2 The Pros and Cons of Top-Down Models

The list of criticism of the top-down approach is undeniably a long one, especially when compared to the 'rival' bottom-up approach⁽¹²⁾. This, however, must not lead to the conclusion that the latter is 'better' than the former. "A "bottom-up" perspective is not an exclusive approach - indeed there is a valid argument to be made for a variety of perspectives" Thrasher and Dunkerley, 1982: 372-3). Although, as will be discussed later, top-down models appear "at odds with a bottom-up approach ..., it might well be more rewarding to regard them as partially complementary rather than mutually exclusive" (Wittrock *et al.*, 1982: 134).

It should be borne in mind that reference to the model of Sabatier and Mazmanian (1981 and 1983) is made merely as a 'representative' of the approach; the arguments apply with almost the same strength to other models that adopt the same view of the implementation process.

Criticism	Reply/Comment
1. The assumption that policy-makers are in command of the organizational, political and technological processes that affect implementation when clearly this is not always the case (Thrasher and Dunkerley, 1982: 351).	Policy-makers have the ability, to a certain extent, to influence the implementation process. For instance, they can select one implementing agency over another or, they can base the policy on valid causal theory (Sabatier, 1986: 25).

⁽¹²⁾ This is due, perhaps, to the fact that top-down models have been applied more than bottom-up ones. Sabatier (1986: 26) listed twenty-one applications of the approach of himself and Mazmanian, whereas the applications of bottom-up models that he could trace did not exceed six.

2. The approach seems to make a distinction between policy formulation and policy implementation. "Certainly, this separation would meet with universal approval if only because it satisfies general notions of democracy and accountability. Reality is somewhat different from this position" (Thrasher and Dunkerley, 1982: 351-2).

3. It is least valuable in explaining why differences and disjunctions between policy intentions and actual outcome occur (Thrasher and Dunkerley, 1982: 373)

4. Top-down approaches place questions of steering and compliance "at the centre of attention" (Hanf, 1982: 160). This precludes consideration of a strong element of 'local presence' evident in many public programmes.

5. Centre/Periphery Relation:

a) "implementation is frequently a more differentiated process - with more relevant actors and policy activities, and more complex behaviours - than is allowed for in the top-down view of things" (Hanf, 1982: 160).

b) It is "the policy-makers' perspective"; "It assumes that policy comes from the top and is the starting point for implementation and action. This, ..., is not necessarily the case: policy may be a response to pressures and problems experienced on the ground. Equally, policy may be developed from specific innovations, that is, action precedes policy" (Barrett and Fudge, 1981: 12).

c) The assumption that "implementers are agents for policy-makers and therefore in a *compliant* relationship to policy-makers" is not always true (Barrett and Fudge, 1981: 12; see also Hanf, 1982: 160-1).

Sabatier and Mazmanian (1983: 146) argued for a division of the policy-making process into three basic stages: formulation, implementation and reformulation. This division "directs attention to this area of inquiry by focusing on the extent to which the legislature (...) modifies its original policy as a result of the implementation experience."

Actually, Sabatier and Mazmanian's framework directs attention, explicitly, to at least some of the factors that may explain such disjunctions when they occur. For instance, adequacy of resources, socio-economic conditions and public support.

Sabatier (1986) admitted that the approach's focus on the Centre's (policy-makers') perspective and its tendency to ignore other actors' perspective(s) is a 'significant flaw'. It is "certainly a potential Achilles' heel of their [Sabatier and Mazmanian] model."

6. Sabatier (1986: 29-30) also acknowledged that the model has other significant flaws:
 - a) It "did not provide a good conceptual vehicle for looking at policy change over periods of a decade or more" although it did encourage a longer time-frame for analysis than its predecessors;
 - b) Top-down models "are difficult to use in situations where there is no dominant policy (statute) or agency, but rather a multitude of governmental directives and actors, none of them preeminent."
7. Browning *et al.* (1981:129) added:
 - a) "nonstatutory local variables are frequently stronger and more difficult to overcome than Sabatier and Mazmanian imply and are not adequately treated in their conceptualization."
 - b) "In addition, when the programs are social, a valid technical theory, clear objectives, and adequate financial resources are also lacking, and these variables are often not readily controllable through statutory design."
 - c) "the focus on compliance ignores interesting questions about the long term effects of federal [central government] programs".

Notwithstanding these criticisms, the top-down perspective has its particular strengths:

1. It "has proved to be helpful in tracking the transition from policy initiation to policy output and finally to its impact" (Thrasher and Dunkerley, 1982: 352).
2. It is "most valuable in measuring the disjunction between policy intentions and actual outcomes" (Thrasher and Dunkerley, 1982: 372).
3. From Sabatier's point of view (Sabatier, 1986: 27-8), the framework has the following strengths:
 - a) "the importance it attaches to legal structuring of the implementation process - one of its major innovations - has been confirmed in numerous studies. ... The evidence suggests that, while fairly coherent structuring is difficult, it occurs more frequently than critics realize and, when present, proves to be very important."
 - b) "the six conditions of effective implementation [see page 155] have proven to be a useful checklist of critical factors in understanding the strategies of program proponents over time."
 - c) "the relatively manageable list of variables and the focus in the framework on the formulation-implementation-reformulation cycle encouraged many ... to look at a longer time-frame than was true of earlier implementation studies (i.e. ten years instead of four). This, in turn, led to the discovery of the importance of learning by program proponents over time"

- d) The "focus on legally-mandated objectives - particularly when combined with the ten-year time span for assessing program effectiveness - helped produce a less pessimistic evaluation of governmental performance than was true of the first generation of implementation studies."
4. The framework of Sabatier and Mazmanian is "more comprehensive and less trivially common-sense" and "more relevant" than some of the other models "since it does devote some (though not enough) attention to an aspect of implementation of which planners have traditionally been keenly aware: the relationship between the structure of the plan or policy itself and the implementation process" (Alterman, 1983: 64).

6.3.3 Bottom-Up Approaches

The bottom-up perspective starts with an "analysis of the multitude of actors who interact at the operational (local) level on a particular problem or issue" (Sabatier, 1986: 22). That is, the perspective takes what is done as central and attempts to find out and understand why various actors (individuals, groups) act the way they do (Hambleton, 1983: 405). The stress in this approach is, thus, on the "importance of the lower ranks of bureaucrats and locally based organizations in dealing with the policies which are handed down from above" (Davies and Mason, 1982: 147-8).

Accordingly, "[t]he crucial distinction between the 'bottom-up' and 'top-down' approaches is as methods for reconstructing who did what in generating programme outputs" (Hanf, 1982: 170). It should not be, however, a question of choosing between the two "as though these were mutually exclusive alternatives" (Hanf, 1982: 171). Researchers, in both streams, "have been motivated by somewhat different concerns and thus have developed different approaches" (Sabatier, 1986: 35).

The development of the approach dates back to the late 1970s. In their analysis of the implementation of manpower training programmes in Sweden and Germany, Hanf *et al.* (1978) found that the implementation process involved a multitude of actors at different levels (local, regional and national) with a fairly complex network of interactions. They concluded that the programme success was far more dependent upon the skills of specific actors than upon the efforts of central government officials. Hence, in contrast to the top-down approach, the focus is no longer on the extent to which central objectives are being achieved. Rather, it is the strategies pursued by various implementing actors in pursuit of their own objectives (Sabatier, 1986: 22 and Hambleton, 1983: 405).

Implementation Structures

The approach (as developed by Hanf *et al.*, 1978; see also Hanf, 1982 and Hull and Hjern, 1982) "involves a 'snowballing' procedure of interviewing" (Hull and Hjern, 1982: 189). The target 'actors' (individuals, groups, firms, ...) are interviewed first, then those named as contacts and further out to identify the implementation networks (Hull and Hjern, 1982: 189). Hjern and Porter (1981) further developed the approach, laying the theoretical basis of "a new unit of administrative analysis", the so-called 'Implementation Structures'. This approach is indeed the most common among the bottom-up models. The two terms, bottom-up and implementation structures, are sometimes used interchangeably. Sabatier (1986: ff. 44) selected it as a representative of the "bottom-uppers", over others, "because of Hjern's superior methodology."

The point of departure for Hjern and Porter's work (1981) was their observation that: "One of the most pervasive findings in studies of programme implementation is that many actors are involved in what outwardly appears to be a confounding set of relationships" (p. 213; the multi-organizational nature of policy-making, see chapter 5, p. 128). Their argument was that there is a need for a framework to analyze such situations where "*parts of many public and private organizations cooperate in the implementation of a programme*" (p. 214, stress in original). The intention of the approach, thus, was "to capture the set of individual actors (and their interactions) involved in a particular functional activity with regard to the programme imperatives in given regions" (Hanf, 1982: 160).

The approach starts with an analysis of the 'administrative imperatives' of a programme. Through this analysis, it becomes possible to identify the 'pools of organizations' within the environment of that programme. Various 'members' of these pools take part in the implementation process. "With a little imagination it is easy to name the main corporate actors in the pool of organizations which will potentially be involved in implementing a ... programme" (Hjern and Porter, 1981: 215). In brief:

An analysis of the objectives of a *programme* suggests an *administrative imperative*. This imperative points to a potential *pool* of organizations, from which an *implementation structure* is formed (Hjern and Porter, 1981: 222, stress in original).

Implementation structures are comprised of parts of many organizations which, in turn, are comprised of many programmes, ie implementation

structures are not organizations. They are "more likely to be *self-selected* than *designed* through authoritative relationships" (Hjern and Porter, 1981: 216, stress in original). There are differences in the way decisions are made in both organizations and implementation structures. The major ones are that in implementation structures:

- there is less formal structure and fewer authoritative relations;
- the social structures which exist are more dynamic and shifting; and
- decisions to participate in a programme are 'fuzzy', based on consent and negotiation. (Hjern and Porter, 1981: 216)

Thrasher and Dunkerley (1982) elaborated on the same approach, with a different theoretical conceptualization. They argued in favour of using the 'Social Exchange' theory in implementation analysis. The focus of their 'model' was on individual implementers, their resources and the exchange of these resources. This framework was intended to complement existing implementation studies not to replace them. Although it appears useful in understanding interactions between individuals, its very detailed nature is its defect. Numerous individuals take part in the implementation process of public policies. This focus on individuals will, thus, require substantial resources and time. It will involve great difficulties in collecting, handling and analysing the vast amounts of data that will be needed.

6.3.4 The Pros and Cons of Bottom-Up Models

The bottom-up approach is not without its limitations. Most notable among these are:

1. "Looking at the bottom, ..., gives one rather a hopeless feeling" (Davies and Mason, 1982: 156). This is perhaps due to the sheer number of actors involved in the process and the very complex networks of interactions between them (Davies and Mason, 1982: 156).
2. The approach is difficult to apply in a context where economic (market) forces are dominant (Davies and Mason, 1982: 156).
3. The bottom-up approaches are in danger of overemphasising "the ability of the Periphery to frustrate the Center" (Sabatier, 1986: 34-6). The "political situation surrounding the formulation of programmes serves to pre-structure the composition of implementation structure and to influence the content of decision-making process within them" (Ackermann and Steinmann, 1982: 184).
- 4.. From a 'top-downer' point of view, Sabatier (1986: 34-6) advanced the following criticism (taking Hanf *et al.* (1978) approach as a representative of bottom-up approaches):
 - a) "one of the most basic shortcomings of the ... approach is that it takes the present distribution of preferences and resources as given, without ever inquiring into the efforts of other actors to structure the rules of the game."
 - b) The approach takes "the present participation in an implementation structure as given without examining the prior efforts of various individuals to affect the participation rates."
 - c) A more fundamental limitation of the approach is its "failure to start from an explicit theory of the factors affecting its subject of interest. Because it relies very heavily on the perceptions and activities of participants, it is their prisoner - and therefore unlikely to analyze the factors *indirectly* affecting their behaviour".

However, the approach has its strengths

1. It provides a solution to the problem of comparability across different constitutional contexts. "By focusing on the phenomenologically delimited implementation structure - ... - it is possible to construct comparable functional units in the different geographical settings under investigation" (Hanf, 1982: 171)
2. It provides a neutral mapping of who participates to what effect in the implementation process (Hanf, 1982: 171).
3. Sabatier (1986: 33-34) admitted that the approach has some "notable strengths":
 - a) It provides an explicit and replicable methodology for identifying a policy network ('implementation structure');
 - b) Because it does not begin with a governmental programme but rather with actors' perceived problems and the strategies developed for dealing with them, it is able to assess the relative importance of a variety of governmental programmes vis-a-vis private organizations and market forces in solving those problems. "In contrast, a

top-down approach is likely to overestimate the importance of the governmental program which is its focus."

- c) As it does not start with a focus on the attainment of formal policy objectives, it is free to see all sorts of (unintended) consequences of governmental and private programs;
- d) The approach "is able to deal with a policy/problem area involving a multitude of public (and private) programs, none of them preeminent. In contrast, such cases represent substantial difficulties for top-down-approaches."
- e) "because of their focus on the strategies pursued by a wide range of actors, bottom-uppers are better able to deal with strategic interaction over time than top-downers - who tend to focus on the strategies of program proponents, while neglecting those of other actors."

6.3.5 Synthesis of Approaches

Having reviewed both the top-down and bottom-up approaches, Sabatier (1986: 38-42) sought to produce a synthesis that "explicitly attempts to develop such a general model of the policy process which combines the best features" of the two. This proposed approach had the following elements:

- a) It takes, as its starting unit, the bottom-up unit of analysis: various actors and their perspectives and strategies.
- b) This is then combined with top-downers concern about the manner in which socio-economic conditions and legal instruments constrain behaviour. Policy change is analyzed over a period of a decade or more.
- c) "Finally, the synthesis adopts the intellectual style (or methodological perspective) of many top-downers in its willingness to utilize fairly abstracted constructs and to operate from an admittedly simplified 'portrait of reality' Sabatier (1986: 39) (see Fig 6.4).

Drawing from the top-down models, the framework recognizes two sets of exogenous variables which affect the constraints and resources of sub-system actors and 'structure' policy-making. "Within the subsystem, the

framework draws heavily upon the bottom-up approach" (Sabatier, 1986: 40); its recognition of the existence and influence of the multitude of actors. However, it is assumed that those actors can be aggregated into a number of 'advocacy coalitions'. It is also assumed that at any particular time, each coalition adopts a strategy which is perceived to further its (the coalition's) objectives. 'Policy Brokers' is a third group of actors who mediate conflicting strategies. The result is a compromise leading to an action programme. This, in turn, produces policy outputs and, ultimately, impacts.

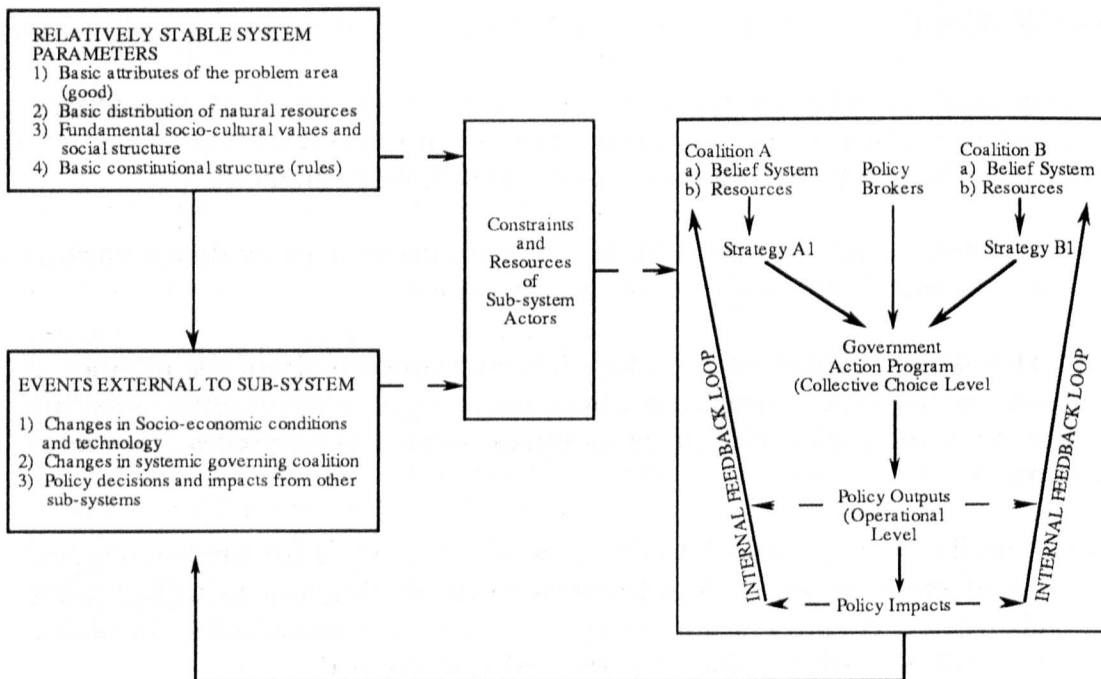


Fig 6.4: General Overview of Conceptual Framework of Policy Change

Source: Sabatier, 1986: 41.

6.4 THE CHOICE OF A MODEL

If evaluation was to fulfil its prime function and inform policy-makers, in order that better informed and more rational decisions are made, two crucial questions need to be answered: "what happened?" and, "why did it happen this way?". That is, what is the impact of the policy and has it achieved its objectives and, what were the influential factors behind this achievement, or otherwise? In other words, 'comprehensive' evaluation should combine both outcome/impact evaluation *and* implementation analysis. The two concepts thus become closely related to each other; yet, the distinction between the focus of each and the tools it requires should always be borne in mind.

Unlike outcome evaluation, there seems to be a need to reiterate the reasons why implementation analysis should be an integral component of policy evaluation. As mentioned above, failure to examine how a policy is being/has been implemented is a threat to the four principal dimensions of validity: internal, construct, statistical conclusion and external (see page 148). In brief, failure to examine the policy process both casts considerable doubts on the results of outcome evaluation and removes the possibilities of learning from past experience. The sheer cost of "wrong" public decisions is believed to more than cost-justify the concern with the reasons behind a policy's success or failure.

The question to ask, then, is what model should be adopted for the analysis: top-down, bottom-up or the synthesis? Unfortunately, and since

circumstances differ widely from one case to the other, there can be no one single answer to this question. The 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) left the question about the model to adopt unanswered, though apparently for a different reason. "The rudimentary state of the art of implementation process evaluation does not as yet provide ... a set of distinct and tested methods" to answer the questions implementation analysis asks (Alterman *et al.*, 1984: 383-384).

While we agree that "each implementation setup and process is unique" (Alterman *et al.*, 1984: 383), we seem to be in a better position to answer this question. As has been shown in the previous sections of this chapter, analytical models have long been established and applied in numerous cases. The strengths and weaknesses of these models have also been long recognised, and attempts for synthesis have been made. The state-of-the-art of the field is no longer rudimentary.

The focus of implementation analysis is on the political and administrative context within which the planning and implementation process takes place. To choose an analytical model, it seems very plausible therefore to take this context as the starting point. As discussed in chapter 2 (section 2.3), there is a number of planning and policy-making modes, or styles (Hill, 1985a):

1. Command/Central Planning
2. Policy Planning
3. Corporate/Transactive Planning
4. Participatory Planning

Within each of these modes, a particular style of implementation is likely to prevail. Table (6.1) summarises both the planning process within each of those four policy-making modes and the form of implementation each entails.

The table shows that it is only within the first mode (central) that 'compliance' of target groups becomes the dominant variable. In other words, it is only within the command/central mode of policy-making that implementation will veer towards the top-down model. In each of the three other policy modes, implementation can be in either of the other models: bottom-up, adaptive or evolutionary; there is no way to predict which of the three models is likely to dominate. However, it has also to be borne in mind that, although in any given situation one policy mode may be dominant, it is very likely that it will be supplemented by another one, or more (Hill, 1985a: 13). Consequently, there will be a hybrid of planning and implementation modes. For reasons of simplicity, and because hybrids differ widely from one case to the other, it will be assumed for the moment that each mode can be treated separately.

Given that assumption, the characteristic features and focus of both the top-down and bottom-up approaches and, the attributes of each policy mode, the following preliminary conclusions emerge:

1. The key variable in the control/central mode (the two types) is compliance of lower-level agencies and target groups. There is a dominant agency with a dominant policy. Such situations appear ideal to apply the top-down perspective of implementation analysis.

2. The situation is entirely different in the three other modes. To a varying degree, implementing the policy is dependent on the lower-level agencies and target groups; there is no dominant agency. Consequently, analysis should focus on the various actors and their interactions. Since this is the focus, and strength, of the bottom-up models, they seem more appropriate in these situations.

Table 6.1: Decision-Making and Implementation Modes

Decision-Making/Planning Mode	Implementation Mode
<p>1. Command/Central: In situations where power is highly centralized. Control is largely sought by means of sanctions.</p> <p><u>1.1. Command-regulative</u> planning is undertaken by authorized governmental agency(s). This central body has statutory control.</p> <p><u>1.2. Command-initiatory</u> A public, or private, agency undertakes the task of planning. This agency has the resources to implement its plan. Control is thus achieved by budget allocation.</p>	<p>Implementation is viewed 'first and foremost' as a process of compliance (by both governmental agencies and target groups).</p> <p>"The agency in question initiates the process, allocates the resources, and has considerable control over timing" (Alterman, 1981: 17).</p>
<p>2. Policy Planning: A characteristic of "weakly centralized governmental systems" (Hill, 1985a: 13). Planning is "formulated as statements about what the agency considers desirable" (Alterman, 1981: 18). Control is sought through guidelines and decision rules and criteria for lower-levels with no ability of enforcement; material incentives may be used to encourage compliance (Hill, 1985a: 13).</p>	<p>"Full compliance is not to be expected, and failure to comply in full should not be regarded as failure of implementation" (Alterman, 1981: 18).</p>
<p>3. Corporate/Transactive: Power is "decentralized among a small number of corporate bodies" (Hill, 1985a: 13). Planning is a "process of give and take among semi-autonomous groups" (Alterman, 1981: 18). "Control is achieved by means of a normative compliance with agreements reached through bargaining and negotiation" (Hill, 1985a: 13).</p>	<p>There is no single trajectory of the implementation process. It is relative to the point(s) of view of each of the participating groups (Alterman, 1981: 18).</p>
<p>4. Participatory Planning: Power is dispersed among many actors; voluntary compliance is the predominant method of 'control' (Hill, 1985a: 13).</p>	<p>Implementation is a process of interactions between the participants. Full compliance may be expected only from participants who agree with the policy's objectives.</p>

Source: compiled by the author drawing on material from Hill (1985a) and Alterman (1981).

However, exceptions to this pattern are not unknown. Individuals, in a central policy mode, may be willing to undergo sanctions not to comply with

what they perceive as unjust or unfair constraints. On the other hand, when target groups share the same objectives among themselves and with the central policy, intact compliance can be expected in any of the other modes. In addition, policy-makers have the ability, even if to an extent, to structure the implementation process (Sabatier, 1986). Ignoring target groups' objectives, in the first instance, is as much a mistake as to ignore the centre's in the second. In other words, attention should be paid to both ends of the scale: Centre and Periphery. Just where exactly should the balance be is a seemingly unanswerable question. The mixture of planning and implementation modes, socio-economic conditions and a multitude of other factors differ from one case to another. There are bound to be several other policies at various levels that affect both the implementation process and policy outputs. Although the policy may be initiated by one central agency, a host of agencies, public and private, will almost certainly take part in the implementation process.

In sum, the recommendations advanced earlier are deemed inappropriate. To focus on one end of the scale is a mistake. It is believed a mistake to ignore all the other forces and factors that influence and shape the environment within which the policy of concern operates.

This leads us to the synthesis of approaches (top-down and bottom-up) proposed by Sabatier (1986). As cited before, the proposal combines the focus of the bottom-up models on target groups (the Periphery) with the top-down concern about the way socio-economic conditions and legal instruments

constraint behaviour. The question, however, remains, whether emphasis should be placed on either of the two sides, and if so, which side and how much? Unfortunately, there is no one simple answer to that question; and that is, in fact, as it should be. Each case is different from the others; what is appropriate in one case is not in another.

However, there remains a need for at least a set of guidelines to help analysts decide on which way to go. A logical starting point seems to determine the dominant policy-making mode. This will indicate, in later stages of the analysis, where emphasis should be placed. Borrowing from the implementation structures approach, and since we are dealing with ex-post evaluation, it should not be difficult to identify all parties and actors involved, and interested, in the implementation process. If the control/central mode of policy-making is dominant, more attention is to be paid to the Centre and its objectives and the organizational structure(s) through which implementation is undertaken (top-down perspective). That is, attention will still be paid, though to a lesser degree, to the lower-level agencies and target groups (the Periphery). The available models of the top-down approach offer a 'pool' from which a 'checklist' of the factors to be included in the analysis can be drawn. The seventeen elements of Sabatier and Mazmanian's model, for instance, serve as a good starting point (see page 154).

If, on the other hand, any of the other three modes was dominant, the situation is to be reversed; more attention is to be given to the Periphery than

the Centre. In such cases, the bottom-up approach with its 'snowballing' technique of interviews, is to be adopted. In addition, the influence of the Centre (and if any, the central policy) on structuring the process and the participation rates should be taken into account. Finally, and since any policy is not operating in a vacuum, the impacts of both the socio-economic conditions and other policies, and change in any or both, should be considered as part of the analysis.

The apparent problem with such a wide-scope approach of analysis is that it will almost certainly require a long period of time to be carried out. In fact, the wider the scope of the policy, or the longer it has been in operation, the longer the time required for implementation analysis becomes. Criticism of delay and irrelevancy will be mounted from almost every interested party in the evaluation. Unfortunately, there is no best way out of such a dilemma. Perhaps the only resolution lies within monitoring. As a regular, systematic exercise, an extended view of monitoring can bridge that gap by providing information, on a regular basis, with regard to the most 'urgent' aspects of the implementation process (see chapter 5).

PART 2:

CONTEXT-DERIVED METHODS

INTRODUCTION TO PART 2

Part 2 focuses on 'context-derived' evaluation methods. Despite the severe lack of research on a systematic approach to ex-post evaluation, numerous case studies have been, and are being, undertaken in various fields. The question then arises: How is the task being approached in practice, and why? What are the methodologies used; what are the problems encountered and how have they been overcome, if at all? Part 2 is an attempt to find answers to these questions. Given the main objective of this research (to develop a systematic approach for comprehensive ex-post evaluation), the need for these answers should be self-evident. To put it rather briefly, Part 2 aims to find out the strengths to be adopted, the weaknesses to be avoided and the problems to be overcome.

These findings are supplementary to and as important as those of Part 1. The proposed framework should, and will, build on theory as much as it will build upon practical experience. Ignoring this experience carries the inherent danger of putting forward an 'impractical' framework; one that founders on practical difficulties that have been overlooked. It also carries the risk of overlooking 'good practice'; practice that has overcome difficulties or at least pointed out the directions for future research.

Part 2 starts with the first hand experience in urban policy evaluation. Chapter 7 follows the experience of the first bidding round of the Single Regeneration Budget from the appraisal of bids to the start of their implementation and the first quarterly assessment of progress and payment of grant. The chapter is based, in the main, on information and material obtained while working at one of the Government Offices for the Regions during the appraisal process (September - October, 1994) and maintained contact with the GO since then. The importance of this experience lies mostly in exposing the constraints imposed on evaluation research.

In chapter 8 we move closer to the focus of this part. The chapter is a review of three studies of the most recent research in the field of urban policy evaluation. The

first case study (evaluating the Programme for the Valleys) claimed to have dealt with the problem of 'vague objective statement'. The second (Assessing the Impact of Urban Policy) is perhaps the most ambitious and comprehensive research to date. City Challenge, the subject of the third case study, marked a clear shift in urban initiatives and brought about several innovative features that are likely to underline future policies. The first and third case studies have each benefited from meetings with members of the evaluation teams, and in the third case a senior member of a City Challenge team.

The review of these studies reveals a lack of a systematic approach to urban policy ex-post evaluation and a host of conceptual and practical difficulties. The question is then raised whether similar problems have been encountered in evaluation research in other fields, and if so, how they have been overcome, if at all. In other words, does experience in other fields offer any 'transferable' lessons into urban policy evaluation? This question is at the heart of chapters 9 and 10.

Chapter 9 is a review of the experience of assessing the impact of regional (economic) policy in Britain. Chapter 10 is a review of the appraisal and evaluation of trunk road and motorway schemes in the UK, USA and the Netherlands. In addition to information gained from the literature, this chapter benefits from contacts with key figures in the fields of transportation and planning on both sides of the Atlantic. The information on the Dutch experience was gained through a field trip (May - June, 1995). In addition to meetings with senior officers at the Dutch Ministry of Transport, valuable material was also obtained.

Each of the four chapters is a free-standing one. However, the underlying structure of them all is as follows:

1. A brief account of the policy at hand;
2. A summary of the evaluation methodology adopted; and
3. A critical review of this methodology.

The review of the methodology in each case study is the prime focus of each chapter and centres around the following questions:

1. **Measurement of impact:** How were the outcomes of the policy measured? What indicators were used? Were these indicators the 'most appropriate'? Were there any constraints on the development and/or use of these

indicators or others? Were outcomes measured in aggregate form or were the distributional effects taken into account?

2. **The treatment of the counter-factual problem:** What was the approach adopted to treat the counter-factual situation, if at all? What are the advantages and disadvantages of this approach?
3. **Assessing the achievement of objectives:** Were measured outputs/outcomes related to policy objectives? How was the achievement of objectives assessed, if at all?
4. **Explanation of the results:** Was an explanation of the results provided? On what ground was the achievement, or otherwise, of objectives explained? Was 'implementation analysis' an element of the evaluation methodology?

**CHAPTER 7:
THE SINGLE REGENERATION
BUDGET**

7.1 INTRODUCTION

In April 1994, the British Government launched its latest urban regeneration initiative: the Single Regeneration Budget (SRB). The objective is to achieve "sustainable local development" in England through partnership between the public and private sectors. The initiative is intended to bring together the various regeneration programmes currently in operation under one umbrella aiming at greater value for money. The Budget is administered on a bidding basis by Regional Offices that were restructured, also in April 1994, to combine the activities and responsibilities of four Government Departments.

This chapter follows the experience of the first bidding round of the Budget from the appraisal of various bids to the start of their implementation and the first payment of grant. It is based, in the main, on insights and information gained through working at one of the newly restructured Government Offices (GOs) during the appraisal process (September-October, 1994) and maintained contacts with the GO throughout the following stages of the process. This first hand experience has, no doubt, significant reflections on evaluation research, both ex-ante and ex-post. In an ex-ante context, the appraisal process raises an important question about the applicability of established methods such as CBA and GAM. It reveals a mixture of encouraging signs and potential difficulties for any future ex-post evaluation

of the SRB. In both contexts, the appraisal process exposes several of the constraints imposed on evaluation research. These are the issues this chapter sets out to examine. In so doing, the focus shifts more towards urban policy evaluation in general. The reflections of the first bidding round on the SRB initiative have been examined elsewhere (Bakr, 1994).

The chapter starts with a review of the SRB initiative, its underlying concepts and objectives and the guidelines (published and unpublished) for assessment of individual bids. Section 3 is a detailed account of the workings of the GO through the appraisal process and, more importantly, the constraints and influential factors that shaped the process and its outcomes. These are grouped under three main headings: the bids submitted, political constraints and information and time constraints, of which political considerations are believed to have been the most influential.

Section 4 covers the next phases of the process, following the submission of regional packages to the DoE Headquarters (DoE HQ). Prior to, and after, the announcement of the successful bids, DoE HQ and GOs were engaged in an almost continuous dialogue for various reasons that are explained. The final decision was announced before Christmas 1994. Shortly after, further guidance on the management of successful bids and their monitoring and evaluation was issued. Successful bidders prepared their Delivery Plans and many started implementation, as initially planned, around April 1995. In July 1995, many of those bidders submitted their first progress report and claim

for grant. The assessment of these reports and claims represented the first quarterly monitoring task which took place in early August 1995. This is covered in brief at the end of this section.

The final section is devoted to the main purpose of this chapter: to examine the issues the appraisal process raised in regard to urban policy evaluation, in addition to revealing several of the constraints usually imposed on the working of evaluators. As no evaluation has yet taken place of either the SRB or any of its schemes, the conclusions on such a topic are no doubt speculative. Nevertheless, it is argued that there are as much encouraging signs as potential problems in the evaluation of the SRB. It is also argued that an equal emphasis must be placed on 'process evaluation' and that the focus on quantified outputs/outcomes alone may indeed be misleading. These conclusions are informed by the first-hand experience of working at the GO, contacts with the GO, interviews with several practitioners in the field and the review of other case studies (see chapters 8, 9 and 10).

7.2 THE SINGLE REGENERATION BUDGET⁽¹³⁾

7.2.1 The Initiative

The Single Regeneration Budget (SRB) came into operation in April 1994, underlined with a governmental wish to encourage "sustainable local development" in England (DoE, 1994a: 1). Exactly what is meant by

⁽¹³⁾ This section draws on the Bidding Guidance (DoE, 1994a and 1995e) and information from the GO, both written material and the outcome of interviews and discussions with members of staff.

"sustainable development" remains, however, unclear. The SRB is operated through a unified network of Government Offices (for the Regions). These GOs were themselves the result of a restructuring process whereby the regional establishments of four departments - Transport, Trade and Industry, Environment and Employment - have been combined together. A new Ministerial Committee (known as EDR) has also been set up to coordinate regeneration policy among the various departments.

The Budget is a fund of public money which is intended to complement or attract private sector and other resources. Concern to maximise the leverage of private sector investment and intensify the impact of public expenditure is a significant aspect of the initiative. Individual bids are also expected to harness the talents and resources of the voluntary sector. In its first year, 1994/95, the Budget is said to have promoted regeneration through established programmes (eg City Challenge, Urban Development Corporations, ...). The commitments of existing programmes will be met from the Budget until they cease operation (see Table 7.1). This, however, limited the scope of the first bidding round of the Budget. Of an estimated £1.3 billion available in 1995/96, only some £100m were available for new projects (raised to £125m by mid-November, 1994); the bulk of the funding being directed to meeting the commitments of the existing programmes

Table 7.1: Programmes Contributing to the SRB, 1994/95 (£m)

From the Department of the Environment (DoE)	
Estate Action	373
Housing Action Trusts	88
City Challenge	213
Urban Programme	83
Urban Development Corporations	286
Inner City Task Forces	16
City Action Teams	1
English Partnerships	181
From the Employment Department (ED)	
Programme Development Fund	3
Education Business Partnership	2
Teacher Placement Service	3
Compacts	6
Business Start-Up Scheme	70
Local Initiative Fund	29
TEC Challenge	4
From the Home Office (HO)	
Safer Cities	4
Section 11 Grants (part)	60
Ethnic Minority Grant	5
Ethnic Minority Business Initiative	1
From the Department of trade and Industry (DTI)	
Regional Enterprise Grants	9
From the Department for Education	
Grants for Education Support and Training (part)	5
TOTAL	1441

Source: written information from the GO.

7.2.2 Objectives

Allowing for local circumstances, the Bidding Guidance set the main objectives of the Budget as to:

- enhance the employment prospects, education and skills of local people, particularly the young and those at a disadvantage, and promote equality of opportunity;
- encourage sustainable economic growth and wealth creation by improving the competitiveness of the local economy, including business support;
- improve housing through physical improvement, greater choice and better management and maintenance;

- promote initiatives of benefit to ethnic minorities;
- tackle crime and improve community safety;
- protect and improve the environment and infrastructure and promote good design;
- enhance the quality of life of local people, including their health and sports opportunities. (DoE, 1994a: 4-5)⁽¹⁴⁾

7.2.3 Partnerships

Aiming to encourage a joint approach to regeneration, bids are supposed to be supported by a partnership representing the appropriate range of interests. The existing partnerships involved in programmes such as City Challenge were seen as a potential basis for new initiatives. Local authorities and Training and Enterprise Councils (TECs) were considered to have a central role to play, alongside other partners, in submitting bids. Local authorities or TECs were expected to lead bids and act as convenors of partnership efforts.

7.2.4 Assessment of Bids

Guidelines as to the format, content and even the number of pages of the bid are also provided. In the 'initial' version of the Bidding Guidance (DoE, 1994a) a 'Bid Checklist' was also provided against which bids would be assessed. This list included a total of 30 questions to which partners should ensure that clear answers were provided. The questions were grouped under the following headings:

⁽¹⁴⁾ The objectives were re-arranged in the revised version of the Bidding Guidance (DoE, 1995e). Objective no. 6 (to protect and improve the environment ...) was ranked third; the other objectives were ranked accordingly.

1. Concept or vision
2. Impact
3. Quality of evidence
4. Need for Budget support
5. Risk and realism
6. Exit strategy
7. Pilot projects
8. Implementation and management
9. Monitoring and evaluation

A similar set of questions was also provided in the revised version of the Guidance (DoE, 1995e), though in a slightly different format. Regional Offices were urged to take account of the detailed information about local conditions based on the 1991 Index of Local Conditions (DoE, 1994b) and other data sources (eg Employment Services for information on labour markets). The Index of Local Conditions is a replacement for a series of urban deprivation indices from the 1981 population census (known as z-scores). The 1991 index combines a range of indicators (economic, social, housing and environmental) into a single deprivation score.

The monitoring and evaluation of bids was one of the issues emphasised in the Guidance. A list of output measures, grouped under relevant objectives, was given (see Appendix B). Bidders were also required to demonstrate that effective management and monitoring arrangements would be put into place. It was made clear that continuation of funding would depend on a satisfactory outcome of periodic, at least annual, performance reviews. Regional Offices could carry out additional reviews. It was also envisaged that the SRB as a whole could become the subject of an independent national evaluation.

7.2.5 Bidding and Decision-making Arrangements

The time-table of the first bidding round was initially set as follows (DoE, 1994a: 45):

1. 14 April 1994: Bidding round opens
2. May-June: Partners consultations; outline discussions with Regional Offices
3. 7 September: Submission of bids
4. Oct. - Nov.: Recommendations from GOs are made to Ministers
5. January (1995): Announcement of decision
6. April (1995): Start of implementation of approved bids.

The Secretary of State for the Environment is accountable to the Parliament for the Budget but is guided by the new Ministerial Committee on Regeneration (EDR) which will make the final decisions on bid selection.

7.2.6 Further Guidance

During the first bidding round, the SRB Steering Group (inter-departmental group - London) circulated several, unpublished, reports concerning both the appraisal of bids and the submission format (from GOs to DoE). Most, if not all, of these reports were in response to GOs' queries.

Although this may be interpreted as a lack of coverage in the Bidding Guidance, it seems inevitable, given the national scope of the initiative, that local clarification of the guidelines would be required. Local characteristics and needs differ considerably from one region to another. It would be unrealistic to expect a 'national-level' document to cover such a wide range of regional circumstances. Previous experience with other regeneration programmes could, however, have been put to more use in setting a

framework as how to choose between bids. In practice, the onus was left to GOs. The Bidding Guidance concentrated mainly on assessing individual bids. The Steering Group responses focused on the same issue in addition to the submission format (from GOs to DoE) but they also included further clarification of the decision-making process expected to take place following the submission. This amounts to an admission that there is more to the appraisal process than its technicalities. The 'Centre' is apparently aware of the diverse local political contexts of the initiative, which could be an influential factor over the outcome of the process (as shall be discussed later).

Of great importance to the appraisal of individual bids was the report circulated from the Steering Group to GOs around mid-August 1994. The report detailed the following:

1. Definition of the key output measures (of Annex C, the Bidding Guidance - DoE, 1994a),
2. The Summary Format of each bid, to be prepared by bidders for publication,
3. A tabulated format of the Bid Checklist to be completed for each bid by the Regional Offices,
4. Draft guidelines on Regional Offices Submission,
5. A revised version of the time table for handling bids.

This Checklist (see Fig. 7.1) contained a total of 43 'requirements' based on the Guidance. 39 of these 'criteria/requirements' were to be measured on a four-grade scale: A - very well, B - well, C - adequately and D - poorly. The other four criteria were 'yes/no' questions. The results of this check were then to be summarised in an 'Overall Assessment Sheet' leading to a 'Final Grade' of the bid (see Fig. 7.2).

Fig. 7.1: An Illustration of the SRB Bid Checklist

Requirement	Comments	A	B	C	D
Concept and/or vision					
1) A clear idea of the problems to be tackled or opportunities to be exploited					
Impact					
4) A viable assessment of the beneficial impact on the quality of life of local people and on sustainable local regeneration the project will bring					
Quality of evidence					
10) Robust evidence of problems, opportunities and needs					
Need for SRB support					
12) Maximisation of contributions from other resources					
Risk and realism					
16) realistic identification of risks and contingency plans					

Source: Unpublished report of the SRB Steering Group, August, 1994.

Fig. 7.2: The Overall SRB Bid Assessment Sheet

OVERALL ASSESSMENT	
Main Strengths of the bid	
Main Weaknesses of the bid	
Conclusion	
Recommended next steps	
Final Grade (A-D)	Date of assessment

Source: Unpublished report of the SRB Steering Group, August, 1994.

7.3 THE APPRAISAL PROCESS

7.3.1 Assessing Individual Bids

Both the 'Bid Checklist' (Fig. 7.1) and the 'Overall Assessment Sheet' (Fig. 7.2) were the basis on which individual bids were assessed. Copies of the submitted bids were circulated to 2-person teams, each representing either an involved department or other programmes with a stake in the decision-making process. Copies were also circulated to other parties, ie European Funding Programmes' Secretariats and, of course, the Regional Directors of the Office. However, not every bid was examined in detail by every member of the Working Group. The checklist (Fig. 7.1) was retained for the appraisal process by members for bids that fall within their departmental interest. Yet, every bid was considered against the 'Overall Assessment Sheet' (Fig. 7.2) by all members of the Group. At the Working Group meeting, each bid was introduced by a member of the relevant 2-person team. Main strengths, main weaknesses, conclusion and an overall grade (on an A-D scale) were all presented. Comments were raised and a marking was agreed upon. To every one's apparent relief, very few significant differences of opinion emerged.

Despite its apparent objectivity, the 'Bid Checklist' left plenty of room for value judgement and subjectivity. Given the nature of the checklist, it seems inevitable that different individuals would come up with different rankings for the same project based on individual values, background, expertise and interpretation of the requirements and the A-D scale. The format of the 'Overall Assessment Sheet' was also problematic in relation to its requirement

to award a 'Final Grade' to each bid. The question, that remains unanswered, is how was such a grade achieved? One might argue that the most likely answer is value judgement about the 'overall worth' of the bid. Although members of the appraisal teams drew on the resources and expertise available through their relative departments in conducting the task, an overall grade for each bid was clearly a very subjective issue.

7.3.2 Choosing Among Bids

The Working Group, in reaching their conclusions, were inevitably influenced by external pressures and considerations (most notably the political ones - see below). A package of recommended-for-approval bids was agreed upon and the conclusions were then passed upwards to the Regional Directors. To a large extent, this package was determined by political considerations. In its meeting, the Working Group first considered the local authority-led bids. When the largest local authority failed to produce sound, coherent proposals, members became aware of the constraint imposed over their decision. It was recognised that it would be difficult to refuse funding to certain prominent authorities whatever the quality of their bids. This imposed a pressure to base decisions on geographical as well as quality criteria.

In the case in question, this meant that no more than one bid could be recommended for each local authority and in some cases high scoring schemes had to be relegated to the reserve list - or even rejected. The next

four bids to agree upon were all within the area of one department and the fact that they were region-wide in coverage was one of the reasons behind their approval. They seemed to provide a compromise for 'satisfying' all the local authorities of the region. However, the decision was influenced by the argument of this department's representative in favour of the four bids. The remainder of the package included three bids; a regional-wide one and two pilot projects. Of all the decisions taken at the meeting, the choice of these bids was indeed the most objective.

7.3.3 Budget Constraints

The total SRB funding requirements of the agreed-upon package exceeded the then indicative budget of the region. There was a need, thus, to adjust the expenditure profile of some of the bids especially given the apparent overlap in activity between some of them. The indications of over-bidding in some of the bids was also a reason behind these adjustments. This suggested the need to contact bidders to clarify their intentions; something that was seen by some members of the Working Group as involving negotiation. Eventually, the decision was made not to contact bidders, on the grounds of fairness (and avoidance of legal problems). This decision was in accordance with the advice of the national SRB Steering Group. Provisional financial adjustments were thus made, mostly on the basis of overlapping between bids, but there was still some scope for value judgement.

The Regional Directors' final decision was, by and large, in line with the

Working Group's package, except for two issues. Firstly, there was a call for further reduction in expenditure, across almost all the bids. This was due to the realisation that the indicative budget was less than the GO had been led to expect (something that happened only on the very last day before submission was due). Secondly, there was a change in the 'reserve' list of bids; a substitution of one bid in place of another.

7.3.4 Influential Factors

This may have not been the best appraisal approach to adopt. Some might argue for a more systematic, 'scientific' approach to the task. Indeed, many of the Working Group members expressed their dissatisfaction with the guidelines and constraints they had to work within. Nonetheless, there were several crucial factors, in addition to budgetary constraints and central guidelines – or rather lack of it, which suggested there was probably no better method of selection.

The Bids Submitted

Several of the bids submitted to the GO failed, in one way or another, to comply with the format set in the Guidance. Thus, it became difficult to extract the information needed to base any judgment, especially about value for money. One of the main weaknesses was the untestable assertions about partnerships and outputs to be achieved. In one case, parties were listed as partners in a bid without their prior knowledge, not to mention consent. Members of the Working Group expressed considerable doubts whether the

levels of output claimed in many of the bids can actually be achieved. One of the bids, for instance, claimed a gearing ratio of 1:40. Most important, many of the bids failed to address the issues of risk, uncertainties and realism efficiently. There was also a fairly widespread failure among bids to address the issues of monitoring and evaluation.

The quality of many of the bids as a whole was also criticised. Many bids lacked a strategic approach to the issues they tackle. In one of the bids, it was difficult to relate the strategy (text) to the activities (tables) proposed; therefore it was not clear what the bidders intended to do.

In addition, there was the diverse coverage of the bids. Each bid aimed at a different area, population and activities. There was no common ground between the bids; not even the output measures as the combination of outputs varied considerably from one bid to the other. To reach an overall value for money index of each bid, for instance, a common unit has to be established in which all the outputs can be measured; a far from easy task not least for the lack of detailed information.

Political Considerations

Political considerations were probably the most crucial factor influencing both the appraisal process and the working of the GOs as a whole.

- *Local Conflicts:* Regional Offices, by definition, encompass a number of local authority areas. Despite limited resources, decision-makers were

inevitably concerned to exhibit a degree of 'even-handedness' across the region. This led to decisions being taken which did not coincide with an "objective" analysis of the bids – including one case of a bid being recommended despite having been judged poor at the start of the process. This very fact, however, raises considerable concerns about the 'fairness' of the process as a whole.

The relatively small number of both the authorities within the region and the number of total bids submitted to the GO was something quite unique in the case in question. However, this represented another source of pressure. The regional office appeared to be unable either to exclude a local authority or to support more than the same number of bids for each local authority regardless of the quality of the bids or of the varying local needs as this could have been taken as some form of favouritism. In other regions, the large number of local authorities involved and bids submitted and the limited resources available, when taken together, meant that some local authorities, sometime prominent ones, had none of their bids recommended for approval. Whether it is favouritism or exclusion, the repercussions of such a notion over the relations between the various levels of government and between local authorities and GOs could be quite damaging. Unfortunately, there seems to be no way out of such a dilemma. Such political considerations will always leave GOs in the same position, whereas some local authorities may exploit the situation for their own advantage.

• *Departmental Conflicts:* Another consideration that made the appraisal task more difficult was the fact that the SRB, and consequently many of the bids, cut across the domain of several government departments. Whether explicitly or not, members of the GOs staff are likely to pursue, or in a sense protect, their own departments' interests. The decisions taken may in turn be based on subjective rather than objective judgment. This seems to have taken place. For example, the bids within the area of a particular department were argued for so forcefully by the member representing this department that despite a departmental team assessment which put the bids involved within a third class category (overall grade: C), the member insisted on placing them within a second class category (overall grade: B), compared with the rest of the bids. The argument became apparently biased when compared with the same member's critical assessment of other bids, and actually of the bids themselves⁽¹⁵⁾. However, this is not to say that all the members' judgments were necessarily subjective. Other members were highly objective in assessing those bids falling within their area of interest.

• *Other Commitments:* One of the issues that remained largely unresolved in the appraisal process was the commitment of other, currently in operation programmes. Many of the bids were in partnership with, and required matching funding from, programmes such as English Partnerships. Although the partnership was endorsed by such programmes, no final decision was

⁽¹⁵⁾ At the same time as most of the successful bidders were claiming their first quarterly grant payment, one of those four bids had not yet produced an approved Delivery Plan. This was achieved in August, 1995.

taken on whether to provide the funding required or not; the issue had to be referred to HM Treasury to decide upon. Such an issue is further complicated in regions eligible for EU grants (ERDF, ESF, ...). It is well known that these grants are being administered on a different time table than the SRB. Some bids have required matching, sometimes substantial, funding from these grants. Both issues remained unresolved until well after the final decision on successful bids (see below).

- *Central Pressures:* Furthermore, there remains the 'national' scope of the SRB initiative which exerts some 'central' pressure over the regional offices. GOs work within the limits of an indicative budget. Combined with the concern over local authorities' reaction if excluded, the GOs may have had to relegate very sound bids, but not local authority-led, to their 'reserve' list in order to meet this limit. Ironically, the lack of clarity as to the size and status of the indicative budget, in the case in question, led to the need to make reductions to the recommended bids (both their expenditure and outputs) at the very last minute. Most of these adjustments were based on very sound, objective criteria (eg scaling-down or re-phasing). Yet, subjective judgement was inevitable given the lack of details in many bids.

At the time the decision was being taken, it was unclear whether successful bidders would agree with such adjustments. Either way, the adjustments were seen likely to be unpopular as it would become clear that the bids approved are different from what was submitted. It was feared that

other, unsuccessful bidders and third parties may question the equitable treatment of bids. It was also feared that if many of the successful bidders did not accept the scaling-down of their bids, this would call for further modifications of the final package.

However, the 'central' view was that bidders are unlikely to look a gift horse in the mouth(!). In our case study, this turned out to be the case. Successful bidders have not objected to the reductions in their funding profiles. Moreover, they have actually committed themselves to delivering the initial level of outputs for the reduced funding. Despite growing attention (and criticism) of the first bidding round, these issues have gone unnoticed. In our case study, in particular, none of the unsuccessful bidders has questioned the output of the process. Two questions remain unanswered, though. Firstly, on what grounds have such indicative budgets been determined? Secondly, how has the choice between the various packages submitted by the different GOs been made?

- *The Decision-making Process:* Finally, the 'multi-level' nature of the decision-making process is another factor that was expected to influence the final output of the bidding round, on the national level. In addition to the Working Group, there are three, and possibly more, levels of decision-making: the Regional Director(s), the Ministerial Committee and the Secretary of State for the Environment. As mentioned above, the Working Group recommendations were altered, though slightly, by the Regional Directors

(again in response to political considerations). It was expected, given the various levels of decision-making involved, that the packages submitted by GOs may not remain intact by the end of the process. This, however, was not the case (see below).

Information and Time Constraints

As might be well expected, the various levels of the decision-making process had different information requirements, and these kept changing till the very last moment(!). However, the decision on which information had to be provided was taken centrally by the national SRB Steering Group. The Group seems to have adopted an ad hoc approach in its response to GOs' queries about the format in which packages were to be submitted. Indeed, many of their 'explanations' were unclear and contradictory with each other.

As most, if not all, GOs are not yet well computerised, meeting these changing requirements simply meant more, needless pressure especially given the time constraint imposed. The Bidding Guidance initially allocated the period extending October to November for GOs to evaluate bids and raise their recommendations to Ministers (which is by any standard a very short period). However, the actual time GOs had to conduct the task was much less than that. Bids were submitted, as scheduled, on the 7th of September. Recommendations had to be reported to London by the 3rd of October (or the week beginning 3 October); less than a month. The structure of the GOs meant that the time Working Groups had to appraise projects was even much

less than that. Working Groups had to reach a decision early enough for the Board of Directors to consider it and prepare their report to the Ministers. In the case in question, the Working Group had only 8 working days to accomplish the task.

Having assessed the merits of individual bids, GOs were left with the task of choosing among those which scored a high overall grade. The Bidding Guidance referred to the 'Index of Local Conditions' (DoE, 1994b) as a possible aid. But the Index was not put to use in our case study, and probably others (senior members of the GO have indicated this possibility on the basis of their contacts with other GOs). Whether it would have been of use or not remains largely to be tested. The Index, nonetheless, does not substitute for information on local communities' needs, preferences and objectives - information that might have been of help in balancing the political pressures. Such information was, unfortunately, unavailable and what bidders claimed was taken, in many cases, for granted. Many members have expressed their belief that a need-assessment would have been of great help in the appraisal process. Some believed it may have been a way out of the political constraints imposed on their choice. Lack of such information meant that value judgement was the only resort left in several cases.

7.4 THE AFTERMATH

7.4.1 Before the Final Decision

Between the submission of regional packages and the announcement of the final decision, GOs and DoE HQ remained on an almost continuous contact initiated for several reasons:

Information Requests

Both the Ministerial Committee (the EDR) and the DoE HQ requested, on several occasions, additional information on the packages recommended for approval; information that was not initially requested and could not be extracted from the packages (eg total expenditure per main objective). Some GOs had already supplied such information in their packages. Others, for one reason or another, did not. In our case study, neither was this information initially supplied nor was the GO in a position to respond to every such request. The reason simply was the vague nature of many of the bids (see above).

The Budgetary Limits

One of the major problems was that the exact amount of available funding only became known at a very late stage. GOs entered the process on the assumption of an indicative budget limit. In our case study, however, this limit was not clarified until the very last day before submission of the recommended package was due. Yet, GOs were saved another, very late, shift in policy. By mid-October, less than two weeks after submission, the EDR

requested the GOs to re-consider their recommendations in the light of possible increase in available resources. By mid-November, the GOs were notified of a uniform 25% increase among them, bringing the overall first year SRB funding to £125m rather than £100m.

In our case study, this increase was put to use in two ways. First, the addition of two more bids to those recommended for approval. The first of these two bids was the second submitted by the largest authority; one amongst the worst scoring bids. The second, in the Working Group's words, was not an entirely SRB type of project, as clearly indicated on their meeting around mid-September. Second, the funding profile of many of the bids was re-readjusted, mostly in the form of increasing that profile. It is not clear at all on what grounds were such decisions made, though they undoubtedly cast considerable shadows over the objectivity of the whole process. Or else, on what grounds were these two bids rejected in the first place and others scaled-down?

Appraisal, Monitoring and Management Guidelines

The DoE HQ, in consultation with the GOs, was seeking to produce guidelines for the appraisal, monitoring and evaluation and management of projects of successful bids. Much of the attention was initially drawn to the management of successful bids. The first proposal of the DoE HQ came around the end of October in the form of a 'contractual agreement' between GOs and bidders. In the words of one senior officer, the proposal "was given

a pretty hard time". Many, if not all, GOs in their response to the proposal were highly critical of its subsequent legal, personnel and financial complications. The latest proposal, which was subsequently approved, was an explicit agreement between GOs and successful bidders as to what outputs to be delivered, at what timescale and at what cost. It should also be made explicit who is accountable for delivery. This agreement is to be expressed in a 'Delivery Plan' the nature of which is intended to reflect the nature and complexity of the scheme.

7.4.2 The Final Decision

Although GOs were notified of the outcome sometime earlier, the final decision on the first bidding round was formally announced on the 6th of December, 1994. This is a month earlier than was initially planned, and there is no given reason why the whole process was brought forward.

In our case study, the package approved is identical to that submitted by the Office (having been modified in accord with the latest funding increase). This brings us back to the question of how, and on what basis, was the central decision made. Although it may not be the case for other GOs, it appears that adherence to the budget limit was condition enough to grant approval. On the one hand, this may be justified by the sheer size of the task forwarded to the EDR. It may also serve to indicate the Government's move towards decentralisation. On the other hand, however, it may indicate a meaningless exercise of 'rubber stamping'.

7.4.3 After the Final Decision

The consultation between the DoE HQ and GOs materialised, in February 1995, with the publication of four guidance notes intended to help with the management, implementation and monitoring and evaluation of successful bids (see Appendix D):

1. SRB Guidance Note No. 1: Partnership Delivery Plan (DoE, 1995a).
2. SRB Guidance Note No. 2: Monitoring and Periodic Review (DoE, 1995b).
3. SRB Guidance Note No. 3: Project Appraisal and Approval (DoE, 1995c).
4. SRB Guidance Note No. 4: Financial Guidance (DoE, 1995d).

Partnerships were given until the end of March 1995 to prepare their Delivery Plans (in consultation with GOs) so that implementation would start, as planned, in April. However, not all partnerships were successful in meeting this target; as mentioned above, one of the approved bids only produced an agreed-upon Plan in August 1995. There are as well variations in the 'quality' and details of those approved Plans. By April 1995, only a few schemes were able to start implementation. Those schemes that required matching funding from EU grants (where applicable) have got underway already without a decision being made in regard to this funding. Other schemes that required substantial funding from existing programmes (eg English Partnerships) also had to wait for the final decision to be made by the Treasury (July, 1995). In other words, many of the schemes have started implementation with a considerable degree of uncertainty surrounding their funding profiles.

The final decision on EU grants was finally made in October 1995. Questions, however, remain about the detailed arrangements and payment of grant (December, 1995). Some schemes have actually been refused EU matching funding. In one case, this resulted in a 25% deficit in the bid's financial profile. On the one hand, bidders are seeking funding from other sources. On the other hand, the GO has indicated that a flexible approach will be adopted in such and similar cases if bidders had to scale down their schemes. Sometime between September and October, 1995, HM Treasury approved, in principle, the agreement between successful schemes and programmes such as EPs. However, the exact amount of funding for 2 of the large schemes is still the subject of negotiations (December, 1995).

Payment of grant is on retrospective, quarterly basis. Partnerships were to submit their first claims (April-June), supported by progress reports, by the end of July 1995. More than half the schemes have submitted their claims, and all have been approved. Some have not achieved their set targets of outputs whereas others have claimed to have exceeded their targets. With regard to those who have not met their targets, the GO has apparently exercised a degree of flexibility in recognition of the late start they had. The assessment of claims was no more than a comparison between targets set in the Delivery Plans and 'actual' output.

In line with Guidance Note No. 1 (DoE, 1995a), approved schemes have set out their plans to establish their partnerships; many envisage a series of

meetings and consultations among the different partners. The GO, in arrangement with partnerships, have appointed one member of staff for each scheme to attend these meetings only as an observer. The GO is clearly taking a close interest in the evolution and working of partnerships; it is not clear however whether other GOs are adopting the same approach. It is also unclear yet how, if at all, will schemes be assessed in regard to their organisational aspects. The GO *may* get involved if, for instance, one of the partners raised complaints about their representations. Yet again, it is still unclear in what form will this involvement be. Neither is it clear whether such incidents will be considered 'poor performance' for which partnerships may be penalised.

7.4.4 The Second Bidding Round

The second bidding round was launched on 12 April 1995. This round was the subject of a modified version of the initial Bidding Guidance (DoE, 1995e). The principles of the initiative remain essentially the same. This revised guidance, nonetheless, has addressed some crucial aspects, apparently in recognition of the difficulties encountered in the first round. First, it provided a 'proforma' for outline bids in an attempt to secure a degree of consistency between all the outline bids (in terms of clarity and information). It was hoped to facilitate the preliminary sifting of those bids. This, unfortunately, turned out not to be the case. Many of the second round bidders have interpreted this proforma in different ways leaving the GO with the task of extracting relevant information.

The criteria for assessing individual bids, on the whole, have been made 'slightly' clearer. However, the onus of choosing among bids remains for the GOs. More important perhaps is the requirement for bidders to consider explicitly the possibilities of modifying their bids. Potential bidders are requested to indicate effectively the scope of scaling down (by 10%, 25% or 50%), scaling up, removing discrete elements and deferring/extending expenditure (DoE, 1995e: 18). This provides GOs with a great deal of flexibility both in choosing among bids and in preparing their recommended-for-approval packages; flexibility that was largely missing in the first round. This ensures, to a large extent, that modifications of bids are based on sound (objective) criteria, rather than subjective judgement as was the case in the first round.

Outline bids were due no later than 16 June 1995 whereas the final bids were submitted by 18 September; as with the first round, GOs were expected to make their recommendation to Ministers sometime between October and November 1995 (DoE, 1995e: 5; 38). The final decision was announced on the 12th of December 1995.

7.5 CONCLUDING REMARKS

7.5.1 The Appraisal Process

Despite being an 'ex-ante' experience, the appraisal of SRB first round bids has revealed several concerns of wider relevance, many of which are equally crucial in ex-post evaluation. At the outset, there is the oft-cited problem of the vague statement of objectives. It seems that the objectives of the SRB are a collective statement of the objectives of those programmes it brought together. These vague statements provided no benchmark against which bids could be assessed. They provide no yardstick against which any future progress can be judged. The appraisal process has as well demonstrated that time and resources constraints are becoming almost the norm in evaluation research. Furthermore, it is evident that evaluation is intertwined with politics; political considerations have been the single most influential factor on the process and its outcomes.

The experience, moreover, has a significant bearing on the first part of this research, namely the applicability of 'established' evaluation methods, eg CBA and GAM. In choosing among the bids, it became clear that these methods are of little, if any, help. It may suffice to recall that these techniques were conceived to choose among alternatives of the same strategy. They were designed to compare among different means of achieving the same objectives for a particular area/group. Clearly, that was not the situation in the first bidding round; and it may never be assumed in the SRB. There is hardly any common ground among the various bids — apart from the very broad goal of

urban regeneration. Each bid aimed at a different area, population and activities with varying emphasis on different objectives. Even the output measures were no longer common ground for comparison. The combination of outputs varied dramatically from one bid to another.

Putting aside the conceptual difficulties of CBA (see chapter 3), it was not possible to arrive at a 'net present value - NPV' or a 'benefit/cost ratio - B/C' for each scheme. On the costs side, the total funding requirements are clearly not the 'actual' costs of a scheme. This is simply because neither the potential 'negative' impacts nor the 'opportunity costs of resources' was known. On the benefits side, the hoped for outputs may not be all the 'benefits' of a scheme due to spill-over effects. Also, the exact nature of many of the outputs was unknown (eg types of job to be created). Thus, it was not possible to value even many of the quantified outputs – not to mention intangibles. In addition, costs were not broken down by outputs. Consequently, a 'cost per unit output' was impossible to arrive at and cost-effectiveness analysis became no substitute.

Neither was it possible to identify 'producers' and 'consumers' within the targeted communities nor the incidence of costs and benefits on each. A planning balance sheet analysis was thus ruled out. The vague statement of the objectives of both the SRB and the bids was the main reason why GAM was inapplicable. There was no means by which the contribution of each bid towards the achievement of different goals could be assessed. This was also

a problem for MCE methods since 'evaluation criteria' have, at least to an extent, to reflect the goals and objectives of a scheme. In both GAM and MCE the analysis would as well be partial since the goals, preferences and needs of targeted areas/groups were unknown.

In short, there was no means to 'systematically' assess the merits of individual bids. The Working Group had to rely on their judgment (1) whether the proposed measures in each scheme were the 'most appropriate' to tackling what it claimed to be the problems and needs of the targeted area/group and, (2) the 'soundness' of the financial claims of each bid. This judgment was based, in the main, on personal knowledge of local circumstances (and bidders) gained through involvement in previous government programmes within the region. On the basis of this judgment, each bid was given an overall score. The Working Group chose the recommended-for-approval package guided by these scores within the constraints imposed on their work (see above).

7.5.2 The Evaluation of the SRB

The SRB process and guidelines reveal a mixture of encouraging signs and potential difficulties for the evaluation of the initiative. There are three encouraging signs. First, from the earliest stage of the initiative, there has been a strong commitment to regular monitoring and periodic review at the various levels of the process (DoE, 1994a, para 51-53). Subsequent Guidance Notes have reiterated and emphasised this commitment to the point that

continued funding is related to compliance with monitoring arrangements regardless of the outputs actually achieved (DoE, 1995b, para 13). Second, the distinction between outputs and outcomes was clearly recognised (DoE, 1994a, para 54 - see Appendix C). The guidelines indicated clearly the need to go beyond counting outputs to identify outcomes (DoE, 1995a, para 4.3). Partnerships are also required to indicate improvements in relation to baseline conditions (DoE, 1995a, para 4.1). This distinction, together with the commitment to regular monitoring, lead to the expectation of an extended approach to monitoring; to measure outcomes as well as outputs. If that was the case, evaluation may have been made relatively easier.

Finally, there has also been a commitment to evaluation, though slightly less strong than that to monitoring. "The operation of the Budget as a whole *may* be subject to independent national evaluation" (DoE, 1994a, para 57, stress added). The Bidding Guidance and the various Guidance Notes, in between them, emphasised, on numerous occasions, the commitment to evaluation of SRB-assisted schemes and the need to set their arrangements from an early stage in the scheme preparation stage.

Nevertheless, it is also clear that the evaluation of the SRB is likely to face several problems. The first critical difficulty stems directly from the objectives of the initiative (which are themselves a direct result of the nature of the Budget and its underlying concepts). These objectives are obviously very vague; they are merely a collective statement of the objectives of those

programmes put into the SRB. At the national level, there is no means of judging whether the SRB has achieved those objectives or not. There is no 'benchmark' or 'yardstick' against which movement in any direction could be assessed.

In the first place, the initiative as a whole relates to no definition of the problems it aims to tackle. This lack of problem definition makes it virtually impossible to assess the 'relative' magnitude of improvement the SRB may bring about. This is similar to the critical problem in assessing the impact of regional policy (as shall be explained in chapter 9). The absence of any quantified estimate of the scale of regional problem made it "quite impossible to judge how effective regional policy must be over a given period to solve the problem, and more importantly, the extent to which previous policy measures have fallen short of solving it" (Moore *et al.*, 1982: 3). Therefore, the evaluation of the SRB will fall short from assessing the achievement of its objectives; it may, somehow, provide an estimate of its outputs and, possibly, outcomes. Whether these are "achievements" will be up for grabs! No matter how many *new* jobs are actually created, for instance, there is bound to be those who claim it to be an achievement.

The comparison between baseline conditions and improvements, though crucial, is not sufficient in itself to assess achievements at the local level. The targets of these improvements have been set by the partnerships in the absence of any regional development strategy. Whether these targets are the

'best' estimate of what, realistically, should schemes strive for is open to dispute. Furthermore, and cynical as it may sound, partnerships may have exacerbated the needs and problems of the areas they targeted in as much as they might have inflated their output targets in order to secure success in the competitive bidding process.

The approval of a scheme is an approval of its targets. Assuming then that, at the scheme level, realistic targets have been set, it follows that achievements at the scheme level can be assessed. It still remains to assess this achievement at both the regional and national level. Will the total achievement be the sum of its individual components? The answer is very likely to be 'No'. The introduction of the SRB, as a catalyst for investment, is likely to have wider impacts than the sum of individual schemes.

Another immediate problem is the range of output and outcome measures proposed by both the SRB and the partnerships. It is questionable whether any of the two sets is comprehensive enough to reflect closely the objectives of the SRB (vague as they are). The definitions underlying some of the output measures are open to dispute. For instance, a permanent job is defined as one that is likely to last for at least six months. If the aim is 'sustainable' economic growth, how much would short-lived jobs contribute to this aim? Might not that encourage a high turnover rate of the workforce and an over-estimate of outputs and outcomes? Perhaps more important is the range of outcome measures and whether or not they represent the closest proxies possible of the

objectives. Although they might, they all lack a very crucial element in their definition. That is, the duration and stability of these outcomes. Again, if the aim is sustainable development, these outputs have to endure for a period of time enough to counteract persistent negative trends.

It may appear that the evaluation of the SRB, by focusing on project-related monitoring and evaluation, has escaped the counter-factual problem. This is not entirely true however. The operation and performance of any project at a locality is susceptible to the influence of other local, regional and indeed national trends. The local economy is part of the regional economy which is a segment of the national economy. The national economy in itself is influenced by global trends and shifts. Moreover, the SRB, by its very nature, requires a level of integration between proposed schemes and current regeneration programmes and projects. Projects' performance is then likely to be influenced by the performance of those programmes, even if not in partnership with any of them. Furthermore, it should go without saying that the contribution of the private-sector is dependent on factors other than and in addition to government grants. It is equally true that there is likely to be social interactions between the various localities and regions (eg migration). These affect, and are affected by, economic trends. It is quite naive then to assume that projects are operating in complete isolation of their surroundings. The question 'what would have happened in the absence of the policy?' is equally important in the evaluation of the SRB as it is in other cases.

A critical issue that appears to have been overlooked is the spill-over effects of SRB-assisted schemes. The focus merely on scheme-related outputs/outcomes may result in an over- or under-estimate of the overall impact of the SRB. Due to economic (and indeed social) linkages, activities in one area are likely to influence socio-economic conditions in other neighbouring localities. In addition, the SRB-assisted schemes may have indirect effects, within their locality, that are clearly not covered by focusing monitoring solely on those schemes. It is as well likely that these schemes may have induced effects on the area they target. There is apparently a lot of thinking to be done about these likely chains of actions if an 'accurate' estimate of the magnitude of effects is to be achieved.

A distinctive feature of the SRB (and its predecessor, City Challenge) is the highly selective approach to targeting of funds. The Bidding Guidance (DoE, 1994a and 1995e) have made it clear that schemes should be responding to local need. GOs were urged to make use of such information sources as 'The Index of Local Conditions' (DoE, 1994b) to help in assessing local needs and priorities. Guidance Notes 1 and 2 (DoE, 1995 a and b) emphasised the need to ensure that those who are intended to benefit from a scheme do so (and actually are involved in the scheme). By emphasising such issues as 'Ethnic Monitoring' (DoE, 1995b, para 14) and, at the outset, 'equality of opportunity' (DoE, 1994a, para 9), the SRB appears to be tackling the equity consideration head on.

That, however, does not eliminate the problems of 'equity', neither in the appraisal nor the evaluation of schemes. Due to limited resources, a choice still has to be made even between the most 'needing' localities. In evaluating these schemes, it is equally important to assist their indirect and/or induced effects on non-targeted neighbouring areas and populations. That is, the appraisal and evaluation of the SRB may involve the (almost impossible) task of weighing different groups of the society against each other.

The focus on quantified outputs/outcomes in the evaluation of the SRB may, however, be misleading. The SRB has carried forward a distinctive characteristic of its predecessor, City Challenge. That is, the emphasis on creating partnerships at the local level encompassing all parties of interest in regeneration: the public, private and voluntary sectors and community groups. The establishment and sustainability of these partnerships is clearly an objective in both initiatives. It follows, then, that the evolution and working of partnerships should be among the topics addressed in any evaluation of the SRB. The illustrative example of Delivery Plans (DoE, 1995a) did enlist the progress of setting up partnerships and their workings among the activities that will be monitored. Indeed, almost all approved Delivery Plans have followed this example and detailed their envisaged plans to create and maintain their partnership arrangements.

There is, however, another equally important reason why partnerships should be evaluated. The involvement of several parties — with different and

even conflicting interests, preferences, backgrounds, ... – in the process simply means that decisions are most likely to be the outcome of compromise rather than consensus. The only way to understand why things happened, on the ground, the way they did is to follow the 'process' through which they have been delivered. In other words, equal emphasis must be placed on 'process evaluation' (see chapter 6) as much as it is placed on quantified outputs/outcomes.

CHAPTER 8:
URBAN POLICY EVALUATION -
3 CASE STUDIES

8.1 INTRODUCTION

One of the conclusions to emerge from Part 1 is the severe lack of research on ex-post evaluation methods for urban policies. Yet, numerous studies have been carried out over the last decade or so. These two facts put together raise a crucial question about the methodology adopted in this practice. How did practitioners approach the task? What were the problems they faced? What were the constraints imposed on their work? How did they overcome these problems and constraints? Is there some sort of an unpublished 'code of practice' by which evaluators are guided in their work or is it a matter of improvisation and personal endeavour in every case?

Given the methodological orientation of this research, the above questions warrant a closer examination. Therefore, this chapter is, in a sense, an evaluation of previous experience intended to inform this research in order to avoid past failures and build on success. In focusing on methodological questions, this chapter, as with the following two, centres around the four issues detailed at the introduction to this Part: 1) measurement of outputs/outcomes; 2) the treatment of the counter-factual problem; 3) assessing the achievement of objectives and, 4) the explanation of results. Undoubtedly, an assessment of all previous studies of urban policy ex-post evaluation is virtually impossible. Therefore, this chapter examines the experience of three case studies: the Programme for the Valleys (PFV) (Victor

Hausner & Associates, 1993), 'Assessing the Impact of Urban Policy' (Robson *et al.*, 1994) and City Challenge (Russell, 1994 a and b).

These three cases are among the most - if not *the* - recent research in urban policy evaluation. That has certainly been one of the reasons behind their choice, but not the only one. The first case study was claimed to be one where 'vague policy objectives' - a major problem in evaluation - have been dealt with (Personal contact, 1995). The comprehensiveness - at least in coverage - of the second case study was one reason to include it in this review. It may have a significant bearing on 'comprehensive' policy evaluation as defined within the context of this research.

The experience of City Challenge (CC) evaluation is perhaps the most important of the three case studies (despite the lack of detailed information). Many of the 'innovative' features of City Challenge have been carried forward into the latest central government initiative: the Single Regeneration Budget (SRB). The emphases on partnership, strategic and integrated approaches to regeneration, targeting disadvantaged groups and areas and competitive bidding are all key propositions of the SRB as its predecessor the CC. It seems that these features are set to underpin future government initiative; they are here to remain. Even if not, an evaluation of the SRB will still benefit from any lessons emerging from CC, if only for those common features.

This chapter has four main sections; each of the first three is devoted to

a case study, the final section brings them and the issues they raise together. Section 2 covers the evaluation of the PFV. It starts with a brief account of the Programme followed by the terms of reference of the study. It then moves to the methodology adopted to accomplish the task. The discussion - the final part of this section - relates primarily to the methodological difficulties and deficiencies of the study. The complete absence of any consideration of the counter-factual problem is the most critical deficiency of the study. This means that the 'actual' impact of the PFV remains largely unknown. The crucial issues of displacement and equity were not covered. Neither were the observed changes related to the objectives of the PFV. The indicators used seem to be those readily available, or measurable, not necessarily the most appropriate. Finally, the study lacks any attention to people's perceptions of the Programme's impact on their communities.

'Assessing the Impact of Urban Policy' is perhaps the most ambitious evaluation research to date. Section 3 starts with a summary of the broad objectives of the research and the conceptual and methodological difficulties these objectives raised. It then moves to the research's sample of local authorities and its 'multi-stranded' methodology. This is followed by a review of the outcome indicators employed and the objectives on which they were based. In the final part of this section, it is argued that despite the Team's efforts there remain several crucial issues that may undermine the validity of the research's findings and conclusions. Some of these issues result from the outcome indicators, whereas others are the outcome of the quantitative and

qualitative analyses. In more general terms, the research founders on two further grounds. Firstly, the lack of explanation of reasons behind failure. Secondly, the apparent neglect of the mechanisms behind policy outcomes. It is recognised, however, that many of these issues are due mainly to the very wide focus of the research. The attempt to overcome any of these concerns would call for a 'finer-grain' research that looks at individual programmes and areas separately; something apparently beyond this research.

Section 4 looks at the experience of evaluating City Challenge at both the national and local levels. Similar to section 2, it starts with a brief review of the initiative and its key distinct features. This is followed by a summary of the government guidelines on monitoring and evaluation of CC projects. The main characteristics of the interim national evaluation of CC are then outlined. Unfortunately, the information available is not sufficient to allow for any further analysis. That has also been the case with local evaluation. However, the evaluation of CC at both levels has a distinct feature, with no matching in the previous two case studies. That is, the clear and strong emphasis on process evaluation.

As the case with the previous two sections, section 4 also concludes with a discussion of the issues which local evaluation raises. This discussion is based mainly on the outcome of meetings with a member of the local evaluation team and a senior member of the CC project that was the subject of this local evaluation. At the outset, the two seemed to disagree on the level

of commitment to evaluation at the early stages of the initiative. They share, however, the same views when it comes to such concerns as data and information, measurement and shifting objectives and the problems associated with each and all of them. The relation between the two teams was one of co-operation; many of the evaluation's recommendations have actually been put to use. This serves to indicate that tensions between evaluators and 'those being evaluated' can be resolved. A continuous dialogue between the two parties, right from the start, and explicit consideration of issues of concern to the CC team was apparently the key to this resolution.

The discussions, at the end of each section, have been informed by several sources. There is of course the literature on policy evaluation, to start with. There is also the insights gained from examining evaluation research in other fields (ie regional policy and transportation; see chapters 9 and 10) and the personal contacts initiated during the different stages of the research. Each of these three case studies actually informs the discussion of the other(s). Perhaps more important is the contacts established with some of those who have been involved in these studies. The first case study (PFV) benefited from a meeting with a senior member of the evaluation team (July, 1995). For the CC experience, as mentioned above, two meetings were held, one with a member of the evaluation team (May, 1995) and the other with a member of the CC team itself (August, 1995). These meetings shed considerable light on each of the two cases, and more importantly, on several of the concerns and tensions likely to surface in almost any urban policy evaluation.

The final section of this chapter first compares the three case studies and then draws these separate strands together in an assessment of their bearings on urban policy evaluation. The three studies varied dramatically in their treatment of issues such as the counter-factual problem and displacement and equity concerns. The vague statements of objectives was a problem for the three studies; it was further complicated in the second and third by the change and shift in these objectives. However, the vague objective of 'partnership' represented no problem for the evaluation of CC. Another problem common to the three studies was the definition of the "most appropriate" performance indicators. The development of such indicators is bound to require further research, and hence resources. It is argued that the benefits of such research, in the sense of better-informed decisions, cost-justify the endeavour.

Of the three case studies, that of the CC is seen to be the most important. Many of the key propositions of CC have been carried forward into the SRB. Therefore, it is argued that an evaluation of the SRB will benefit from the experience of assessing these propositions. It is finally argued that evaluation will have to strike a balance between the traditional, quantitative output/outcome measurement and the qualitative, process issues; not only what is being done but also how.

8.2 RETROSPECTIVE EVALUATION OF A SINGLE POLICY - PROGRAMME FOR THE VALLEYS

8.2.1 The Initiative

The Programme for the Valleys (PFV) was launched in June 1988 in response to the social and economic problems faced by the communities of the South Wales Valleys. The Programme was designed to address the economic, social and physical conditions found in the Valleys following the decline of its primary industrial base - coal extraction and steel production.

The stated aims of the Programme were:

- to improve significantly the prosperity and attractiveness of the South Wales Valleys and the well being of the people within them;
- give people in the Valleys a confidence in the future of the Valleys as places in which to live and work;
- instil in people elsewhere a new perception of the Valleys as places in which to invest and visit (Welsh Office, 1988 - quoted in Morgan, 1995: 205).

In more concrete terms, the Programme aimed to reduce unemployment by up to 30,000. It also aimed to improve education and training, remove environmental degradation, enhance the housing stock and raise the quality of life. Another aim was to promote partnership between the various parties involved in social and economic regeneration. (Morgan, 1995: 205).

The Programme took as its starting point the substantial level of activity already existing in the Valleys by various government departments; it was developed on the basis of bending, or tilting, existing expenditure provision

(Victor Hausner & Associates, 1993: 15). The management of the Programme as such was expected to require "the development of a broad strategic approach, the setting of task related objectives and the creation of mechanisms for the co-ordination and monitoring of activity" (Victor Hausner & Associates: 19). In practice, however, individual programmes continued to operate under existing management arrangements with each agency responsible for developing its own strategies.

The Programme was initially for a three years duration. In July 1989, the then Secretary of State confirmed the intention to extend the Programme to five years. The extended Programme was formally launched in December 1990 under the title "Partnership with the People". However, while previous policies were to continue into the fourth and fifth year, the emphasis was on people and community based activity and projects with a high level of public involvement.

The five year Programme was completed at the end of the 1992-93 fiscal year. The Welsh Office, conscious of the fact that a five-year programme cannot reverse sixty years of decline and anxious to maintain the momentum, launched another five-year programme in April 1993 at a total cost of over £1 billion in public sector investment (Morgan, 1995: 208). The three main principles of the new programme were:

1. "to deepen and extend the *partnership* approach to urban regeneration by harnessing the resources of the WDA";

2. "to ensure a more *co-ordinated* approach to urban renewal by integrating the full range of services";
3. "to move towards a more *decentralised* model of urban regeneration by tapping local initiative to a much greater degree than in the past" (Morgan, 1995: 208, stress in original)

8.2.2 Terms of reference

The study was commissioned by the Urban Affairs Division of the Welsh Office to conduct a high level evaluation of the PFV and report on the impact of the Programme and whether it:

- achieved its overall objective
- represented value for money
- was well focused

Subsidiary to this, the study also aimed to derive lessons for the further regeneration of the Valleys (Victor Hausner & Associates, 1993: 1).

8.2.3 Methodology

It was claimed that the achievements of the PFV were evaluated in five broad and inter-related respects:

- **strategic objectives** and final outputs (outcomes);
- **quantifiable performance targets:** inputs and direct outputs;
- **innovation** in area regeneration
- **process objectives:** coordination, management and milestones, institutional and community capacity-building, targeting, dissemination of good practice;
- **sustained regeneration:** the promotion of an on-going development and private investment process (Victor Hausner & Associates, 1993: 1-2, stress in original)

It was recognised that no single methodology was adequate to conduct such a wide-ranging evaluation. "Some aspects required a quantitative approach, others a qualitative, perception-based approach" (p. 2).

Establishing a causal relationship between Programme inputs and direct outputs on the one hand, and statistical socio-economic trends in the Valleys on the other was "not possible, nor was it intended" (p. 2). The reason was the large area and population coverage of the Programme which - it is claimed - has limited "the ability to demonstrate the effect of the Programme on the overall character of the area particularly in the short term". Instead, key statistical indicators on a 'before' and 'after' basis had been used "only to provide a context for evaluation and a basis for identification of follow-up initiatives" (p. 2)

The Study had three other main strands:

1. A review of the various elements of the Programme in terms of expenditure and quantified financial and performance outputs.
2. A survey of a small sample of firms to assess the impacts of the Programme on perceptions and process objectives (coordination, management, capacity-building, ...).
3. Four case studies, based apparently on interviews, were examined to assess the impacts of key aspects of the Programme.

The progress and achievement of the Programme were assessed first in terms of the proportionate expenditure across the main programmes and activities. Four main elements of the Programme were then "evaluated in terms of their primary achievements, perceptions and key issues" (p. 40). These elements were: Economic Development, Education and Training, Transport and Housing. This evaluation, however, turns out to be no more than descriptive statements of expenditure, observed changes (based on

'before-after' comparisons) and untestable assertions of quantified outputs. For instance, financial assistance schemes to encourage inward investment (a total of £91.8m) were claimed to have attracted around £700m of private investment involving 24,000 jobs (p. 41). These figures, however, were based on an unsubstantiated statement by the then Secretary of the State for Wales. Moreover, the study admitted that it "has not been possible ... to establish how much investment would have occurred or how much Welsh Office assistance to the Valleys there would have been without the PFV" (ff. 41). Also, in considering the element of 'creating a new economy' it was asserted that "[t]he 'old' economy of the Valleys is long gone" (p. 74). This assertion was based on a rather simple comparison of employment in the PFV area, classified by sector, in September 1987 and September 1991 (p. 74).

The assessment of both perceptions and the managerial aspects of the Programme were gauged through interviews with "45 senior officials in the public sector and managers in the private sector" (p. 43).

8.2.4 Discussion

The evaluation of the PFV, as such, raises several crucial concerns which are of wider relevance, and have significant bearing on the evaluation of urban policy in general. Perhaps the most critical deficiency of the methodology adopted is the complete absence of any consideration of the counter-factual problem. "Evaluating the impact of the Valleys Programme is beset by all the standard methodological problems, like the counterfactual

problem of what might have happened in the absence of the programme and what causal significance can we attribute to the programme when it is just one among many factors influencing social and economic trends in the sub-region" (Morgan, 1995: 206).

The study quoted the large area and population coverage as the reason why the effects of the PFV on the overall character of the area could not be demonstrated, especially in the short-term. This is hardly a convincing argument. While it is well recognised that the *impacts* of a programme may take a longer period of time to materialise, that in itself is no reason to entirely overlook the relation between the programme and observed changes on the ground.

Perhaps more revealing, a senior member of the study team maintained that it is useless to pursue this causal relation, and strongly questioned its relevance to evaluation (Personal contact, July 1995). What is relevant, in his point of view, is to establish this causal link/rationale in the policy-making process, right from its inception. The vision behind any policy should relate to the aspirations of the community it is targeting. This vision should then be translated into "operationable, actionable objectives". These objectives, in turn, are translated into programmes and projects. Monitoring and evaluation should both be carried out, then, at the three levels: the relevance of the vision to the community's aspirations; the relevance of the objectives to this vision and, the relevance of programmes and projects, and their achievements.

Yet, how to ascertain that observed changes on the ground are the result of this policy remained an unanswered question.

Equally important, and related, is to disentangle the effects of other policies and trends. On the one hand, Morgan (1995: 208) noted three central government policies that ran counter to the aims and aspirations of the Programme:

1. The reduction in regional policy expenditure in the programme area by some 50 per cent in real terms during the first three years of the PFV;
2. The cuts in TEC budgets and, consequently, the inability of the Programme to prevent the attrition in training places in the Valleys;
3. The weakening of the local authorities' financial position.

On the other hand, it is well known that the diversification of the Valleys' economy has been well underway since the late 1970s. The Valleys were already targeted by several programmes when the PFV was launched (The Valleys Initiative, Urban Programme, ...).

Failure to address this issue simply means that the impact of the Programme remains largely unknown. What is known is a portrayal of pre- and post Programme social and economic conditions in the Valleys. Whether the changes in these conditions are the result of the Programme or other factors has yet to be established. It is, thus, totally unrealistic to attribute changes on the ground solely to the PFV.

Closely related and equally important is the issue of displacement. Have

the observed outputs/outcomes in the Valleys been at the expense of other neighbouring areas? Has inward investment been at the expense of the indigenous sector? (This also relates to equity considerations; see below).

The same senior member considered this to be no problem. Provided that there has been no loss in either jobs or output, and that the area in question has benefited, such a situation does not lead to any economic loss at the national level (though it does not result in any gain either). However, it can be argued that not only nothing is being achieved but that the situation may also result in extra public expenditure. Transferring the problem from one area to another (which is being partly paid for by public investment through that policy in question) may induce the introduction of certain measures in the negatively affected locality to offset these trends. In other words, things will go in circles; public expenditure being incurred to offset the negative impacts of previous allocations and policies in another area.

Another crucial concern is the achievement of objectives. It is equally unknown whether these changes have any bearing on the achievement of the Programme's goals. This is mainly due to the fact that the aims and objectives of the Programme "were sufficiently general to prohibit a definitive evaluation of their success" (Victor Hausner & Associates: i). The only objective to have been stated in quantified terms was jobs creation. However, claims that 24,000 jobs have been created have to be treated with caution; "the record suggests that these claims are exaggerated" (Morgan, 1995: 213).

The study, indirectly, pointed out to another problematic concern. That is, the choice of indicators. Performance indicators, especially in the absence of quantified objectives, should be the closest proxies possible to those objectives, not simply what is measurable or available. For instance, it was recognised that although the quality of life is closely associated with income levels, other factors (eg local services, environmental quality, ...) are also important (Victor Hausner & Associates: v). In measuring improvements in this respect, however, another indicator was employed. It was contended that the age structure and size of population in an area is a key indicator of its quality of life. It was asserted that an "area with a poor quality of life will tend to have a net migration of people of working age" (p. 85). In 1991, the PFV area had a slightly higher share of the 16-44-age group, compared with Wales despite the fact that earlier trends of net out-migration continued into the 1980s.

Nonetheless, it can be argued that a higher share of population in working age is not necessarily an indicator of an area's well-being. It can be argued that many of those living in many of the inner city areas are simply trapped there. The issue has to be considered in a wider context, especially in regard to long-term unemployment, before any conclusions can be drawn (eg Bradford and Robson, 1995: 42-44 and, Burton and Boddy, 1995: 29).

Another important example is that unemployment rates were among the indicators used to assess economic regeneration. It was noted that

unemployment has fallen over the period of the PFV (by 0.2% between 1988 and 1993 - p. 77). It was also noted, however, that employment and wage levels have also fallen, over the same period. It is thus very clear that "crude" unemployment rates are not a suitable indicator of economic regeneration. As the experience with Regional Policy has indicated (as will be explained in chapter 9), unemployment is subject to a multitude of factors and forces, especially on the supply side of the labour market. For instance, outward migration may decline when a policy is successfully creating jobs. Therefore, labour supply will increase over and above what it would have been in the absence of this policy.

In explanation of this situation it was suggested that "while new inward investment has benefited some members of the Community, overall improvements in prosperity remain elusive. For instance, unemployment amongst men has increased during the Programme Period from 17.8% to 19.8%" (p. iii). This raises another problematic issue; namely, equity considerations. Apparently, the Programme has affected different groups of the society in different ways. This issue has also to be considered in a wider, regional context. Achievements in the Valleys may have been at the expense of neighbouring communities (ie displacement; see above). If an overall picture of the impact of the Programme is to be achieved, these effects have to be measured first. Equity consideration were not taken into account.

There is no doubt that surveys, despite their limitations, are a crucial

source of qualitative information, particularly in both gauging perceptions and process evaluation. The results of the survey component of the study, however, have to be treated with extreme caution for two reasons. Firstly, the very small sample size (45 interviewees) and the non-random selection of this sample. The study actually admitted that the results "are not necessarily representative" (p. 3). Secondly, and perhaps more importantly, the complete exclusion of the targeted population of the communities. There is no indication whatsoever on people's perceptions of the Programme and/or its effects on their living communities.

8.3 COMPREHENSIVE EVALUATION - ASSESSING THE IMPACT OF URBAN POLICY

8.3.1 Objectives and problems

'Assessing the Impact of Urban Policy' (Robson *et al.*, 1994) was commissioned to "evaluate the overall impact of central government urban policy in England over the period of the last decade" (Robson *et al.*, 1994: vii). It was "specifically enjoined to look, not at any one of the elements of urban policy, but at the impacts of policy across the board" (Robson *et al.*, 1994: 2). The Team recognised the numerous problems such a very ambitious and difficult task involves.

Strategic Problems

The difficulty of identifying, unambiguously, policy aims at one time and to characterise one set of objectives over the whole period. Throughout the 1980s, new programmes came at different times reflecting the change in balance of government objectives. The multiplicity of programmes further suggests the shift in focus of policy (Robson *et al.*, 1994: 2). The Action for Cities (AfC) package, the main focus of the research, originally included a total of 33 programmes, the combination of which varied considerably both over time and between areas (Robson *et al.*, 1994: 3, 5).

Conceptual Problems

The research also faced several conceptual problems:

1. Counterfactual: assessing what might have happened in the absence of the government intervention;
2. Confound: outcomes can be affected by many other public policies in addition to those within the AfC package;
3. Contextual: local authorities started the period from very different conditions which is likely to affect their capacity for improvement;
4. Contiguity: an intervention in one area may have positive, or negative, spillover effects on adjacent areas which are not subject to intervention;
5. Combinatorial: public assistance has been delivered in differently constituted packages, some combinations may have worked better than others;
6. Choice: areas targeted for preferential treatment have altered over time and across different programmes, "any decision to assign particular authorities to a 'policy-on' or a 'policy-off' set cannot apply unambiguously over the whole time span and for all elements of policy." (Robson *et al.*, 1994: 4)

Of the above, the counterfactual problems were considered to be the most complex and ultimately unsolvable. The comparison between 'experimental' and 'control' groups, a traditional approach in social studies, was not an option "since there are no control cities. All of the big cities in Britain, to greater or lesser degree, have been the recipients of urban policy" (Robson *et al.*, 1994: 5). Furthermore, it was also difficult to distinguish and balance between the various reasons that might account for failure.

8.3.2 Sample and Methodology

The research was carried out at both the national and local level using a sample of 123 authorities which was divided into three categories:

1. all of the 57 Urban Priority Areas (UPAs),

2. 40 authorities which had either been included earlier in parts of the Urban Programme or whose socio-economic circumstances were little different from the UPAs (**marginal**)
3. 26 authorities which have never been part of the targeted focus of urban expenditure (**comparator**) (Robson *et al.*, 1994: 14)

All of the 123 authorities "showed some degree of socio-economic distress, but some of which had received more and some had received less financial assistance" (Robson *et al.*, 1994: 14). The three conurbations chosen for detailed quantitative analysis, and on which the qualitative analysis was based, were Greater Manchester, Merseyside and Tyne and Wear.

The Team designed a multi-stranded research methodology to address the above mentioned difficulties (Robson *et al.*, 1994: vii):

1. A quantitative analysis aimed at establishing the relationship between expenditure and socio-economic outcomes (input-outcome analysis);
2. Qualitative information derived from a formal questionnaire of a sample of residents and discussions with a sample of businesses in the three conurbations;
3. Qualitative information derived from discussions with experts, in the three chosen conurbations

The quantitative analysis used a variety of forms, from highly sophisticated multi-level modelling, to more mundane analyses of the changing gaps between the 57 UPAs and the rest of the 123 LAs and between inner city wards and the rest of the conurbation (Bradford and Robson, 1995: 47). Policy inputs were measured in expenditure at the district level whereas outcomes were measured in five indicators (see below).

The residents survey involved a total of 1299 interviews (Robson *et al.*, 1994: 33), whereas the employers survey involved a "small number" of businesses in each of the three conurbations (Robson *et al.*, 1994: 35). The experts survey involved some 30 to 40 interviews in each of the three conurbations (Robson *et al.*, 1994: 16). Both residents and employers surveys aimed "to investigate views about current socio-economic conditions in the conurbations; to see to what extent these conditions had altered over the recent past and were perceived as likely to alter in the immediate future; to identify whether government policy had impinged on its intended recipients and, if so, whether the experiences of those recipients supported or refuted the evidence culled both from ... statistical analysis and from the views of experts" (Robson *et al.*, 1994: 33). The aim of experts survey was "to explore perceptions of the areas' problems, the ways in which policies have been implemented and the impacts of those policies on the conurbations over the last decade" (Robson *et al.*, 1994: 37).

8.3.3 Outcome Measures

The terms of reference gave particular salience to the AfC package (Robson *et al.*, 1994: vii). The core of the research was thus based on the most explicit statement of AfC objectives spelled out in its original document:

- to encourage enterprise and new business, and help existing businesses to grow stronger;
- to improve people's job prospects, their motivation and skills;
- to make areas attractive to residents and to businesses by tackling dereliction, bringing buildings into use, preparing sites and encouraging development, and improving the quality of housing; and

- to make inner city areas safe and attractive places in which to live and work (Robson *et al.*, 1994: 3)

The published objectives of 30 of the 33 programmes originally comprising the AfC were examined. This showed that there were well over 100 such programme objectives. These were grouped into a *lower-level* set of ten principal objectives and a *higher-level* set which identifies two principal objectives:

- I. the creation of employment opportunities; and
- II. the creation of cities which are more attractive places in which to live. (Robson *et al.*, 1994: 6)

On the basis of these two sets five socio-economic indicators were identified (see Table 8.1).

8.3.4 Discussion

Despite the Team's efforts to ensure the validity of the research and both its findings and conclusions there remain several crucial issues that may undermine both. It is, however, believed that this is due, first and foremost, to the very ambitious aim of the research. These concerns can be grouped under two main headings: outcome indicators and the research methodology.

Outcome Indicators

Outcome indicators and their choice, coverage of and relation to the objectives raise several important issues. First, in selecting the indicators, the Team opted for "relatively simple" socio-economic variables for which data were relatively accessible (Robson *et al.*, 1994: 8). Simple measures, no doubt,

Table 8.1: The Socio-Economic Indicators

INDICATOR	Objectives Addressed		Data Sources	Limitations
	Higher-level	Lower-level		
1. Unemployment and long-term unemployment, 1983-91	1	1,2,3,4,&6	DE	1. Frequent definitional changes 2. The difficulty of deriving appropriate denominator for calculating rates from the raw data
2. Net job changes, 1981-89	1	1,2,3,4&6	Census of Employment	1. Being work-place based, it fails to distinguish whether beneficiaries are city-based or commuters
3. Percentage change in the number of small businesses, 1979-90	1	1,2,4&5	VAT registration records	1. Only businesses with a sufficient turnover to warrant VAT registration are captured 2. It cannot distinguish, in firm deaths, between new firms of short life from those of longer duration nor does it isolate deaths caused by displacement from those of competition
4. House price changes, 1983-90	1&2	2,6,7,8,9&10	Building Society	1. As it reflects the intersection between supply and demand, local knowledge of number and type of available properties is required to make sense of price movements 2. It reflects only those houses that have come to the market 3. Price movements of this subset may not be typical of the whole stock 4. It necessarily excludes housing outside the private market 5. Most problematical, robust and interpretable data are notoriously difficult to collect
5. Net change in the number/proportion of 25-34 year-olds	2	1,7,8,9&10	OPCS mid-year population estimates	

Source: compiled by the author drawing on material from Robson *et al.*, 1994: 9-10.

facilitate the interpretation and communication of results. Nevertheless, it is questionable whether they can accurately reflect complex contexts and issues. The lack of information also led to changes in each indicator to be examined over varying time periods (Robson *et al.*, 1994: 25). Can a consistent, overall picture be drawn upon inconsistent comparisons? However, problems of the availability, accessibility and comparability of data are issues that no one research can be criticised for.

Second, there is the correlation between these indicators and the objectives of the AfC, the core of the research. Although there is a great deal of overlap between these indicators and the four explicitly stated objectives of the AfC, it remains a task for the reader to establish which indicator(s) relates to which objective(s) of the four. The concluding discussion of the research was however drawn on the basis of the two higher-level objectives.

Third, the coverage of these indicators, of any set of the objectives, is somewhat partial. For example, none of the indicators provides information on the 'strength' of existing business (AfC obj. no. 1). Nor do they inform on the change in people's motivation or skills (AfC obj. no. 2). (These issues were not covered either in the qualitative analysis). An increase in the number of people obtaining jobs does not necessarily result from an increase in the number of those seeking job. Furthermore, it is not clear how a change in the number of small businesses (indicator no. 3) can measure, as claimed, inter-agency co-ordination (lower-level obj. no. 5).

Fourth, there is the interpretation of the changes in any of these indicators (apart from the input-outcome relationship). For instance, an increase in house prices may sometimes indicate a housing problem, whether or not the area is becoming more attractive. In effect, not to examine outputs alongside outcomes may be considered a mistake. In this instance, the number of houses built may help explain the change in prices, and in cases can be an indicator of area's attractiveness.

Finally, and in broader terms, there are two more important issues: the number of indicators employed and the relative weight given to each. There is room to argue for the use of more than five indicators. The multiplicity of programmes and the diversity of their objectives and activities all justify the call for more than five measures of policy impacts. The Team quoted data availability, consistency and comparability as one, evidently major, reason why other indicators were not included. We may add both time and resources constraints as another.

The analysis and its conclusions clearly draw equally on each of the indicators. Again, there is room to argue otherwise. Some of the indicators may reflect policy objectives closer, or better, than others. Although all the programmes of the AfC package may share the same broad goal, some have definitely placed more emphasis on particular objectives than others. When evaluating these programmes, the indicators closely reflecting their objectives may be given more consideration than others. That is, some indicators may

be assigned more relative weight than others. This, however, calls for a 'finer-grain' research that looks at individual programmes and areas separately; something apparently beyond this research.

In sum, and to put it in the Team's own words: "the indicators are far from perfect; nor would we claim that they provide a comprehensive overview of the impacts of policy" (Robson *et al.*, 1994: 8).

The Research Methodology

In general terms, the multi-stranded methodology of the research, in itself, is an attempt to overcome the limitations of its individual elements. A quantitative analysis alone is not an adequate approach to gauge policy impacts, not least because of problems of quantification. On the other hand, a qualitative analysis may not meet all the information requirements of all interested groups. It lacks, by definition, quantification - an appealing feature for many decision-makers. In effect, an evaluation research that draws on both approaches is likely to answer more questions and meet more information requirements than a purely quantitative or qualitative approach (though it is also likely to face more problems). However, each of the two methodological strands of the research had its serious limitations.

1. The Quantitative Analysis

- *Sampling*: At the national-level, the research was based on quite a comprehensive sample of areas. To an extent, as the Team claimed, this

sample provided the basis to tackle some of the conceptual problems (Robson *et al.*, 1994: 14):

- Counterfactual: the geographical spread allows some comparison between areas that faced some degree of urban distress but have received varying amounts and kinds of support;
- Contiguity: the swathes of continuous areas allows the examination of spillover effects;
- Choice: the sample allows examination of the relationship between amounts of expenditure and outcomes regardless of whether areas are 'policy-on' or not.

However, other, equally important, conceptual problems have received little, if any, consideration. The Team, at the outset of the research, recognised the fact that

The economies, environments and social structures of cities form a seamless interconnected web and the effectiveness of expenditure on any one of these distinctive elements of policy clearly intersects with the impact of expenditure on many of the other policy instruments (Robson *et al.*, 1994: 1)

They were also well aware, as quoted before, of the possible effects of other public policies, especially given the proportionally little expenditure on regeneration. The impact of other policies (the confound problem) and different local contexts and circumstances (the contextual problem) have been given some consideration. The same, however, cannot be said with regard to the interaction between the elements of the policy or the combination of programmes through which assistance was delivered (the combinatorial problem). Most probably, this is due to the very ambitious scope of the research which makes such a task virtually impossible. For one reason, the wide geographical coverage of the research and the vast number of localities

involved. For another, the multiplicity of policies and their different combinations that might have well varied over time within the same area.

At the local-level, there is no given or apparent justification for this particular sample (the three conurbations), though that does not undermine the valuable information these case studies yielded. The fundamental concern remains, however, as to the representativeness of this sample of all the conurbations, and of all the urban areas. The Team explicitly admitted

Clearly, the areas were not intended to be representative of the conurbations as a whole; because of this, the results of the survey cannot - and were not intended to - represent views across each conurbation (Robson *et al.*, 1994: 33)

The inherent danger here is the tendency, on the part of policy-makers, to generalise these findings perhaps even to the national level. Despite the Team's clear warning, there is simply no guarantee that this would not be the case.

- *The Analysis:* The Team summed up the limitations of the quantitative analysis as follows:

the amounts of public resource cannot be a very sensitive measure of policy on the ground; the outcome indicators are a very imperfect and partial measure of socio-economic change; treating all of the component instruments of policy as equal parts of policy delivery does not recognise that different combinations in different circumstances can work better or worse; the input/outcome relationships act as black boxes which cannot tell us anything about the underlying processes or mechanisms of change. (Robson *et al.*, 1994: 16)

Apart from the outcome indicators, we may elaborate further on these limitations. Firstly, the total public expenditure on the AfC package represented only 2% of total government expenditure (Robson *et al.*, 1994: 20).

It was concluded that this "merely reinforces the concern about the 'confounds' problem; that the impacts of AfC may have been swamped by the effect of other non-targeted public resources" (Robson *et al.*, 1994: 20). However, no attempt seems to have been made to take the effects of those 'other policies' into account. Whether impact, or lack of it, can be attributed solely to the AfC, other policies or both remains unclear. This largely undermines the causal relation between input and outcome which was the main focus of the research. There is no way to ascertain the validity of any of the findings.

Another question is whether or not financial assistance in itself is enough to achieve impacts. The answer, most likely, is 'No'; or else, why may a policy succeed in one area and fail in another? Contextual circumstances and local characteristics vary, sometimes considerably, from one place to another. Among possibly many other factors, this is very likely to influence the performance of a policy and its impacts.

Secondly, in addition to the effect of different combinations of programmes, which has not been tackled either, there is the relative weight of each programme. Some programmes may have targeted more resources or lasted for longer, or both, in an area. In this case, such programmes are likely to have more impact than others. This calls for relative weighting of programmes based, among other things, on their resources and duration. That is, to investigate individual programmes and their combinations in each area,

separately. Again, this was something beyond the scope of the research.

Thirdly, the input-outcome relationships are subject to diverse interpretations. A negative relation may suggest either a slow rate of an area's response or an ineffective policy. The relation in itself is not, by any means, enough to judge its own direction. To do so, those "black boxes" have to be opened; the processes and mechanisms through which policies have been delivered have to be examined. That is, implementation analysis, also known as process evaluation (see, for instance, Alterman, 1987). The experts survey may indeed have shed some light on the process. However, one can still argue for implementation analysis as a distinct, yet integral, component of the evaluation. The scale and complexity of the implementation process mean that a full understanding of the outcome demands a closer examination of the procedures and mechanisms involved.

2. The Qualitative Analysis

- *The Recipients Survey*: A major concern with this survey, as was with the local-level quantitative analysis, is the representativeness - and, consequently, the generalisability or otherwise - of the sample chosen for the study. First, the three conurbations do not represent the whole array of conurbations, neither were they intended to (Robson *et al.*, 1994: 33). Second, the size of the sample, both of residents and employers, appears less than adequate to represent fully the views of either group within the three conurbations, not to mention nation-wide. The residents survey involved a total of 1299

interviews (Robson *et al.*, 1994: 33), whereas the employers survey involved "a small number of businesses" (Robson *et al.*, 1994: 35).

Moreover, it has to be borne in mind that the residents surveys were aimed at a specific section of the society with a particular socio-economic profile. These surveys were conducted "in some of the very worst of the inner-city areas"; different views may have emerged had these surveys been conducted in other areas (Robson *et al.*, 1994: 48). This also applies to the employers survey. Furthermore, it should be noted that employers survey was undertaken at a time of deep recession "where most employers were preoccupied by macro economic conditions and that such concerns tended to swamp the evaluation of the more subtle impacts of policy interventions" (Robson *et al.*, 1994: 35).

These issues combined together cast shadows over the validity of conclusions and their generalisability. Nevertheless, both surveys provide a valuable insight into people's and businesses' perceptions of government policies and their impacts, at least within their local level.

- *The Experts Survey*: There is not much disagreement with the experts survey. It is equally important to gauge the views of decision-makers on the performance of policies as it is to assess people's perceptions. The findings of this survey are quite important, and of major relevance. However, it might have been better to spread the survey over a wider geographical area, if only

to ensure the validity of the findings.

In more general terms, the research founders on two further accounts. First, although it may have pointed to failure, the research fails to account for its reasons. Failure, it was reckoned, can be attributed to one or a combination of the following elements:

theory failure (the incorrect identification of the genesis of problems and hence the development of an inappropriate set of policy instruments); implementation failure (the inadequate or ineffectual translation of policy into practices on the ground); or measurement failure (the absence or inadequacy of data or of techniques for measuring the impacts of policy). (Robson *et al.*, 1994: 4)

No measure has been taken within the research to tackle any of the above.

And again, this appears to be due to the very wide scope of the research.

Second, and in our view, the main weakness of the research is the insufficient attention paid to the mechanisms behind the outcomes. The input-outcome relations, the intersection between policy elements themselves and between them and both other policies and varying contexts and, the reasons behind failure all will remain 'black boxes'. And the key to these boxes is implementation analysis. In the words of Alterman (1987: 348):

Many of the problems with traditional evaluation research arise from its focus on the assessment of outcomes and its concern with providing a causal relationship, while ignoring the process that produced these outcomes.

A causal relationship between inputs and outcomes was exactly what the research aimed at, without sufficient consideration of the mechanisms behind these changes.

8.4 INTERIM EVALUATION OF INNOVATIVE POLICY - CITY CHALLENGE

8.4.1 The Initiative

City Challenge (CC) was launched in May 1991. In its first round (1991/92), fifteen local authorities were invited to draw up programmes for action to tackle their key neighbourhoods. Having said that, the (then) Secretary of State for the Environment made it clear that LAs were expected to attract private finance and thoroughly involve the private sector in managing the programmes. Eleven authorities were successful in that round, each will receive £37.5 million over a five-year period. The second round of City Challenge (1992/93) was open to all 57 Urban Priority Areas (UPAs). In the event, a further twenty councils were successful so that there is now a total of thirty-one CC authorities.

City Challenge has certainly a number of distinctive features. After a decade of antagonism to local government, it has clearly strengthened the role of local authorities in urban regeneration — though others argue that this has also involved relinquishing overall control to arms-length boards or trusts (Oatley and Lambert, 1995: 142). This strategic role is, however, conditional on plans to be developed and delivered through partnerships between the local authorities and all other stakeholders in the area, including public, private and voluntary sector bodies and the local community. The initiative also marked a clear shift from the major underlying principle of inner city

policy throughout the 1980s, namely 'trickle down'. There is an emphasis on re-integrating disadvantaged areas and groups into the mainstream economy of the city. The initiative encouraged a more strategic approach to regeneration that attempts to tackle problems in an integrated fashion linking projects dealing with economic, environmental and social concerns. After a decade of project-based, annually allocated, incentives, CC called for a five-year strategy and action plan.

Perhaps the most controversial feature of CC (and the SRB as well) is its competitive bidding process. The government claimed that the stimulus of competition would transform the way in which local authorities and their partners approached the task of urban regeneration. "Competitive bidding was seen as a way of promoting an entrepreneurial culture in local government and as a way of producing bids which conformed to the government's objectives of creating innovative approaches to economic and social development through partnerships which institutionalised the influence of a wider set of actors, most notably, those in the private sector" (Oatley and Lambert, 1995: 142). However, it is "probably still too early to start drawing sensible conclusions about the effectiveness of competition but we can say that it encourages both bold strategies and bold claims for the prospects of future success" (Burton and Boddy, 1995: 35). (see for instance, Davoudi and Healey, 1995: 158; Hambleton and Thomas, 1995: 6; and Russell, 1994b: 4).

8.4.2 Monitoring and Evaluation Guidelines

The DoE issued several guidance notes on project appraisal, financial arrangements and monitoring and evaluation of CC projects and strategies. At the outset, a distinction was made between monitoring (checking inputs and outputs of individual projects and aggregate outputs for strategic objectives), periodic review (to assess achievements and to check whether corrective action is needed) and evaluation (to test, at intervals, whether the action is having the desired impact; are the strategic objectives delivering the vision of CC?) (ICGD, 1993, para 3). A distinction was also made between output measures and impact/outcome indicators (ICGD, 1993, para 10).

Partnerships were required to monitor their projects regularly (ICGD, 1993, para 8). They were also required to devise and establish their own information systems - with a clear emphasis on ethnic monitoring of projects (ICGD, 1993, para 17). They were expected to show, over the five year period of CC, "what progress has been made towards achieving the vision and strategic objectives" (ICGD, 1993, para 19) (core impact indicators were identified - see Appendix E). The first priority was given to establishing, and regularly updating, the baseline position, which "should help determine impact indicators" (ICGD, 1993, para 23, 24, stress in original - see Appendix F). The main responsibility of monitoring and evaluation was laid on the Partnerships with the DoE's (and its regional offices') role limited to the monitoring of:

- expenditure
- key milestones and aggregate outputs
- core impact indicators
- a sample of projects on the ground (ICGD, 1993, para 28).

8.4.3 National Evaluation of City Challenge

Evaluating an initiative such as City Challenge is extraordinarily difficult, not only because its objectives are wide-ranging and its focus unstable and diffuse. Its 'targets' primarily focus on making relationships, for example, establishing links between training places provided and routes to jobs, or active involvement in developing and managing projects in order to build the confidence to seek training or a role in formal politics. *This suggests that policy evaluation needs to take a qualitative and multi-perspective approach*, preferably over a range of time-scales. The tension between the government's narrow output monitoring regime for City Challenge projects and its encouragement of local evaluation teams looks set to head into a confrontation on appropriate evaluation approaches (Davoudi and Healey, 1995: 163, stress added).

The DoE has already commissioned an interim national evaluation of CC (Russell, 1994b). The Team recognised, from an early stage, the multi-faceted nature of the initiative which requires a multi-stranded approach to its evaluation. The methodology of the study, thus, combines quantitative and qualitative research; "It is examining quantitative data and reviewing the basis on which City Challenges can be measured for their economy, efficiency and effectiveness. But it is also looking at less or non-quantifiable aspects, in particular to see whether City Challenge is changing the relationships between different sectors and making a difference to the way in which cities are governed" (Russell, 1994b: 7).

This evaluation will follow CC in its four main stages: design, implementation, programme progress and programme impact (Russell, 1994b:

1). The first report concentrated on the first two stages; much of the report's

focus was on process issues. It is believed that in order to evaluate the outcome of CC "and understand the reasons for whatever impact it has – the process of partnership and project development must first be tracked and understood" (Russell, 1994b: 5).

The research draws on interviews with Chief Executives of CCs and, in some areas, other team members, board members and local evaluators. It also draws on interviews with a range of community representatives and DoE regional officials. The Team has also visited a local authority which failed to win CC but is trying, nonetheless, to pursue its plans (Russell, 1994b: 1).

The features of CC examined in this evaluation were those specified in the research brief as key propositions to be tested. These were:

1. Partnership: The concept, the diverse forms, measuring partnership, parity of partners and the role of different partners.
2. Competition: Criteria for selection, bidding and the losers.
3. Targeting: Which areas? and who benefits?
4. Time limited: Certainty and continuity, annuality, vertical take-off and exit strategy.
5. Output driven: Judging performance, DoE management, suitability of the management framework, the range of concerns, process evaluation and explaining change.
6. Economy, efficiency and Effectiveness.

A sample of 14 case studies has as well been chosen to further examine these features; "Are these appropriate and desirable features which are

conducive to good practice?" (Russell, 1994b: 43). The case studies were selected on the basis of a number of criteria that reflects the variety of CCs in operation. These included:

1. Date of start (first or second round) and regional coverage
2. CC context: Type of LA, area population, area size and presence of other programmes (eg UDC, EZ, EPs)
3. Type of programme: Area- or people-based, degree of risk in the strategy, action plan ingredients and ethnic minorities
4. CC organisation and structures: Relations with LA, monitoring systems, company status and chief executives
5. Partnership arrangements
6. DoE ranking (following the mid-year reviews).

Work has also been, and still is being, carried out on the quantitative component of the research (inputs, outputs and outcomes). Unfortunately, no information on this side of the evaluation is available.

8.4.4 Local Evaluation of City Challenge

Some CC partnerships have appointed external bodies to carry out their local evaluation. The same institution conducting the national evaluation is actually involved in one of those local evaluations as well. There is a great deal of similarity in the approach adopted in both cases. An interview was held with a member of the local evaluation Team (May, 1995) and a senior officer within the CC Team (August, 1995). The discussion sheds interesting light on many issues and concerns about the evaluation of CC.

The evaluation Team saw this particular evaluation exercise as a departure from past experience. Not only were they looking at ex-post evaluation (ie outputs and outcomes), but also "action evaluation". There were two elements to satisfy:

1. The Government's requirements about outputs and establishing a good baseline against which subsequent progress can be monitored.
2. The "softer" questions in which the CC Team was more interested. These questions related in the main to the policy process. For instance, were they bringing in partners in an integrated way; were they setting up appropriate consultation arrangements; were they defining objectives in a meaningful way and were they up-dating them, etc.

No one approach to evaluation was seen to be ideal. Instead, the methodology adopted was double-stranded: quantitative and qualitative analyses. On the quantitative side, the central government's output measures were taken as a starting point and then supplemented by what the CC Team saw of particular relevance (to the strategy and its objectives and/or to particular partner(s)). Throughout the Study, the Team made a distinction between outputs and outcomes. It was recognised, however, that outcomes will materialise probably after the life time of the CC.

One particular problem was immediately faced. That is, to establish the baseline position. The study had not commenced until about 7 months into the project. The study Team had, thus, the task of trying to assemble available material and impose some sort of structure on what to be collected. Almost inevitably, there remained gaps that either required extra resources to cover, or could not be covered at all. It is to be noted that other CC Teams who set

up beforehand monitoring and evaluation systems could take a much more sensible judgement of what were the gaps and what was available.

The counter-factual problem was treated in a rather qualitative way. At the outset, the Team considered using an 'experimental' approach; to compare the CC area with a 'similar' one but not subject to CC. However, this approach was rejected. It was not possible to identify such a similar area. Instead, the Team looked at whether there has been a convergence between the CC area's conditions and those pertaining to the city as a whole. It was believed that if the comparable position of the CC area has, over time, changed dramatically, one can then get an order of magnitude that something significant has taken place. Beyond that, there is a need to talk to those closely involved with the property market and social situation to get a 'softer' impression of how things changed on the ground. The Team realised that this was not an ideal approach, but it was seen as the most meaningful measure of additionality and that any quantitative assessment would be more questionable.

In addition, the qualitative analysis relied on information gained through interviews and discussions with a variety of actors who were involved in the process and "sort of cross-compare their perceptions." The discussions related to both the additionality of CC and, more important for the CC Team, the policy process itself and how it evolved over time.

8.4.5 Discussion

The experience of evaluating CC, at both the national and local level, raised a few more issues of wider concern. First, the evaluation team member maintained that although official arrangements were made for baseline studies to be conducted before the start of CC, in practice – because of shortage of time and obsession with getting things done – questions of evaluation tend to be put aside. In the first bidding round of CC, LAs had only six weeks to draw their plans and negotiate partnerships. The winners were then under pressure to start spending and implementation from the first day. This led to tensions between setting up partnerships and procedures and producing more tangible outputs quickly.

The CC officer, somehow, seems not to share the same view. It was emphasised that there has been a firm commitment to evaluation, particularly process evaluation right from the outset. The reason why an external body was appointed was the lack of clear government guidelines as to what is meant by evaluation. CC Team felt that there is a need for an independent, external body to conduct the task. Nevertheless, it was admitted that building up monitoring and evaluation systems from the beginning was quite a difficult task. It took the CC Team a long time to learn how to do them and they are still learning how to do them well.

Second, there were concerns with data and information, particularly availability, compatibility and replicability. The evaluation team member's

view is that it would have been preferable had the study commenced before the start of CC. Monitoring, evaluation and information systems could have then been put in place and a more accurate baseline been established. In any event, data collection and establishing and maintaining a data base has been, and still is, a costly and time consuming task for the CC Team.

Third, there were problems of measurement, most notably of concepts like 'quality of life'. In the first place, there are the problems of definitions and thresholds. It was also difficult to establish a baseline for this indicator because the CC area, in the local evaluation, had a very small population. The study Team opted for retrospective evaluation; to track back the situation by asking people if and how the area has changed (and also determine where they initially come from).

Another set of problems was the changing objectives and shifting projects between different headings. For instance, after about a year-and-half some transport projects were shifted under the heading of economic infrastructure.

Finally, there was the question of measuring 'success' and achievements. The starting point was whether targets have been met; the study team was much aided by the quantified targets of the project. The member's view is that outputs are a perfectly valid measure of success if they were set against wider contextual evidence. The team, however, was always keen to assess the experience against the key propositions of the initiative: partnership,

integrated approach to evaluation, etc, and whether these have materialised on the ground and to what extent.

Some conclusions and recommendations of the evaluation were described as "painful". Nevertheless, the relationship between the two teams has been one of co-operation, and the evaluation as a whole is regarded as 'excellent' and of great use and relevance to the working of the CC team. The reason, apparently, is the emphasis given to the qualitative, process evaluation of CC which is of crucial importance to the CC Team. Many of the recommendations has actually been put into practice changing the working of the CC team.

8.5 CONCLUDING REMARKS

There is no doubt that the three case studies in between them raise several issues of wider relevance to evaluation research in general, and of urban policy in particular. At the outset, the three case studies put together demonstrate clearly the lack of a systematic approach to ex-post evaluation of urban policy. Each of the three studies approached the task from a different perspective and each adopted a different methodology. This is at the time when they all had the same objective: to assess the impact of the policy in question. The differences between the three studies are made clear when the issues they raise are considered one at a time.

Urban policies are usually launched against, and to offset, existing negative trends. A new policy is usually introduced against a background of previous and current government policies in various fields. These existing trends and policies will no doubt have an influence on the operation of the 'new' policy. More important, changes on the ground are likely to be the outcome of more than one single factor or policy. In other words, it is important to disentangle the effects of prevailing trends and other policies from those of the policy at hand. Equally important, and related, is to estimate what would have happened in the absence of the policy in question - the counter-factual position. That is, it is important to establish a causal relation between observed changes/impacts and the policy at hand; which of these impacts are the result of this particular policy and which are not.

It is not the statistical relationship *per se* that is the issue here. What is of concern, if the impact of a policy is to be assessed, is to establish, beyond reasonable doubt, which of these impacts is solely attributable to this policy. Failure to establish this relation will lead to either an over- or under-estimation of policy impacts. If other policies' impacts are attributed to the policy at hand, this is an over-estimation of its effects (and under-estimation of others'). On the other hand, policy impacts may have been offset by negative trends, or other policies that ran counter to it.

The problems are no doubt further compounded the larger the targeted area/population becomes. This is due, in part, to the varying conditions and trends, and the problems, prevailing in each of the parts/groups of this area/population. It is also due, consequently, to policies that may already be in operation in response to these problems and trends. Nevertheless, that should be no reason why the counter-factual problem, and the causal relation between the policy at hand and changes on the ground, are not to be addressed. Indeed, it can be argued that it is more reason to consider, more carefully, such issues rather than spurious conclusions of achievements.

The three case studies varied dramatically in their treatment of this problem. At one extreme, the first case study simply did not address this issue. A senior member of the evaluation team, as mentioned before, considered such a pursuit fruitless. At the other extreme, the second case study used rigorous statistical techniques and models in attempt to establish

this causal relationship. In between these two extremes, the local evaluation of City Challenge opted for a somewhat qualitative approach to assess its additionality. While this approach avoids the difficulties of statistical analyses, it is wide open to subjective judgment and several errors associated with surveys and questionnaires. Whereas it is important to gauge people's perceptions of change and causes of change, it is equally important to ensure the validity of any conclusion. In turn, it seems appropriate to use some degree of statistical analysis to disentangle the impact of the policy from other exogenous factors.

Displacement, and its bearings on equity considerations, is another crucial concern in evaluation. The positive achievements of a policy at one area may indeed be offset by its negative impacts on a neighbouring locality. New jobs created in one area may be at the expense of firms closure in another. Both issues actually further complicate the task of evaluation. If both are to be taken into account, a wider scope of the task has to be adopted. In addition to the resultant work-load, there are two other problematic concerns. Firstly, to identify, as accurately as reasonable, the area over which the policy has had its impacts (in terms of either community groups or geographical area, or both). Secondly, if an overall picture of the policy impacts is to be achieved, its effects on those different communities have to be aggregated somehow. This amounts to weighting different groups of the society against each other.

Again, the three case studies differ widely in handling any, or both of the two issues. The first case study virtually ignores both. The second case study, by encompassing a wide range of LAs, has claimed to have accounted for displacement. There is no explicit consideration of equity, though the study, by focusing on aggregate impacts at targeted areas, has eventually overlooked this concern. As for the third case study, there is no indication available to us as how equity and/or displacement may have been/are being tackled. Nevertheless, the initiative of City Challenge is quite distinct from previous policies in explicitly focusing, from the outset, on disadvantaged groups and areas. It appears to be tackling equity "head on" through some sort of "positive discrimination". Yet, it is crucial to assess whether benefits to these groups/areas have been at the expense of others.

Assuming that policy impacts have been assessed, it then remains to judge whether these constitute an achievement of its objectives. The success in this task is entirely reliant on the manner in which these objectives have been stated in the first place. Objectives have to be stated in a way that provides a yardstick, or a benchmark, against which movement in any direction can be judged. Unfortunately, it appears to be the norm that policy goals and objectives are so vaguely stated that any 'objective' assessment of their achievements is virtually impossible.

The evaluation of the PFV is a 'classical' example where objectives are stated in such a vague fashion that any assessment of achievements is

virtually impossible. The evaluation relied, in the main, on 'before - after' comparisons. In assessing the impacts of government policies in the 1980s, as in evaluating City Challenge, a further problem was encountered. That is, the change and shift in policy objectives over time. Combined with the difficulties of estimating time lags accurately, assessing the achievement of objectives may prove almost impossible.

Perhaps surprisingly, the vague statement of one particular objective appears to represent no problem. That is, the key proposition of City Challenge (and the SRB) of 'Partnership'. Process evaluation, in the case of City Challenge, appears to have no difficulty in tracking achievements in this regard. The only possible explanation is perhaps the fact that this type of evaluation is reliant on qualitative information derived from interviewees who provide their perceptions of success; in a sense, doing the evaluation.

Closely related is the choice of indicators to assess the performance of the policy. On the one hand, policy objectives should provide the basis on which these indicators can be identified. That is, objectives have to be stated in an operational way which facilitates the definition of these indicators. This, however, may have its drawback in inducing policy-makers to 'dilute' the substance of policies by focusing on easily measured aspects. It may as well induce them to under-estimate their targets in order to guarantee success in advance. On the other hand, and especially in the absence of such operational objectives, performance indicators should reflect as closely as possible policy

objectives; they should be the closest proxy possible of these objectives not simply what is measurable or available. This is very likely to require further research both to identify such indicators and, consequently, to secure their data and information requirements.

The first two case studies have, to varying degrees, foundered on the difficulty of identifying the "most appropriate" performance indicators. The government guidelines, in the case of City Challenge, provide a starting point, though they do not cover "process" outputs. These guidelines, however, are being supplemented by local City Challenge Teams. The lack of information prohibits any conclusion in this respect.

Any call for 'better-defined' performance/outcome indicators is very likely to require further research both to identify such indicators and, consequently, to secure their data and information requirements. Questions are bound to arise in regard to the justifications of both the requirement to state objectives in a measurable manner and further research on performance indicators and data sources. The answer is, indeed, quite simple. Government policies, over the last sixty years or so, have had almost the same substance and, on average, have had minor impact (or else, how to explain the fact that the latest policy is addressing the same problems of the 1930s?). If a clearer picture of both the achievements and failures of past government policies was ever available, and if such a picture was used in formulating new initiatives, it is very likely that substantial amounts of public expenditure would have

been saved, and problems would have eased quite a lot.

In broader methodological terms, the experience of City Challenge is perhaps the most important among the three. The 'innovative' features of CC (partnership, strategic and integrated approach to evaluation, targeting disadvantaged groups and areas and, competitive bidding) are all key propositions of the Single Regeneration Budget. It seems that these features are set to underpin future policies. Even if not, an evaluation of the SRB will no doubt benefit from any assessment of those same concepts.

One crucial concern that emerges clearly from both the documents on the evaluation of CC, and the discussion with members of the two teams, is that of "process evaluation". The CC team member asserted that it is of crucial concern to them to assess, on regular basis, the working of their partnership and the co-ordination and co-operation (or otherwise) between the different partners involved in the project. It should go without saying that the involvement of several actors in the initiative will sometimes give rise to conflict; compromise rather than consensus is likely to be the norm. In other words, the outputs/outcomes are no doubt the result of a highly politicised process.

An assessment of outputs/outcomes alone will yield a partial picture. Indeed, if City Challenge was to be a catalyst for regeneration efforts and if it were to achieve its objective of encouraging a partnership approach to

regeneration, it follows, the evaluation of how things are being done is as crucial as what has been done. Evaluation of such initiatives as CC or SRB will, thus, have to focus equally on the process through which the initiative is being delivered as much as it focuses on its outputs and outcomes.

Another concern is the aggregate impact of such policies at the national level (a major concern for central government). The aggregate output/outcome of individual projects is only one element of any such assessment. However, any comparison between individual projects in terms of these outputs/outcomes is "unfair" (CC officer, personal contact, August 1995) and, indeed, misleading. Individual projects differ widely in their scope, nature, the base-line conditions they started from and the problems they addressed. More important, they differ in their organisational arrangements - their partnership workings. Performance assessment of individual projects has to take account of the forces behind this performance as much as a national evaluation will have to bear these different characteristics in mind.

CHAPTER 9:
ASSESSING THE IMPACT OF
REGIONAL POLICY IN BRITAIN

9.1 INTRODUCTION

Chapter 8 has indicated that ex-post evaluation of urban policy lacks a systematic methodology and faces a host of difficulties. However, evaluation research is not confined to the field of urban and regional planning. It has been, and is being, applied in several other disciplines. The question then arises whether similar problems have been encountered in other fields, and if so, how they have been overcome, if at all. In other words, does experience in other fields offer any lessons that are transferable into urban policy evaluation? To answer this question, a choice has to be made as to which field(s) to be reviewed.

Prior to the current emphasis on urban regeneration and inner cities initiatives, regional (economic) policy was a predominant feature of government intervention programmes. This is probably one of the longest-standing government programmes; its beginnings can be traced back as early as 1920s. Regional policy has been the subject of some fifteen years of ex-post research in an attempt to estimate its effects (compared with virtual absence of ex-ante appraisal of policy packages and instruments). This well-documented research is one reason to choose the research on the effects of regional policy as one of the fields to review.

Another, equally important reason is the economic drive increasingly

underlying urban regeneration initiatives. The first of the SRB objectives centred around employment prospects; the second emphasised economic growth and wealth creation (see chapter 7, p. 185). City Challenge, the predecessor of SRB, has also emphasised economic growth and regeneration. A characteristic feature of the central paradigm within which policies are being shaped is the belief that the market is the most important factor in determining what needs to be done (Aaronovitch, 1995⁽¹⁶⁾). On several occasions, economic criteria have taken precedent over social ones.

This similarity in objectives is the second reason why the experience of assessing the effects of regional policy has been chosen for review and comparative analysis. It is believed that this experience will have its implications for urban policy evaluation.

This chapter, then, is a review of the experience of assessing the effects of, and evaluating, regional policy in Britain. It aims to find out the pros and cons of different methodological approaches in order to inform future conclusions on comprehensive ex-post evaluation of urban policy. The chapter starts, in section 2, with a very brief account of the historical origins of the policy and how it is apparently becoming an important tool in structural reforms at the EU level. Despite its long history and the substantial resources

⁽¹⁶⁾ Aaronovitch, S (1995) *The Evaluator - The Policy Maker: How Do They Connect?*, a presentation at the "Evaluation of Inner City Regeneration Projects", seminar held at LEPU, South Bank University, London, 6 June 1995.

devoted to it, research on the effects of regional policy has focused on a limited number of indicators. These indicators and the reasons why they were chosen are explained followed by a review of the evaluation approaches that dominated the field: the micro (industrial surveys) approach and the macro approach.

Section 3 covers what has come to be known as partial econometric methods. These methods attempt to isolate the effects of regional policy on changes in one or more of the indicators, eg employment growth and industrial movement. The partiality of these methods stems from their focus on a limited number of indicators only within the assisted-areas. Partial methods have been used to assess the effects of either the policy package or its individual instruments. In the first instance, a modified shift-share analysis was the method most commonly employed. Regression analysis was the technique used to disaggregate policy impact into the likely effects of its individual instruments. Section 3 reviews, in brief, the main characteristics of both approaches.

In contrast, comprehensive evaluation attempts to estimate the effects of regional policy from the viewpoint of the society as a whole; ie it attempts to measure the effects of regional policy on all sectors/groups within and outside assisted areas. Two approaches could be identified of comprehensive evaluation: statistical modelling and cost-benefit analysis. Cost-benefit analysis could be further classified into two categories: social cost-benefit analysis and

the less-comprehensive Exchequer approach. Research in this direction has, however, been very limited; only three attempts of comprehensive evaluation could be traced. These are reviewed in section 4.

The macro approach (industrial surveys) is the second major approach to assessing the effects of regional policy. It employs questionnaires, interviews and surveys to measure the micro level effects of regional policy on the 'micro units', ie industrial firms. The type and nature of questions have undoubtedly differed from one study to the other. However, five main questions could be identified. These, and the main findings of this research, are outlined in section 5.

Given the primary objective of this chapter, section 6 focuses on two issues. Firstly, the pros and cons of different methodological approaches to assessing the effects of regional policy: modified shift-share analysis; regression analysis; industrial surveys and, comprehensive evaluation. In addition to problems associated with each of them, all have faced several other difficulties: measuring the changing strength of policy; identifying precisely periods of active and passive policy; time lags; the differential impact of policy instruments on different recipients; the choice of indicators and, measurement of impact.

The second issue upon which section 6 focuses is the implications of this experience for research on ex-post evaluation of urban policy. A distinction

is first made between 'evaluation' as defined within this research and the experience of assessing the effects of regional policy. The latter, it is argued, is better placed within the category of output assessment, rather than evaluation. Nevertheless, there are several issues which remain equally important for urban policy evaluation. First, the treatment of the counterfactual problem. The balance of argument is considerably against shift-share analysis. Regression analysis is perhaps a more appropriate alternative, though it has to be borne in mind that qualitative information cannot be incorporated in such analysis. Another crucial concern is distributional effects. Unfortunately, the experience of regional policy does not offer any resolution to the several problems associated with equity considerations.

It is argued that regular, systematic monitoring represents probably the best resolution to the difficulties associated with measuring the changing strength of policy. Combined with regression analysis, both seem to offer a useful tool to experiment with time lags. The choice of indicators is another crucial concern for both fields. The limited number of indicators, and their relation to policy objectives (or rather, lack of it) leave plenty of questions unanswered about the achievement of regional policy objectives. It is finally argued that implementation analysis should be an integral component of evaluation if a better understanding of policy performance, and hence better informed decisions, are to be achieved.

9.2 REGIONAL POLICY IN BRITAIN

9.2.1 A Brief Background

The regional economic policy in Britain is probably the longest-standing government intervention programme

Economic fashions have waxed and waned, yet regional policy has remained on the statute books for over sixty years (Armstrong and Taylor, 1993: 193)

The beginnings of a regional policy in Britain can be traced back to the late 1920s. The persistent overcapacity and high unemployment in major export industries led to the existence of unemployment "black spots" due to the geographical concentration of these industries. The government response was to set up the Industrial Transference Board in 1928 which aimed to facilitate workers migration to more prosperous regions through direct grants and loan assistance. This policy had a minor effect relative to the magnitude of the unemployment problem in the mid-1930s (eg the unemployment rate in Wales was 37.8% of the total 'insured' workers in 1933).

This situation induced the introduction of the Special Areas Act, 1934 (and amendments 1936 and 1937). Under the 1934 Act, four areas were designated as Special Areas which benefited from (limited) government assistance to industry. Although these powers were extended in 1937, the overall expenditure was quite limited with the result that very few jobs were actually created. Yet, this initiative marked a shift of policy towards a principle that underlined regional policy ever since. In contrast with the Industrial Transference Board, the initiative was based on the principle of "taking work to the workers", not the other way round.

The turning point in Britain's regional policy came in 1944 with the government's commitment to full employment (the White Paper on Employment Policy, 1944). The main thrust of the policy was to create manufacturing jobs in the newly established Development Areas – DAs. A variety of policy instruments were introduced through the Distribution of Industry Act 1945, with varying levels of success. These included: grants and loans to firms, powers to build factories and establish industrial estates and, the provision of basic services for industry.

Since then, numerous changes have taken place. The boundaries of the assisted areas, their types and categories and the incentives they benefited from have all been the subject of several alterations, and at times abolished. Policy instruments also varied in the timing of their introduction, in their nature and in the strength with which they were applied. At the risk of oversimplification, Appendix G provides a brief chronology of the British regional policy.

The predominant view, despite all these changes, is that the main objective of regional policy has been to reduce inter-regional unemployment disparities; its major thrust has been to induce manufacturing industries to move into assisted areas through a package of incentives and controls (eg Moore and Rhodes, 1973a: 87; Diamond and Spence, 1983: 20 and Armstrong and Taylor, 1993: 214).

However, at the time when government commitment seems to have waned, regional policy is increasingly becoming a prominent feature of the EC reforms towards the Single European Market (Armstrong and Taylor, 1993: 290-291). The principal components of the policy came into existence in 1975 with the establishment of the European Regional Development Fund (ERDF). The policy was reformed in 1979, 1984 and 1988. The EU regional policy has three main components:

1. The ERDF;
2. Other instruments and policies with a built-in regional element (eg ESF, EIB and ECSC);
3. Co-ordination policies which aim to ensure that the regional impacts of other EU policies are taken explicitly into account (eg competition policy and Community Initiatives).

9.2.2 The Choice of Criteria

Since the prime objective of regional policy has been to reduce regional disparities in unemployment, it seems logical to use unemployment data to assess whether regional policy has been successful. It should be noted, however, that the effects of regional policy on job creation may not be fully reflected in unemployment data. "Indeed, unemployment may be rising in a locality even though regional policy is successfully creating new jobs" (Armstrong and Taylor, 1993: 331). For instance, indigenous industries in a locality may be declining faster than new jobs are being created.

In addition, there may be more subtle effects at work as a consequence of

changes taking place on the supply side of the local economy. Net outward migration of a locality, for instance, may slow down when regional policy is successfully creating new jobs. This will have the effect of increasing the labour supply in the locality over and above what it would otherwise have been in the absence of regional policy. Regional policy may also affect the supply side of the economy through changes in the participation rate of the population of working age. As a consequence of improvements in the job market, more people tend to seek work. "Some of these will register as unemployed, and so push up the unemployment rate in exactly the same way as a fall in net outward migration" (Armstrong and Taylor, 1993: 331-2 — see also Moore and Rhodes, 1973a: 90; Diamond and Spence, 1983: 21-22 and Moore *et al.*, 1986: 13).

Another indicator that could possibly be used is the changes in total employment in DAs relative to the national changes. However, this is not a suitable indicator. Regional policy has been concentrated in the manufacturing sector which accounted for only about one-quarter of total employment. Employment changes in other sectors of the economy may offset and thus disguise any impact of the regional policy (eg large job losses in coal mining in DAs in the 1960s). (eg Moore and Rhodes, 1976b and 1977; Diamond and Spence, 1983: 23 and Moore *et al.*, 1986: 13).

Since regional policy was directly applied to manufacturing industries, evaluation research has focused on indicators which measure the performance

of this sector. These were: manufacturing employment, industrial movement to the DAs and industrial investment, output and productivity in the DAs. Unfortunately, there is no information as how EU-supported programmes (eg Merseyside 2000) will be evaluated. The appraisal of projects submitted for funding has, however, placed a clear emphasis on output measures and 'value for money' (Personal contact, 1995).

9.2.3 Evaluation Methodologies

Two methodological approaches characterised research on the impact of regional policy on such variables: the micro approach (industrial survey) and the macro (econometric) approach (eg Ashcroft, 1978: 7; Bartels *et al.*, 1982: 9-10 and, Armstrong and Taylor, 1993: 330). The micro approach employs questionnaires, interviews and surveys to measure the micro level effects of regional policy on the recipients of financial assistance (industrial firms). The econometric approach, on the other hand, uses statistical techniques both to predict the behaviour of those variables which the policy sought to influence and to exclude the multitude of other influences that affect the dynamics of those variables (though with varying levels of success, as will be explained later).

Diamond and Spence (1983: 32) classified the methodologies of the macro approach according to their analytical complexity, as follows:

1. Indirect methodologies: These include studies that employ simple statistical techniques to identify a relationship between changes in the

intensity of regional policy and the movement in particular indicators. A major characteristic, and weakness, of such a type of analysis is the absence of any rigorous standardisation to exclude the multitude of other influences that affect the dynamics of these indicators so that a more precise policy effect can be identified. Another major shortcoming of these methods is that they do not provide a proof of a regional policy effect. They merely suggest that one might exist.

2. **Partial methods:** These methodologies were identified from the first category because an attempt is made to isolate the specific effects of regional policy on one or more of the indicators (employment, industrial movement, ...) by means of more rigorous statistical techniques. The partiality of these methods stems from their focus on a limited number of indicators only within the assisted areas.
3. **Comprehensive methods:** These methodologies attempt to estimate the benefits and costs of policy intervention from the view point of the nation as a whole. That is, they aim to measure the effects of regional policy both on other economic sectors within assisted areas and on other non-assisted areas.

It should be noted, however, that the distinction between the first two categories is not as clear cut in practice as it may appear to be. Many of the 'partial' studies took, as their starting point, an 'indirect' method. In most

cases, a time-series analysis of the variable under consideration was examined relative to the changing strength of regional policy over the same period of time. When such analysis indicated a relationship between the change in both the variable and the intensity of regional policy, these studies moved to more rigorous analytical techniques which fall within the second category of methodologies; the partial methods (eg Moore and Rhodes, 1973a and 1976a and, Ashcroft and Taylor, 1977 and 1979).

9.3 PARTIAL ECONOMETRIC METHODS

Narrowly defined, the objectives of regional policy emphasised the use of controls and incentives to balance the demand and supply of labour by safeguarding existing jobs and creating new jobs in the DAs. "The major tactic of such policies has been to attract industry to the assisted areas" (Diamond and Spence, 1983: 43). Therefore, a measure of the impact of regional policy may involve the analysis of its contribution to industrial employment growth, the movement of industries into DAs or, the level of industrial investment in DAs. Other indicators can be used if a broader view of the policy objectives was taken. If, for instance, the objective was to achieve self-contained growth in DAs at acceptable levels of income and employment, then such indicators as productivity and industrial output and profitability may be used (Diamond and Spence, 1983: 43). However, very little work has been conducted on the impact of regional policy on industrial output and productivity in DAs (and usually using the same methodology of modified shift-share analysis, eg Tyler *et al.*, 1979).

9.3.1 The Effects of Regional Policy

The names most closely linked with the evaluation of regional policy effects on manufacturing employment are those of B. Moore and J. Rhodes (eg 1973a and b, 1974, 1976a and 1977). In their work, they distinguished, quite rightly, between the influence of regional policy and other trends on a region's industrial base. The aim was to illustrate how much manufacturing (in terms of employment, movement or investment) in the assisted areas would have changed in the absence of regional policy. It was therefore necessary to establish an 'expected' or 'hypothetical' series that reflects the effects of all factors that would influence a region's manufacturing base, but excludes the effects of regional policy. Under certain assumptions, it was then assumed that the difference between the actual (A) and expected (E) series can be attributed to the effects of regional policy.

The expected employment and investment series were derived through a modified version of shift-share analysis (eg Moore and Rhodes 1973a). The 'non-policy' series was that which would have occurred had the assisted areas witnessed the same national trends (on an industry-by-industry basis) rather than the policy. For industrial movement into these areas, the alternative series was based on the actual number of moves between 1951 and 1958 since it was assumed that regional policy was largely in abeyance during this period (Moore and Rhodes, 1976a: 19). The same approach was used by Rees and Miall (1979) to analyze data on industrial investment and capital stock in assisted areas.

Modifications were later brought to the basic approach (ie 1973a). To account for past trends, the expected series was then based on extrapolating a linear trend from the policy-off into the policy-on period. This trend was a line of 'best fit' through the deviations from the zero of the 'A-E' series. The difference between this trend and the 'A-E' series then represented the cumulative effect of regional policy (eg Moore and Rhodes, 1974).

The differential cyclical movement in the expected and actual employment series was also accounted for (Moore *et al.*, 1977). Cyclical adjustments were made on the basis of regressing the annual change in actual manufacturing employment on the change in expected employment and the pressure of demand measured by the UK unemployment rate.

The impact of North Sea oil developments (on the Scottish manufacturing sector) was also considered (Moore *et al.*, 1977). This was based on an estimate of the actual number of jobs directly attributed to this development.

Another improvement was to distinguish between the 'indigenous' and 'immigrant' firms (Moore and Rhodes, 1974). The aim was to show that the greater part of the observed effect could be attributed to an acceleration in the inflow of immigrant firms into Scotland (while employment in indigenous firms continued to decline). This distinction was maintained henceforth, though for a different reason. That was the recognition of the fact that some policy instruments were expected to influence primarily the building of new

factories (ie the immigrant sector) while others should have a bigger influence on indigenous firms (see Moore and Rhodes, 1977: 27).

Despite these improvements, the work of Moore and Rhodes (and others who adopted the same approach) faced several practical and conceptual difficulties (as will be discussed in detail in section 9.6.1). Most of these difficulties related to the methodology used; the modified shift-share analysis. In addition, Moore and Rhodes came to reckon that their earlier work had two limitations (Moore *et al.*, 1986: 15-16). Firstly, the impact of the policy had been assessed only up to the mid-1970s. Nothing was known about the impact in the later 1970s or the recession that began in the 1980s. Second, insufficient consideration had been given to the methodological difficulties which arose when regional policy had been operating for a number of years. This refers to the 'plateau' effect; as regional policy continued to operate and policy jobs built over the years, there came a point where the first policy jobs went into natural decline. The longer regional policy had been in existence, the more the job losses from earlier policy would be. Moore *et al.* (1986) introduced what they called the 'modified' method distinguishing between the *gross* and *net* impacts of policy.

This modified method shared a crucial aspect with the 'conventional' one. In estimating the impact of regional policy on the indigenous sector, the linear trend of the 'A - E' employment series was extrapolated from the passive into the active policy period. One difference was the definition of the passive

policy period, however. In the modified method, it was taken to be the period between 1951 and 1963 (1951-59 in the conventional method). Another difference was in the use of regression analysis, though very few details were given (apparently due to the particular 'audience' at which the study aimed). The same equations were then said to have been modified and used to estimate the effects of regional policy on the movement of firms into DAs.

9.3.2 The Effects of Individual Policy Instruments

The first attempt to illustrate the effect of one policy instrument, namely the Regional Employment Premium – REP, was that of Moore and Rhodes (1973a). It was noted that the deviation between the actual and expected employment series has increased after the introduction of the REP in 1967. To isolate the effects of REP it was necessary to establish how much of the increased deviation was a function of REP as opposed to other policy measures. "However, a very serious problem arises in connection with time lags" (Moore and Rhodes, 1973a: 101). Several time lags were used in sensitivity tests but it was admitted that there were no clear principles to overcome this problem (Moore and Rhodes, 1973a: 101-2). It was also indicated that there were other factors which may have clouded the results (eg the widening of the DAs geographical coverage and replacing investment allowance with investment grants in 1966).

Regression analysis was, later on, used to disaggregate the shift-share differential component (regional policy effect) into the likely effects of the

various policy instruments (eg Moore and Rhodes, 1976a and b and 1977 — see also Moore *et al.*, 1986). The dependent variable in their work was either manufacturing employment or the number of industrial moves into assisted areas. The independent variables were limited in range (maximum of five — Moore *et al.*, 1986), only one of which was a non-policy factor (usually male unemployment rate). However, variant specifications of the basic model were run to test for the effect of major changes in the policy (eg widening of the geographical coverage and the introduction of SDAs).

The technique of regression analysis was also used by others to disentangle the effects of various policy instruments. Ashcroft and Taylor (1977 and 1979) extended the approach of Moore and Rhodes to account for other variables (they maintained that its industrial *investment* that was on the move) using a total of three models. Begg *et al.* (1976) used a simple model to obtain objective assessment of the timing of improvement in Scotland's investment.

However, these models have all been limited in their scope. Apart from the 'pressure of demand' no other non-policy factor was taken into account in the analysis (Ashcroft and Taylor (1977 and 1979) used indicators of industrial investment and output but still failed to account for any other influences).

9.4 COMPREHENSIVE ECONOMETRIC EVALUATION

A prerequisite of any comprehensive evaluation is to measure the whole range of effects of the policy package (direct and indirect) in terms of the whole society (assisted and non-assisted areas; the industrial as well as other economic sectors) (Armstrong and Taylor, 1993: 356). In the case of regional policy, this is apparently a demanding and daunting task that verges on the impossible (Diamond and Spence, 1983: 76). It is not surprising, then, that very few attempts have been made to 'comprehensively' evaluate the net social gain of regional policy.

9.4.1 Cost-Benefit Analysis

Attempts to measure the overall effects of regional policy in Britain have relied on cost-benefit analysis. Two approaches can be identified in this realm: 1) social cost-benefit analysis which adopts the viewpoint of the nation entire, i.e. all regions and multiplier effects and, 2) the less comprehensive Exchequer approach which adopts the Treasury's point of view.

Social Cost-Benefit Analysis

To date, there has only been one attempt to evaluate the real costs and benefits of regional policy, using CBA, from the viewpoint of the whole nation. Schofield (1976) proposed the following model to estimate the net present value of the Distribution of Industry (D of I) policy:

$$\text{Present Value Benefits} = \sum_{r=1}^n \sum_{t=1}^m \frac{W_{rt} \beta_{wt} + P_{rt} \beta_{pt} + A_{rt}}{(1+i)^t}$$

$$\text{Present Value Costs} = \sum_{r=1}^n \sum_{t=1}^m \frac{D_{rt} + G_{rt} + M_{rt}}{(1+i)^t}$$

where

- W = labour income due to direct job creation under the policy.
- P = capital income due to direct job creation under the policy.
- β_w = national multiplier for labour income.
- β_p = national multiplier for capital income.
- A = real resource costs avoided as a result of inter-regional labour migration forestalled under the policy.
- D = expenditure of factory construction due to the policy.
- G = real resource costs (other than those included in D) incurred as a result of public expenditure under the policy.
- M = private movement costs incurred under the policy.
- i = discount rate.
- n = 1 ... n regions.
- t = 1 ... m years over which costs and benefits run.

Schofield (1976: 183), having applied the model, admitted that it excluded several crucial factors. For instance, regional policy may have (most likely it does) an impact on the balance of payment and hence on income. Such factors, however, were judged to be relatively insignificant. The policy was analyzed from the perspective of only one objective: economic efficiency. That is, there is no consideration of 'equity' or distribution impacts (between the different regions). Moreover, the model as a whole, and many of its variables, were the subject of considerable approximations in computation due to severe lack of data.

The Exchequer Approach

Another more common, but less comprehensive, approach to assess the costs and benefits of regional policy was in terms of its Exchequer costs and revenues using government data. Moore and Rhodes have also been involved

in estimating the net Exchequer costs of regional policy (Moore and Rhodes, 1973b, 1975 and 1977). They maintained that the real resource costs of regional policy were negligible because it brought into use resources that would otherwise have remained idle. This is because regional policy diverted demand from prosperous to depressed regions which allowed an expansion of output and employment without incurring an increase in the inflationary pressure in the country.

The exchequer outlays on regional policy were divided into two categories:

1. Recoverable: These include expenditure on factory building (through rent), loans and reduced income from accelerated depreciation schemes.
2. Non-recoverable: These include grants, expenditure on infrastructure and administration.

The items of income or revenue included tax gains resulting from increased activity and savings in unemployment benefits. The loss in tax receipts was deducted if it was necessary to maintain aggregate pressure of demand in prosperous regions once activity has been diverted to DAs. Savings that may have resulted if regional policy prevented migration into non-DAs (thus resulting in lower infrastructure and public savings in expenditure) were also included.

On the basis of these views and assumptions, Moore and Rhodes concluded that regional policy had actually resulted in a net gain to the Exchequer over the period 1963-70 (eg Moore and Rhodes, 1975: 92).

However, their estimates have to be interpreted with caution (Diamond and Spence, 1983: 84). The distinction between recoverable and non-recoverable outlays is not always as clear cut as it may appear. Certain costs are not included in their estimates, eg infrastructure expenditure in DAs to complement financial incentives and private movement costs. It is also questionable whether all the jobs diverted to DAs can be replaced in non-DAs by government demand management policies.

9.4.2 Statistical Modelling

In contrast to the CBA approach adopted within the British context, Folmer (1980) developed a linear structural equation model to measure the effects of the Dutch regional policy. His point of departure was the problems associated with partial methods which use only one equation. These problems he grouped under the following headings (Folmer, 1980: 1192-1993):

1. *Handling of the multidimensional nature of the regional profile,*
2. *The separation of policy effects from autonomous developments, and*
3. *Handling of theoretical constructs and measurement errors.*

The model he developed is a simultaneous equation model with separate equations for all profile elements under study and for all intermediate policy targets. The model also includes "all the important explanatory variables, whether autonomous or policy variables" (Folmer, 1980: 1193).

For purpose of illustration, the model was applied to one factor:

employment growth in the building material industry in the Netherlands, 1974-75. This means that the model can be applied to measuring the effect of regional policy from the viewpoint of the nation as a whole; at least for one affected variable. In principle, it can be applied to all variables at the national level, though the indications are that it will be a very complex task. Folmer (1980: 1201) has also recognised that the model requires a huge quantity of information at a low geographical order. He failed to reckon, however, that the validity of the results is largely dependent on the identification of explanatory factors. This is a problematic issue and there is no "systematic" method to ensure that *all* variables have been defined. In addition, he failed to explain how "measurement errors" have been taken into account. The sheer volume of observations needed to run the model at the national level simply makes matters worse.

9.5 THE MACRO APPROACH (INDUSTRIAL SURVEY)

Several studies have been undertaken to assess the effectiveness of regional policy by directly asking the firms which have actually received assistance whether these have had any effect on their operations (eg Moore and Rhodes, 1976b and 1977; Allen *et al.*, 1986 and Wren, 1988). The main objective of these surveys was to obtain qualitative information on the impact of regional incentives on a sample of firms. This information was then used to build up a picture of the impact of regional policy in general.

The type and nature of questions differed from one survey to another, yet

five main questions could be identified (Armstrong and Taylor, 1993: 348):

1. Do firms receiving investment incentives incorporate these into their investment decisions?
2. To what extent do investment incentives affect a firm's level of investment, output, employment and productivity?
3. Which type of investment incentives do firms regard as being the most effective?
4. To what extent are investment incentives 'deadweight'?
5. What is the cost per job of investment incentives?

Although the studies differed widely in their research specifications (sample selection and size, questions, ...), a review of several independent undertakings revealed a considerable degree of consistency in their findings (Begg and McDowall, 1987: 467). The overall picture was that regional incentives were important, or even crucial, in the investment decisions of firms in DAs. Yet, there is less of a consensus over the relative importance of individual policy instruments.

9.6 CONCLUDING REMARKS

9.6.1 Assessing the Effects of Regional Policy

Regional policy in Britain has been the subject of quite intense research on its effects, for some fifteen years (early 1970s - mid-1980s). A distinct characteristic of this research is its ex-post orientation. This comes in sharp contrast to a complete absence of any reported ex-ante appraisal of the policy package or its various instruments. This, in itself, contrasts with the experience of a field such as motorways and trunk roads development (as will be seen in chapter 10). The only possible, and perhaps surprising, explanation

is that this ex-post research has been triggered by political interest.

The work of Moore and Rhodes (1973a) is regarded as the seminal piece of work in this field. It is probably the most cited research in the field. To a large extent, this work was the turning point in the development of research on policy effects. Meanwhile, and at the same year, the House of Commons Expenditure Committee (Trade and Industry Sub-Committee) set out to find an answer to the question:

what effect incentives and restraints on investment decisions in industry have had on the distribution of unemployment in particular but also other factors which lead to the imbalance between the regions which successive governments have been committed to remedy (HC, 1973, para 1)

The Committee observed that

There must be few areas of Government expenditure in which so much is spent but so little known about the success of the policy (HC, 1974, para 116)

They concluded that they were

far from satisfied that the continuing search for a viable regional policy has been backed by a critical economic apparatus capable of analysing results and proposing alternative courses (HC, 1974, para 170)

It may have been a mere coincidence of timing and it may have not. The issue is admittedly open to speculation. The fact remains though that, in contrast to several other fields, regional policy has been the subject of a considerable amount of ex-post research on its effects.

The surplus of research that followed from, and including, the work of Moore and Rhodes (1973a) faced several conceptual and practical difficulties. From its earliest stage, this research has explicitly addressed the counter-

factual problem. In their classic, oft-cited work, Moore and Rhodes (1973a) aimed to illustrate what would have happened in the absence of regional policy. A 'hypothetical' series was always established reflecting the effects of all other factors. The effects of regional policy were then the difference between this expected series and the actual one. This was a characteristic feature of their work, and of others. The concept is certainly applicable in the evaluation of other types of policy.

Nevertheless, the method Moore and Rhodes (and others) used to establish this 'expected' series — modified shift-share analysis — raised considerable criticism. From a conceptual viewpoint, and being a variant of shift-share analysis, the method was subject to major criticism of the technique (eg Buck, 1970: 446; Paris, 1970: 491; Stilwell, 1970: 453-455; Richardson, 1978: 19 and 1979: 206 and, Schofield, 1979: 256-7). This criticism centred around the following issues: the inter-dependence between the differential shift components and being highly unstable over time; the policy conclusions drawn from analysis; the choice of weighting system; sensitivity of the results to the degree of industrial aggregation and, information reliability. Richardson, for instance, (1979: 206) concluded that

The only substantive result to emerge from hundreds of applications of shift-share is that industry-mix alone does not explain regional growth differentials.

Another major problem with the technique is that it does not reflect changes in the industrial mix over the study period (Richardson, 1978: 19). It

simply assumes that the industrial structure of the DAs, at the beginning of the active policy period, remained the same over the whole period. No allowance was made for any possible changes that might indeed have taken place as a result of regional policy (see also Rees and Miall, 1979: 8).

Moore and Rhodes, on the other hand, maintained that "these criticisms do not necessarily apply with force to the use of the technique for disentangling the effects of policy" (Moore and Rhodes, 1977: 21). The reason was that the analysis aimed at comparing the differential employment performance at times of passive and active policy rather than measuring the absolute figures. Additionally, the analysis had its proponents. Fothergill and Gudgin (1979: 310-17), for instance, challenged the severe criticism of the technique and concluded

It seems ... that the widely held misgivings about the shift-share are not strong enough to seriously affect its application to the analysis of regional growth in the UK.

More importantly, the extrapolation of the policy-off trend into the policy-on period raised what Armstrong and Taylor (1993: 284) considered the most serious criticism of the method as a whole. "It is hard to accept that the steady downward trend in the A - E series witnessed in the 1950s would simply have continued through 1960-81 in the absence of regional policy" (Armstrong and Taylor, 1993: 284, stress in origin). The further we move away from 1950 the more likely that changes had occurred that invalidate the use of the 1950-59 trend in the A - E series to establish the policy-off position.

Moore *et al.* (1977: ff. 73), on the other hand, maintained that for this concern to be correct other, non-policy factors should have intervened to change the 1950s trend in the A – E series in the active policy period. "Apart from the impact of North Sea Oil and the political troubles in Northern Ireland there is no evidence that other factors could have generated a large regionally differentiated effect on manufacturing employment in either direction" (Moore *et al.*, 1977: ff73). However, Armstrong and Taylor (1993: 294) argued, quite rightly, that considerable economic and political changes had occurred since 1960 which are very likely to have affected the attractiveness of assisted areas – eg membership of the EC in 1973.

The improvements Moore and Rhodes later brought to the basic methodology were also faced with criticism. The use of the unemployment rate as a proxy for the pressure of demand had been challenged. National unemployment rate rose dramatically in the early 1970s. Yet, other pressure of demand variables had not experienced such a decline. "Consequently, it is thought that there were special factors which caused the unemployment rise (such as the substitution of capital for labour) and thus the variable is a poor measure of the pressure of demand" (Diamond and Spence, 1983: 47). Furthermore, the estimate of the North Sea oil development related jobs was later found to be an inaccurate one (Diamond and Spence, 1983: 47).

The modified method (Moore *et al.*, 1986) was not much different from the conventional one, and thus is prone to the same criticism. More importantly,

Moore *et al.* (1986: 34-6) reckoned that being crucially dependent on the extrapolation of the negative trend of the A – E series, the method faced two major difficulties. Firstly, little was known about the factors responsible for the negative trend in DAs in the absence of the regional policy. Secondly, it was insufficient merely to identify possible factors behind the regionally differentiated performance of manufacturing employment in DAs. It was necessary to estimate which of these factors have changed between the passive and active policy periods. Several factors could have affected the performance of manufacturing employment in the absence of the regional policy. These include, (1) differences in profitability of operating in the DAs compared with other parts of the country (due to differences in unit input cost or productivity)⁽¹⁷⁾, (2) the availability of labour and, (3) the acceleration in the urban/rural shift after the 1960. "The problem is that the different factors do not all work in the same direction and there are very severe difficulties in quantifying their relative impact" (Moore *et al.*, 1986: 36).

Regression analysis was the technique used to disentangle the effects of different policy instruments. The credibility of any results of the technique, however, hinges on one particular condition: a comprehensive definition and accurate measurement of all the factors that are likely to have had an effect on the observed change(s). Within the context of regional policy, this condition has simply not been fulfilled. The variables entered into the analysis

⁽¹⁷⁾ This contrasts sharply with their earlier argument that there was no evidence that industrial costs could have fallen relatively in the DAs (Moore and Rhodes, 1973a: 100-101).

have been very limited in range. Moreover, usually only one non-policy variable has been taken into account (often male unemployment rate). It was assumed that shift-share analysis has already eliminated the effects of other factors. The criticism mounted against the shift-share method serves only to undermine such an assumption. This, in turn, casts more shadows over the results of regression analyses.

"In principle, *controlled experimentation* is possible at the micro level" (Bartels *et al.*, 1982: 10, stress in original). An advantage of this research strategy, therefore, is in isolating causal relations between policy instruments and outcomes. Another advantage is that "those to whom the incentives are directed are asked for their views about the effect of these incentives on their own operations" (Armstrong and Taylor, 1993: 347). This provides some information on the possible role of policy instruments (Bartels *et al.*, 1982: 12). The results of micro studies can also contribute to a better understanding of the decision process; they can be helpful in the measurement of the intensity of certain policy instruments and they can yield information on the resource costs of regional policy (Bartels *et al.*, 1982: 13).

However, despite the valuable information and insights it may yield, the macro approach has its drawbacks. "The major drawback of such surveys is that so many factors affect a firm's operation that they may not know themselves what effect incentives have had" (Armstrong and Taylor, 1993: 347). Thus, it is not possible to obtain 'quantitative' answers to such questions

as 'what would have happened in the absence of policy?' (Moore and Rhodes, 1977: 29). It is not possible either to ask firms to attribute certain decisions to policy and then add up the effects of those decisions; regional policy operated as a contributory influence on most decisions (Moore and Rhodes, 1976b: 191).

The approach has its 'technical' difficulties as well: the representativeness of the sample and the interpretation of answers (Bartels *et al.*, 1982: 13); formulating questions in a way that does not influence answers (Ashcroft, 1978: 7); and it only reveals part of the possible impacts, those that relate to the respondents in the period when question were formulated (Bartels *et al.*, 1982: 14).

Hence, these studies should be treated with caution "since it is neither statistically nor theoretically sound to make generalisations about the present and future behaviour of U.K. industry from survey information based on past experience" (Diamond and Spence, 1983: 57). Moore and Rhodes (1976b: 191) contended that the statistical approach is, therefore, "the only potentially viable way of making a comprehensive evaluation of the effects of regional policy."

In Britain, attempts at more comprehensive evaluation (Schofield's model and the Exchequer approach) have relied on cost-benefit analysis – CBA. The conceptual and practical difficulties of CBA have all been detailed before (see chapter 3). Within the context of regional policy, the major limitations of cost-

benefit analysis stem, perhaps not surprisingly, from the multiple and vague objectives of the policy. A major thrust behind the policy has always been to re-distribute economic growth between the various regions, which is an explicit social equity objective. Meanwhile, regional policy had, at several stages, its national efficiency objective in the form of the drive to enhance national economic performance. This raises the oft-cited conflict of equity vs efficiency. CBA, by concentrating on economic efficiency, cannot pertain to resolve such a conflict. Furthermore, not all policy objectives can be quantified, let alone measured in money units, to be incorporated into the analysis.

In an inter-regional system, Richardson (1978: 31) argued that the benefits of one region are frequently the costs of another. This suggests that a comprehensive evaluation of the aggregate economic welfare requires an explicit set of regional weights by means of which income gains and losses to various regions can be valued and weighed against each other (Haveman, 1976: 450).

The first implication of such a requirement is to estimate policy impacts on *all* regions. Such an estimate is simply unavailable. Research on the effects of regional policy has been almost entirely confined to DAs (except for Tyler, 1980 which covers only one non-DA region). A further defect with existing estimates of employment effects of regional policy is that they have been confined to jobs created in DAs. A comprehensive evaluation, be it CBA or

another, should be concerned with the nation as a whole. The multiplier effects which spill over to other regions should be included in the analysis (Armstrong and Taylor, 1993: 359). In addition, the multiplier effects, within DAs themselves, on other sectors of the economy should as well be included. It is likely that jobs have been created in other sectors of the DAs' economy as a direct result of regional policy. As we have seen, research thus far has concentrated on manufacturing employment⁽¹⁸⁾. Schofield's model (1976) went some way towards accounting for these multiplier effects. Nevertheless, the severe lack of data led to considerable approximations which cast doubts on the accuracy of the results. Moreover, nothing is known about the effects of the policy, by use of the model, after 1976. In other words, nothing is known about the overall NPV of regional policy.

The second implication is that of an explicit inter-regional weighting set. To put it rather bluntly, it may never be expected that decision-makers will make explicit such a weighting set. The result was that attempts to evaluate regional policy have simply ignored equity considerations and focused solely on economic efficiency, which clearly represents a lack of comprehensiveness.

In line with Richardson's argument, Oosterhaven (1983: 125-128) argued that an integrated interregional cost-benefit and input-output analysis is most adequate for policies which:

⁽¹⁸⁾ The same argument applies, of course, to other policy effects (investment, output, ...).

- focus on a single sector;
- have large economic indirect impacts;
- serve primarily regional interests;
- represent an investment project (in the sense that costs are incurred early while benefits are spread over the long-term); and
- are decided upon by central government.

There is no doubt that many of these features have been typical of British regional policy and, therefore, such an analysis could have been applied in assessing its impacts. The use of cost-benefit analysis is justified mainly because of the *investment* character of the project. Meanwhile, input-output analysis represents a means by which indirect effects and equity considerations (at the regional level) can be taken into account. In other words, both efficiency and equity concerns are treated explicitly.

There seems, however, to be no application of this approach to the effects of regional policy in Britain. Moreover, the primary difficulty this approach would face is information. As mentioned above, there is hardly any research on the economic implications of regional policy on non-assisted areas. The lack of detailed information on Folmer's model (1980, see p. 293) prohibit conclusions on its comprehensiveness, though it clearly faces the same major problem of securing all the data it requires.

Another difficulty in evaluating regional policy is to estimate the output foregone as a result of job creation in DAs (the opportunity cost or, resource cost). The complexity of the problem stems from the various sources of these jobs and the many ways in which they can be filled (Armstrong and Taylor,

1993: 358 - see Fig 9.1). There are three distinct ways in which jobs can be generated:

1. Diverting jobs from non-DAs,
2. Inducing foreign firms to move into DAs,
3. Creating entirely new jobs in DAs.

Jobs created by foreign companies or new jobs created in DAs entail no loss of output elsewhere and hence has an opportunity cost of zero. The same is not, however, true for jobs diverted from non-DAs. This would involve a social cost since output would be lost elsewhere in the economy⁽¹⁹⁾.

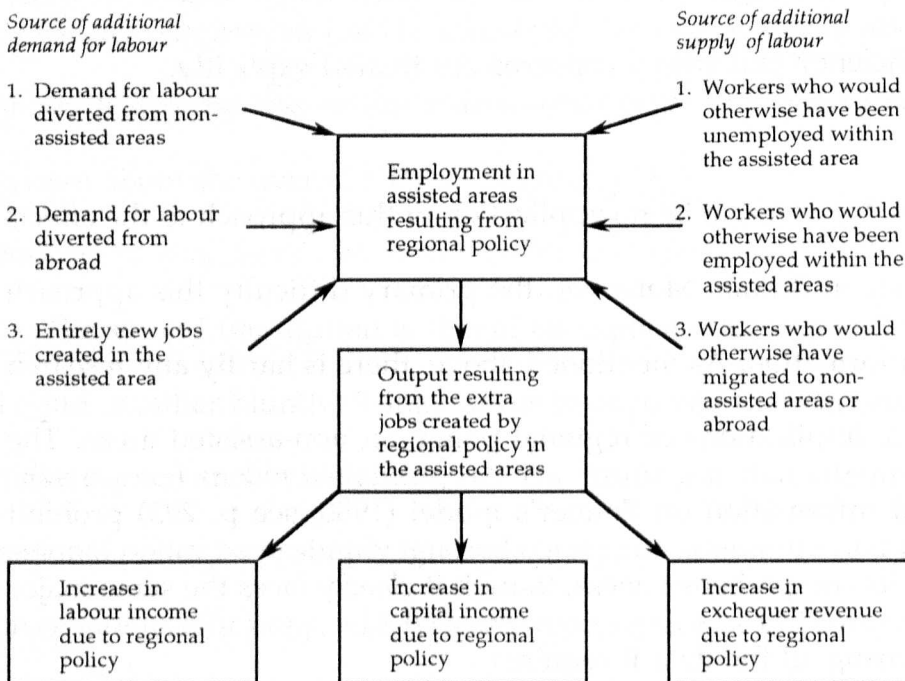


Fig 9.1: The Response of Labour Demand and Supply to Job Creation Schemes

Source: Armstrong and Taylor, 1993: 358.

⁽¹⁹⁾ On the national level, however, this may not lead to loss of output unless the productivity of the moving economic activity has declined after relocation. The equity concern remains the same in either case.

There are three sources of labour supply to fill job vacancies:

1. Otherwise unemployed workers,
2. Poaching workers from other firms within the DAs,
3. Workers who would otherwise migrated from DAs.

In the first case, the foregone output is zero. Poaching workers from other firms within the DAs may happen if unemployed workers lack the skills required by expanding firms. This may entail substantial social costs unless unemployed workers can be trained to fill the vacancies created⁽²⁰⁾. If vacancies were filled by unemployed workers who would have migrated into non-DAs and taken jobs there, this is a gain to DAs at the expense of non-DAs. If these workers would have migrated abroad, the foregone output is zero.

Increased output and forgone output are the most obvious social benefit and social cost, respectively, of regional policy. There is, however, a number of other benefits and costs that need to be identified and included in the analysis (see Table 9.1). Apparently, not all have been measured and not all can be quantified.

Regardless of its scope, or the technique it employed, research on the effects of regional policy faced other difficulties. One problem was to measure the changing strength of policy and precisely identifying periods of active and passive policy. Periods of passive policy have usually merged into periods of

⁽²⁰⁾ The problem may be transferred elsewhere if those firms started poaching workers from other regions.

active policy during a short transitional period. The situation was further complicated by two factors. Firstly, new policy instruments were usually introduced in addition to existing ones. Secondly, and at certain stages, some policy instruments were applied more strongly while others were weakened (eg Moore and Rhodes, 1976a: 17 and 1976b: 191-2; Mackay, 1976: 229 and, Moore *et al.*, 1986: 31).

Table 9.1: Principal Social Benefits and Costs of Regional Policy

Social Benefits		Social Costs	
1.	Additional output and income arising from increased economic activity.	7.	Output foregone as a result of productive resources being diverted from their employment elsewhere in the economy.
2.	Reduced infrastructure costs and costs of public service provision as a result of reduced migration from DAs to non-DAs.	8.	Costs of moving to a new location
3.	Avoided costs of migration to individuals who would otherwise have been forced to leave DAs	9.	Resource cost of constructing new factories
4.	Reduced urban externality costs (eg pollution, congestion, ...)	10.	Infrastructure costs associated with regional policy
5.	Equity or redistribution benefits of regional policy	11.	Administrative costs of regional policy
6.	Non-economic benefits (political, social, environmental)	12.	Damage to the environment

Source: modified from Armstrong and Taylor, 1993: 360.

Another problem was to establish the exact time lag of policy effects. When assessing the effects of any policy at a given point in time, there is an apparent need to establish which previous period of policy life should analysis refer to. This is a very serious problem and little is known in practice

about time lags. Almost all the studies reviewed have assumed a lag period of one year. Few have experimented with other assumptions (eg Ashcroft and Taylor, 1979), though the range of alternatives was very limited (1 or 2 years). Moore and Rhodes (1976b: 201) have found that it took an average of eight years or more for new factories to reach a mature labour force. Problems are further complicated by the facts that different policy instruments may have different lag periods.

The fact that the same policy instrument may have different effects depending on the way recipients put it into use further complicates the analysis. This was particularly the case with REP where firms may have used it to reduce prices as opposed to increasing profits or wages (Moore and Rhodes, 1976b: 212). In such a case, there is a need first to identify those different groups of the recipients and, second, to assess the impact policy instruments have had on each.

The choice of indicators was another difficulty. Although the major thrust behind regional policy was to reduce inter-regional unemployment disparities, unemployment rates were considered not a suitable indicator to measure policy effects (eg Moore and Rhodes, 1973a: 90 and 1976b: 192; Moore *et al.*, 1986: 13 and Armstrong and Taylor, 1993: 331-2). Consequently, little work has been done in this field and little, if any, is known about the effect of regional policy on unemployment in DAs. In a sense, that means that little is known whether the policy has achieved its prime objective.

However, the 'labour market accounts' (also known as 'labour market balance sheet' (Tyler and Rhodes, 1989)) is a methodology which examines the interaction between changes in labour supply and demand (Owen and Green, 1989: 69). This method estimates "the separate effect of demographic, economic activity, employment and migration change upon the imbalance between labour supply and demand in an area between two dates" (Owen *et al.*, 1984: 471-472). The method can be applied both in ex-ante (eg Tyler and Rhodes, 1989) as well as ex-post settings (Owen *et al.*, 1984) for a number of areas at once. Yet, there is no indication that the method has been applied within the context of regional policy in Britain.

Problems did emerge as well in measuring the indicators most agreed upon. In cases, the data needed were simply unavailable or unreliable. In others, the same indicator was measured in different ways by different analysts. The strength of the IDC controls, for instance, was measured in up to three different ways. The use of other indicators (particularly male unemployment rate as a proxy of the pressure of demand) has been challenged.

9.6.2 Research Implications

Regional policy, with its predominantly economic drive, is no doubt a relevant field to examine in search for 'transferable' experience into urban policy evaluation. The reason simply is the economic drive that is increasingly underlying 'urban' regeneration initiatives. As we have seen in chapter 7, the

SRB has clearly emphasised job creation and economic growth in its first and second objectives. City Challenge, the predecessor of SRB, has also emphasised economic regeneration. Economic development, it seems, is more and more seen as a 'precursor' to urban regeneration. This should come as no surprise. A characteristic feature of the (mainly central) paradigm within which urban policy is being shaped is the belief that economic criteria in general should take precedence over social criteria (Aaronovitch, 1995).

Due to this similarity in objectives, it is to be expected that the experience in assessing the effects of regional policy will have its strong implications for urban policy evaluation. At the outset, however, a distinction has to be made between the experience reviewed in the previous sections and 'evaluation' as defined within the context of this research. Evaluation is taken here to mean the systematic exercise that aims to answer the question: Has the policy achieved its goals, and why?

A comprehensive evaluation of regional policy would have had to embrace a wider framework that took account of its impacts on national employment, output and income, on the distribution of income and on the public finance (Moore *et al.*, 1977: 77). Such a view is justified by the national coverage of the policy. Regional policy aimed at reducing inter-regional unemployment disparities. It encouraged growth in some regions while restricting it in others. Thus, comprehensive evaluation would have had to measure the full range of social costs and benefits to all groups/regions

within the nation. To be meaningful, it would have had to relate to the objectives of the policy.

A crucial prerequisite of such an exercise, no doubt, is a rigorous assessment of the policy outputs. Outputs are the means to outcomes. Nevertheless, it remains yet to ascertain whether measured outputs have any bearing on achieving policy objectives. More importantly, it remains to establish the causal relation between policy actions and measured outputs. Are those outputs the result of the policy or are they the result of other forces at play? What policy actions have caused what changes on the ground? In other words, it is essential to establish what would have happened in the absence of policy rather than spurious conclusions about achievements, or otherwise. This is the counter-factual problem.

Although research on the effects of regional policy addressed this problem explicitly, it seems that it was assumed that the very few indicators used were a sufficient proxy for policy objectives. It appears that job creation and industrial movement and investment have all been considered good indicators of achieving policy goals. Nevertheless, no one has yet explicitly addressed the question whether sixty years or more of government intervention have had any impact on regional unemployment disparities - the prime objective of regional policy. Little effort has been expended to disentangle the impact of the multitude of factors and forces that have no doubt been at play over such a long period of time. What is known, to an extent, is the magnitude of

jobs created and generated moves and investment presumably due to regional policy. What is unknown is the impact of these on the economic, let alone the social, fabric of assisted areas. Equally unknown is what impact regional policy has had on other regions of the nation — the so-called 'more prosperous regions'. Even in pure financial terms, the net gain (or loss) of those sixty years of public expenditure remains an issue largely open to dispute.

Therefore, what has been carried out so far on the effects of regional policy in Britain has to be placed within the category of impact — or rather preferably, output — assessment. Nonetheless, the experience of some fifteen years of research still has important implications for urban policy evaluation. Indeed, if one were to substitute 'regional policy' with 'urban policy' in the previous sub-section, much of the discussion would still hold true.

The counter-factual problem is, no doubt, a crucial concern in the evaluation of urban policy. In assessing the effects of regional policy, modified shift-share analysis was the method employed to address this problem. The balance of argument is, however, considerably against the method. Another, more satisfactory alternative has to be found. A potential candidate is regression analysis. The (great) difficulty of identifying and measuring *all* other extraneous factors should not be seen as an inherent deficiency of the technique. This, in fact, is a problem of evaluation rather than the technique used. Nevertheless, regression analysis has another particular drawback. That is, the difficulty of incorporating qualitative data

into the analysis. Not all the effects of urban, or any other, policy can be measured in quantifiable terms. It remains a concern to incorporate such effects into the analysis in a way that ensures both the robustness of evaluation and that these effects are not overlooked, or relegated into a lower category of importance.

Another crucial concern is the redistributive effects of the policy; the equity considerations among various groups of the society. These have largely been ignored within the context of regional policy. This is quite surprising given the fact that regional policy had its explicit redistributive objectives and instruments. Distributional effects, in themselves, raise several difficulties. First, there is the definition of what constitutes an 'equitable' distribution and what does not. Second, there is the classification of the society into groups in a meaningful way that relates to policy objectives. Third, it is difficult to measure policy effects from the viewpoint of different groups. Finally, it will remain to draw an overall picture of policy impact from an aggregate, society-wide perspective. Unfortunately, there appears to be no satisfactory resolution to any of these four problems.

When a policy has been in operation for a long period of time, difficulties also arise in regard to both its changing strength and the lag period. Nothing much has been done while assessing the effects of regional policy in regard to any of these two obstacles. Regular and systematic monitoring represents perhaps the best way out in this regard. An accurate, regularly up-dated,

record of the changes both in policy and on the ground (including as many extraneous variables as feasible) will generate the necessary data required for any further, more rigorous analysis. Regression analysis appears to be a useful analytical tool that can be applied in this respect to gauge the statistical relation between the policy and observed changes. The advances in computer technology make it a relatively easy task to run the specified models for as many times as necessary to establish a more accurate estimate of lag periods. (The application of the technique will remain confined to quantifiable variables unless a way was found to incorporate unquantifiable effects as well).

One critical issue in evaluation is the choice of indicators. It is perhaps the norm that policy objectives are stated in a manner that does not facilitate direct measurement towards their achievement. Performance indicators is the approach usually adopted in these cases. These indicators represent 'proxies' of policy objectives. It follows, then, that these indicators should relate as closely as possible to policy objectives. Within the field of regional policy, the closest indicator of policy objectives was unemployment rate. This, however, was considered not a suitable indicator for it was subject to the influence of forces other than, and in addition to, regional policy. This argument is hardly convincing. The main policy objective was to reduce regional unemployment disparities. Which, then, is a closer indicator of this goal, jobs created or reduction in unemployment rates? It is true that unemployment is affected by other forces (participation rate, migration, ...). It might have been the case as

well that data on these factors were not available. But, these influences have all been recognised for quite a long time (eg Moore and Rhodes, 1973a). Why, then, has not any research been carried out to disentangle the effects of regional policy on unemployment from those other factors? If it was the case of lack of data, why over such a long period of time has not any effort been made to secure these data? The results of some fifteen years of research hardly provide any indication of policy achievements towards its primary objective.

The implications for urban policy evaluation are quite visible. From the earliest stage possible in the policy-making process, indicators have to be defined that relate most closely to policy objectives. This may entail further research to identify more meaningful indicators than might be readily available. It may indeed entail a re-statement of policy objectives in a manner that clarifies the rationale behind them. In either case, data collection systems can, and should, be set up with a better understanding of what needs to be collected and why. The results are no doubt worth the effort. Monitoring will relate more closely to policy objectives. In a climate that is dominated by 'output measurement', the same effort will yield more information. It will also facilitate the use of statistical techniques to gauge the relationship between the policy at hand and observed changes on the ground.

It remains yet to fully answer the question why things are as they are on the ground. Part of the answer lies in disentangling the effects of the policy

from other forces. The other part, which relates to the policy itself, can be answered through 'process evaluation', or 'implementation analysis'; a closer examination of the organizational and administrative arrangements behind policy implementation (see chapter 6). The making and implementation of urban policy do not take place in a political vacuum. Both are susceptible to political and societal pressures. The multitude of stakeholders involved in both activities is another characteristic of urban policy. The interconnectedness, and in many cases conflict, between the interests and activities of these groups is very likely to influence actions taken on the ground.

These inter-related decisions and actions are an important clue as to why things happened the way they did. Therefore, process evaluation (or, implementation analysis) becomes an indispensable analytical tool of any comprehensive evaluation. Admittedly, the more questions asked the more problematic becomes the task. Yet, there are strong indications in programmes such as City Challenge and the SRB of the growing understanding of the importance of process evaluation.

The alternative to evaluation is what is being currently dominant in the field, ie output measurement. Despite any useful information it may yield, this alone is not enough. The persistence of the same problems over the past few decades in inner city areas is ample proof of the minor impact urban policy has had, ie its failure to achieve its goals. It may have been the case that, in the first place, problems were not correctly identified. It might have

been the case that objectives were unrealistic, or had not related to the problems. It might as well have been that the mechanism through which the policy was implemented was not effective. Or, it might have been the case that other factors had their influence. The resolution to all these uncertainties and questions lies within comprehensive evaluation. Otherwise, more failures, and consequently worse problems and more expenditure, are to be expected.

**CHAPTER 10:
TRUNK ROADS & MOTORWAYS
APPRAISAL AND EVALUATION**

10.1 INTRODUCTION

One of the main conclusions to emerge from chapter 8 was the lack of a systematic approach to urban policy ex-post evaluation. It also emerged that practice faces a host of conceptual and practical difficulties. These findings are the main driving force behind the review of experience in other, related fields. In chapter 9, it was found out that similar problems were encountered in assessing the impact of regional policy in Britain. It was also noted that little has been done to overcome these difficulties.

This chapter is a review of evaluation experience in another field, trunk roads and motorways investment. The choice of this field had two inter-related reasons behind it. First, the field has been, and still is, the subject of a well-established, computerised and consistent appraisal process. Second, the scale of the road programme and the vast amount of public expenditure it consumes. One might well expect a similar ex-post tradition to assess, at least, the value for money gained from such a long-standing, large public programme especially given the growing emphasis on the notion of value for money. However, and as will be explained later, there is no tradition of ex-post evaluation of trunk road and motorway schemes or policies on either side of the Atlantic. Ex-post evaluation is partial in its focus, concentrating in the main on economic impacts.

The following section examines the current practice of trunk road and motorway appraisal in the UK. The economic component of this practice is based on a computerised version of a typical cost-benefit analysis. Having detailed the underlying concepts and assumptions of this technique in chapter 3, only a brief review is given here of the elements of the practice. The environmental component of the practice has gone through two distinct stages of revision and updating. These, and the form of environmental appraisal, are reviewed in brief as well. The section also examines two proposals for more 'comprehensive' appraisal frameworks, though these proposals have adopted a narrow definition of 'comprehensiveness' focusing on economic impacts of transportation improvements.

Section 3 summarises the results of a review of a sample of ex-post evaluation studies prevailing in the field. This sample was first chosen at random; only one choice criterion was applied: the availability of information. The sample was the result of an extensive search in the literature on transportation. Two studies were then included: The Severn Bridge (Cleary and Thomas, 1973) and The M40 Case Study (Headicar and Bixby, 1992). These are the studies that key figures in the field, following personal contacts with them, have referred to as among the most important. This review proved the partiality of research on the impact of transportation investment. Economic impacts are the type most commonly assessed. A very limited number of indicators is usually used. Research has focused on either a particular scheme or a segment of the network. In other words, there is no

systematic, let alone comprehensive, tradition of ex-post evaluation in the field.

The Dutch experience is singled-out in detail for it provides a strikingly different, more comprehensive approach to monitoring and evaluation. At the outset, it shows a clear political commitment to evaluation. Systematic monitoring of the national transportation plan is being conducted on an annual basis. A study had been commissioned to design a methodology to evaluate the achievements of the plan. Preparations are underway to secure data and information requirements to apply the methodology in a later stage.

The information on this experience was gained during a field trip to the Netherlands (15 May - 5 June, 1995). Meetings were held with key senior officers at the Ministry of Transport, Public Works and Water Arrangement (in the Hague). In addition to the insights gained through discussions, valuable material was also obtained. In September 1995, a senior officer of the Dutch Ministry visited the Department of Civic Design. An extended meeting was held with Prof. P. Batey, Dr. P. Brown and the researcher. The discussion served, among other things, to confirm the conclusions drawn from the Dutch experience.

In concluding this chapter, the discussion focuses on three issues. First, the scope and coverage of the appraisal process currently in practice in the UK. In addition to the difficulties associated with CBA, the practice is partial and

fragmented. The range of benefits included is very limited and confined to those that accrue to road users. Environmental assessments seem to have a marginalised role and it is unclear how they are incorporated into the final decision. These limitations have been recognised for quite a long time. However, it appears that little, if any, has been done to overcome them. The practice, nevertheless, has several implications for ex-post evaluation. Perhaps most important of these is the need to adopt a broader framework of analysis in both appraisal and evaluation.

Second, the expectation of a systematic ex-post evaluation in the field has not been fulfilled. A number of key figures in the field were invited to comment on this conclusion. They were also asked, if correct, to provide their explanation why it is the case. Responses have all agreed that there is no such thing as systematic ex-post evaluation in the field of transportation investment in the UK. The reasons for such a situation are explained.

The third and final issue is the implications of the experience on urban policy ex-post evaluation. Despite the aforementioned conclusion, there are important issues to note. Practical and conceptual difficulties are the same. It is argued that an understanding of difficulties may ensure a rigorous methodology and encourage further research. Due to its time and resources requirements, political commitment seems to be a pre-condition for comprehensive evaluation. There is, however, an inherent conflict between evaluators and policy-makers caused by the different time scale of each. It is

argued that monitoring can be an immediate step towards the resolution of this conflict. It is also argued that a broadly focused monitoring exercise can be a valuable aid, and even a prerequisite, to comprehensive evaluation. Despite that conflict, and resources requirements, comprehensive evaluation should not be discarded if only for one reason: to avoid past mistakes and build on success.

10.2 APPRAISAL METHODOLOGY

The appraisal of trunk roads and motorways in the UK is the subject of a standard procedure consisting of the following stages:

1. Designing a traffic study,
2. Collecting data,
3. Selecting and building a traffic model,
4. The assessment of errors and the treatment of uncertainty,
5. Validation,
6. Forecasting,
7. Operational appraisal,
8. Economic and environmental appraisal, and
9. Presenting results. (DoT, 1991: para 2.2.2, p. 2-3)

The main concern here is the eighth step: economic and environmental appraisal. The purpose of economic appraisal "is to ensure that money spent on road proposals, in its entirety and in its details, provides value for money" (DoT, 1991: para 2.2.40, p. 2-11). The purpose of environmental appraisal, on the other hand, is "to ensure that the effects of a scheme that cannot be expressed in monetary terms are given due consideration in scheme assessment" (DoT, 1991: para 2.2.41, p. 2-11).

10.2.1 Economic Appraisal - COBA 9

Economic appraisal, in the form of a computerised, typical cost-benefit analysis, is incorporated into COBA 9, the essence of which is to estimate the Net Present Value (NPV) of a scheme (DoT, 1991: para 14.2.2, p. 14-3). COBA (COst-Benefit Analysis) was first introduced in 1972 (DoE, 1972). COBA 9 is the DoT's "standard method of economic appraisal for trunk road schemes" (DoT, 1991: para 14.2.1, p. 14-3). The use of CBA in trunk road appraisal was justified by the scarcity of public resources and the need both to ensure value for money from investment expenditure and to measure that objectively (DoT, 1981, np). The DoT, however, admitted that COBA is a "partial technique; economic appraisal of the sort embodied in COBA does not purport to measure value for money over the whole range of costs and benefits, including those broadly classified as environmental" (DoT, 1981, np).

COBA 9 operates on the assumption that road improvements may only lead to traffic re-assignment. The DoT holds the view that for most schemes "this is a realistic assumption and there is little evidence that the more complex effects [re-distribution, generation and modal split] are significant in most cases" (DoT, 1981: para 1.3.2).

For each scheme, the annual benefits and costs - over the entire road network affected by that scheme - are estimated for two major options: the 'Do Minimum' and the 'Do Something' options (see Fig. 10.1). The 'Do Minimum' option is the base road and traffic network against which other

alternatives are evaluated (DoT, 1981: para 1.2.2). In many cases, though not always, this is the existing network, ie the 'Do Nothing' option. However, even a literal 'Do Nothing' base case is not a 'No Change' one. Future traffic growth will increase the 'Do Minimum' user costs reflecting congestion (DoT, 1981: para 1.2.3).

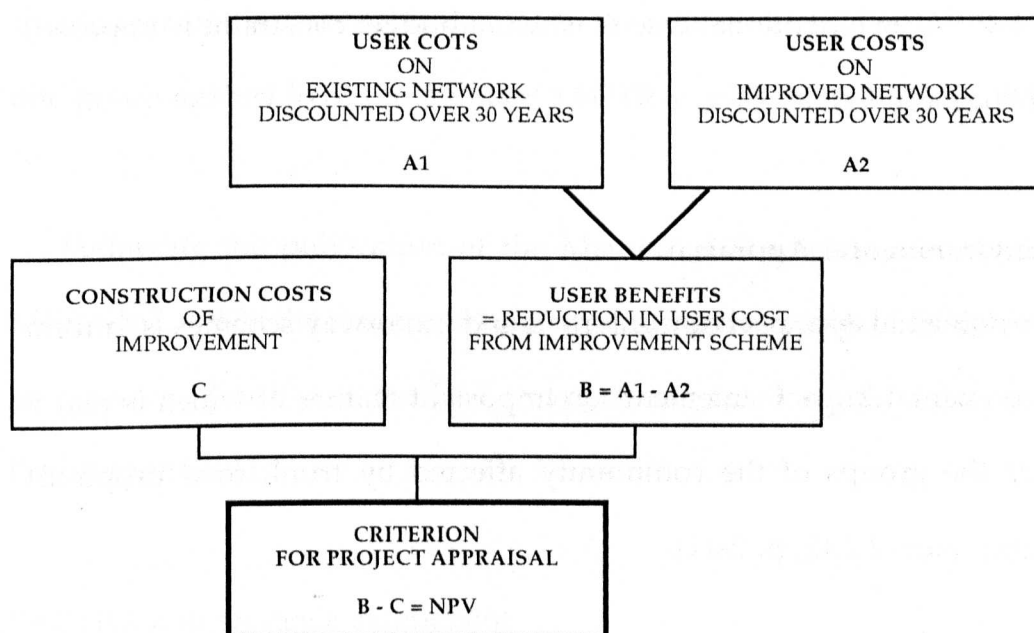


Fig 10.1: COBA Evaluation System

Source: DoT, 1981.

Measurement of benefits in monetary terms has been confined to those which accrue to road users:

1. Time savings (work and all other purposes),
2. Savings in vehicle operating costs, and
3. Savings in accident costs.

Costs, on the other hand, include:

1. Capital costs, including scheme preparation costs, and
2. Maintenance costs (track and user costs). (DoT, 1981: para 2.1.2 to 2.3.5)

The decision criterion is the Net Present Value (NPV), over a lifespan of 30 years. "Considered in purely economic terms, a scheme is justified if it has a Net Present Value (NPV) greater than zero" (DoT, 1981, para 3.1.1). Where several 'Do Something' schemes have positive NPVs that with the highest NPV is the preferred one. The ratio of NPV to NPC (Net Present Cost) may be used to indicate the scale of benefits to costs but it should not be used to rank options (The exceptional case is where a budget constraint is imposed) (DoT, 1981: para 3.1).

10.2.2 Environmental Appraisal

Environmental appraisal of trunk road and motorway schemes is "a form of environmental impact statement" an important feature of which is that it identifies the groups of the community affected by trunk road proposals (DoT, 1991: para 2.2.42, p. 2-11).

ACTRA and the Framework

In 1978, the Advisory Committee on Trunk Road Assessment - ACTRA - found no serious objections to the principle of economic appraisal on the basis of cost-benefit analysis. However, the appraisal approach as a whole was then found to be unbalanced emphasising economic effects over environmental ones (ACTRA, 1978: para 20.44, p. 95). The conclusion was:

Whilst current methods of scheme appraisal, based on COBA, are sound as far as they go, we believe the assessment to be unbalanced and we suggest a shift of emphasis in the whole approach (para 20.44, p. 95)

The recommended alternative took the form of what came to be known

as the 'Framework Approach' (ACTRA, 1978: para 20.36, p. 94). This was a tabular presentation of information on all likely impacts of different scheme alternatives broken-down by groups of the community. It was considered to be a comprehensive approach that "embraces all the factors and groups of people involved in scheme appraisal" (ACTRA, 1978: para 28.4, p. 133). The 'Framework' relied on value judgement. The Committee, however, believed that judgement is "inevitably central to the process of trading off benefits to one group against losses to another" (ACTRA, 1978: para 20.35, p. 94).

Following the publication of the Manual of Environmental Appraisal - MEA (DoT, 1983), the Framework, incorporating the results of COBA, became the standard practice of the DoT in appraising trunk roads and motorways (see Appendix H).

SACTRA and strategic assessment

In 1989, the Standing Advisory Committee on Trunk Road Assessment - SACTRA - was asked to

review the Department's methods for assessing environmental costs and benefits, in particular (to consider) whether a greater degree of valuation is desirable, the appropriate scope and application of valuation and suitable methods for deriving monetary values (SACTRA, 1992: para 1.08, p. 4)

In reviewing the then current practice, the Committee was quite critical of the practice of including the Framework in the environmental assessment. It was believed that such practice should be abandoned as it had both weakened and divided the treatment of environmental effects (SACTRA, 1992:

para 11.13, p. 58). It was also believed that the assessment should not be tied to the six Groups then defined in the MEA (SACTRA, 1992: para 11.15, p. 58).

A major theme that was strongly emphasised throughout the report was the "need for a strategic level of [environmental] assessment as reliance on scheme-appraisal alone may not give the full picture" (SACTRA, 1992: para 9.04, p. 48). In order to take into account the wide range of time-varying environmental effects of road-building and road-use, "the appropriate environmental assessment must underlie every stage in the hierarchy of decisions, from the making of national and regional policy downwards" (SACTRA, 1992: para 16.03, p. 91).

SACTRA also emphasised the need for environmental assessment to derive from explicitly stated policy objectives (para 9.06, pp. 48-49). Such an objective-led assessment should be "a continuing process throughout a scheme's development, and should commence at the earliest possible stage in the formulation of policy" (para 11.01, p. 55). The Committee recommended the preparation of a new manual which should concentrate exclusively on environmental assessment (para 12.03-12.04, p. 61). The Committee examined "the issues of theory and practice involved in the valuation of social and environmental impacts, their aggregation, and the implications of social cost-benefit analysis." Their conclusions were:

- (1) There is no legitimate objection of principle to the use of monetary values for evaluating as many of the environmental effects of road schemes as lend themselves to that technique, even if others cannot be so valued.

- (2) There are great advantages to be gained from extending monetary valuation as far as it is reasonable to do so.
- (3) However, any methods suggested for such valuation must satisfy the criteria laid down by in paragraph 9.06 of this Report. [see above]
- (4) There is a danger, which must be recognised, that as more environmental effects are introduced into cost-benefit analysis, a bias in favour of those effects may result. Those which are not valued must still receive their due importance in the appraisal.
(para 13.11, p. 73)

However, the Committee recognised that

there is a class of environmental impacts - potentially catastrophic changes, losses of unique or sacrosanct assets, and long-term impacts on future generations - for which monetary valuation techniques are unlikely to be helpful to decision-makers. But we must avoid the paradox whereby important impacts which cannot be valued in money might somehow be undervalued in the appraisal (para 14.05, p. 75)

Three valuation techniques were examined: actual costs, shadow costs and revealed values, contingent valuation and stated preferences (see chapter 3). In the Committee's view "the time has come to apply these techniques experimentally, to a sample of actual road schemes, corridors and strategic policy assessments." Immediate adoption was not, however, recommended "because the Department has not yet had the practical experience which would give some guide to whether, at any rate in the case of road schemes, the known criticisms could be overcome" (para 15.08, p. 86).

The Government response

The DoT welcomed SACTRA's suggestions and accepted the main principles and direction recommended in their report (DoT, 1992: 1). The DoT's response took the form of a new manual for the environmental assessment of trunk road and motorway schemes (DoT, 1993a), which replaced the previous manual of 1983. A total of 12 broad categories of

environmental effects were addressed in this manual:

1. Air Quality
2. Cultural Heritage
3. Disruption due to Construction
4. Ecology & Nature Conservation
5. Landscape Effects
6. Land Use
7. Traffic Noise and Vibration
8. Pedestrians, Cyclists, Equestrians and Community Effects
9. Vehicle Travellers
10. Water Quality and Drainage
11. Geology and Soils
12. Impact of Road Schemes on Policies and Plans.

The assessment of each effect was recommended to be carried out in three stages, in increasing detail, reflecting both the development of the scheme and the findings of the previous stage. Findings of each stage are to be presented in a separate report:

Stage 1 - Report 1:

An assessment, in broad terms, of the likely impacts of the broadly defined routes or corridors.

Stage 2 - Report 2:

An identification of the significance of effects of route options.

Stage 3 - Report 3:

A full assessment of the impacts of the preferred route (Environmental Statement).

The recommended form of presenting the results of assessment was the Environmental Impacts Table (EIT). This is a tabular presentation of data summarising the main likely direct and indirect impacts of a scheme, taking account of any mitigation measures (see Appendix I). It was stressed that the contents of an EIT are not rigidly circumscribed. However, it was advised that

the basic structure of an EIT should be as follows:

A. Appraisal Groups:

1. Local People and their communities
2. Travellers
3. The Cultural and Natural Environment
4. Policies and Plans

B. Land Use Table.

C. Mitigation Table.

10.2.3 Towards a Comprehensive Appraisal Framework

The narrowly-focused appraisal process of trunk road and motorway schemes has prompted calls for a broader framework. As early as 1979, Heggie (p. 65) called for "a less rigorous approach based, perhaps, on a planning balance sheet or on ... "multi-variate cost-benefit analysis"." To an extent, the Framework approach represented an attempt to broaden the process taking account of all likely impacts of a scheme on various groups of the society. Whether Heggie's argument was among those behind ACTRA's recommendations or not is unclear. However, both the Framework and the current practice of the DoT fail to account for any other economic benefits but for those accruing to road-users. For instance, and although the argument over the link between transportation improvement and economic development is inconclusive, the current practice fails to account for any possibility of wider economic impacts of a scheme.

Quite recently, attempts have been made to present broader, more

comprehensive *economic* appraisal frameworks. These attempts had, however, adopted a narrow definition of 'comprehensiveness'. It was confined, in the main, to the economic impacts of a scheme as distinguished from road-user benefits, the focus of appraisal processes. Two of these approaches are presented hereafter.

In the first framework (Perera, 1990; see Fig 10.2), transport user benefits were measured in terms of:

1. Time savings,
2. Savings from avoidance of delays at bridge or river crossings⁽²¹⁾,
3. Savings in vehicle operating costs, and
4. Savings in accidents costs.

Economic impacts were broken-down as affecting the following areas:

1. Business and Industry,
2. Residential,
3. Tax revenue,
4. Regional and community activity,
5. Resources, and
6. Appreciation of land values.

Table (10.1) summarises the impacts on each of these areas indicating whether they are direct, indirect or induced and whether they are permanent or temporary. The table clearly indicates the wide range of economic impacts a major transportation scheme is likely to have. In so doing, it also shows how current practice is narrowly focused.

In applying the framework, the initial stage is to select the objectives and criteria that will be applied in the choice between alternatives. The selection

⁽²¹⁾ Although this category is a subset of the first, it is unclear why it was singled-out as such.

and definition of criteria form an important part of the public participation process. Only those alternatives that meet the minimum standard of social and environmental criteria are to be selected for economic appraisal. Guided by these objectives and criteria, the second stage is to set and review the supply, demand and performance characteristics of the system within the study area. The outcome of this stage is the application of the economic analysis and evaluation of viable alternatives. The results of the above stage are then made public. Feedback on public perceptions is used to further improve the project selection. "Where possible and feasible, conflicting stakeholder interests are resolved" (p. 50). The fourth and final stage of the entire process is the reconciliation of any residual stakeholder interests and the selection of the project for implementation, which is a political decision.

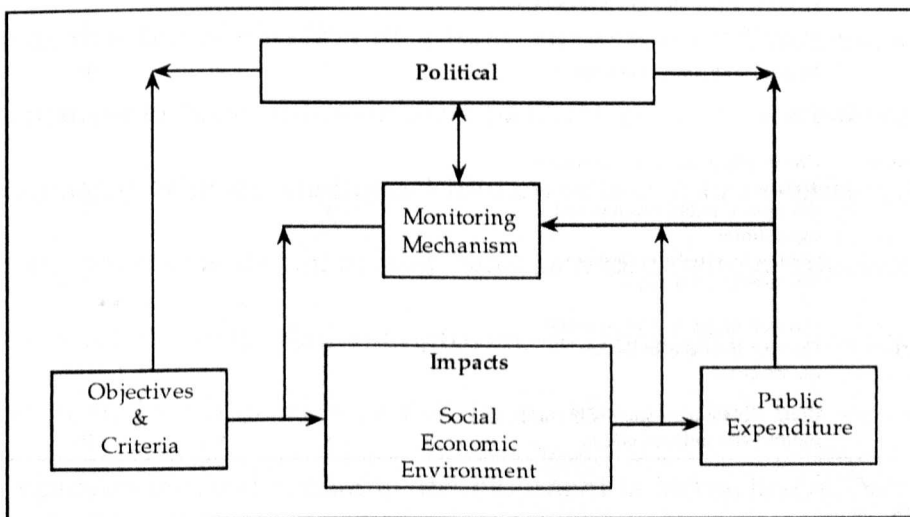


Fig 10.2: Evaluation Framework

Source: Perera, 1990: 49.

Table 10.1: Classification Summary of Economic Impacts

Class	Category	Effects	Direct	Indirect	Induced	Temporary/ Permanent
Business & Industry	Facility Construction	Expenditure on labour and materials for construction	x			T
		Secondary effects induced by direct expenditure		x	x	T
		Losses to firms in the vicinity		x		T/P
	R-O-W ^a Acquisition	Loss of jobs and services due to relocation	x			T
		Redistribution of jobs and services within the corridor		x		
		Loss of land	x			P P
	Business Growth	Expansion of existing businesses	x	x	x	P
		Attract new business or labour	x	x	x	P
		Deter businesses that depend on remoteness	x	x	x	P
	Tourism & Recreation	Expansion of existing businesses	x	x	x	P
		Deter businesses that depend on remoteness	x	x	x	P
		Divert potential business	x			P
	Agriculture	Increase or decrease in productivity and profit	x			T
		Encourage conversion of land to other use		x		P
		Improved accessibility to markets	x			P
Mining & Forestry						
Residential	Regional Economy	Placement & Relocation housing needs		x	x	T
		Attracts additional workers and families		x	x	P
Tax Revenue	Property Taxes	Loss of tax revenues due to acquisition	x			P
		Property value changes and associated tax revenue		x	x	P
	Public Service Needs	Require additional expenditure			x	P
Regional & Community	Community Region	Changes to pattern of community growth				?
		Changes to public revenue and expenditure		x		?
		Gain or loss in direct incomes	x			?
		Environmental changes				T
Resources	Land Material & Labour	Covered under R-O-W acquisition	-	-	-	-
		Covered in effects of facility construction	-	-	-	-
	Energy	Consumption associated with direct, indirect and induced effects	x	x	x	P

^a Right-Of-Way
Source: Perera, 1990: 48.

The second framework (Seskin, 1990) identifies three broad categories of economic benefits⁽²²⁾:

1. Business Expansion
2. Business attraction
3. Tourism.

The impact on business expansion is based on savings in travel time for commercial vehicles. A regional economic model is then used to determine the indirect and induced effects of business travel time savings and of changes in operating costs and of direct safety benefits to commercial vehicles (see Fig 10.3). The second category of benefits is the result of changes in the type of businesses and rates at which they are attracted to the region. The third and final category is the impact on tourism. This is measured in terms of the likely increase in tourism revenue following transportation improvements.

This framework was applied in several case studies. One of the findings was that the application of a more comprehensive framework generated a stream of benefits approximately 50 to 150 per cent of what would have been estimated with the traditional frameworks of user benefits (p. 32). A second conclusion was that both user benefits and economic effects are sensitive to the level of transportation improvement. These conclusions should, however, be treated with caution as the debate over the link between transportation improvement and economic development is inconclusive. Nevertheless, it is

⁽²²⁾ User benefits were identified in terms of the three common categories: travel time, operating costs and safety.

clear that the narrowly-focused appraisal process may lead to an under- or over-estimate of the economic benefits of transportation schemes.

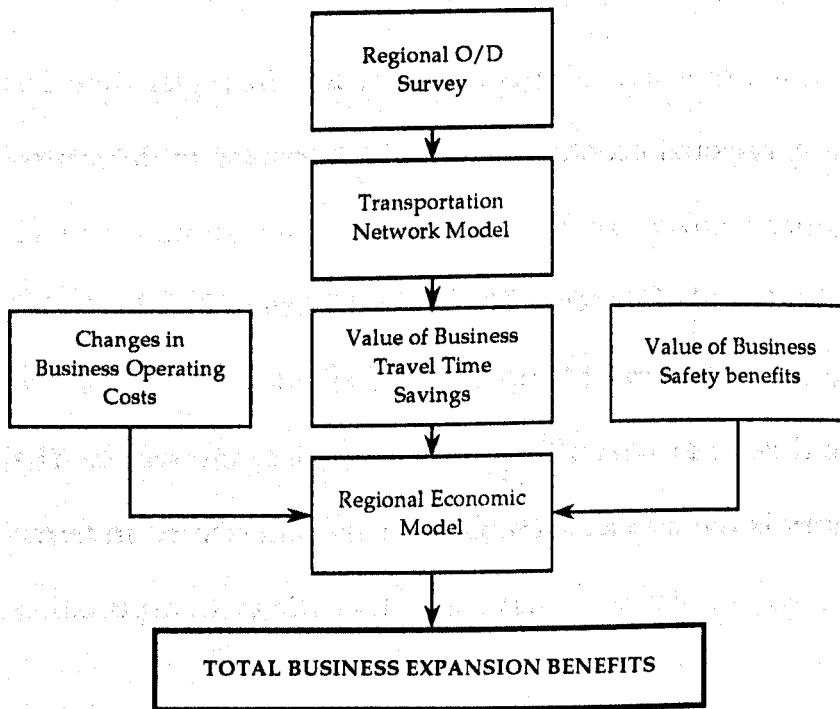


Fig 10.3: Business Expansion Analytic Framework

Source: Seskin, 1990: 29.

10.3 EX-POST EVALUATIONS

10.3.1 Experience in the UK and USA

Despite the well established practice of computer modelling and programming in trunk road and motorway design and appraisal, there is simply no such tradition of systematic, comprehensive ex-post evaluation of transportation investment, neither in the UK nor the USA. Perhaps we should first explain what is meant by 'comprehensive'.

Trunk roads, motorways and freeways have far reaching effects other than merely alleviating congestion. Even by only doing so, travelling costs are likely to decrease making movement of people and goods relatively cheaper. An improvement in the road network will also increase the accessibility of several locations. Accessibility is one of the determinant factors in location decisions for many types of firms. Although far from conclusive, the literature is quite rich in arguments over the role of transportation investment in economic development⁽²³⁾. Whether this economic development materialises or not, transportation improvement is very likely to have its impacts - positive or negative - on the socio-economic structure of those communities linked by, or adjacent to, the new or improved route. Thus, transportation investment may result in socio-economic and hence land-use impacts in addition to its traffic and environmental effects.

A 'comprehensive' evaluation, in our view, is an exercise that examines all of those "likely" effects, bearing in mind the inter-relations in between them and between them and other exogenous factors related to the wider environment (other policies, trends, ...). More importantly, and given the definition of evaluation adopted in this research, such an exercise should explicitly address the question whether transportation investments have achieved their goals, and why.

⁽²³⁾ For a review of the arguments on both sides of the fence see Grieco, M (1994).

Admittedly, it is a daunting task, but there is hardly any attempt towards such practice – again, neither in the UK nor the USA. To put it rather simply, comprehensive evaluation of transportation investments is severely lacking in both countries. This conclusion is based on both an extensive search in the literature and contacts with key figures in the fields of transportation and urban planning on both sides of the Atlantic. Studies on the impact of transportation investment in general, and motorways and trunk roads in particular, are partial in their focus; usually one type of impact is examined, most likely the economic. In both countries, there is no tradition of comprehensive evaluation of motorway schemes.

Despite ACTRA's recommendations almost 20 years ago (ACTRA, 1978), and recent supporting evidence on the land-use effects of trunk roads, transportation and land-use development are each prepared in almost complete isolation from the other. It is striking to find no research on the impacts of trunk road construction on the environment in its broad definition. Attention, primarily from within the medical profession, has focused on the health-related impacts of air pollution, mostly in urban areas.

It is even more striking to find an almost complete absence of follow-up studies of completed schemes. Given the explicit emphasis on 'value for money' in the DoT's practice of trunk road and motorway appraisal, one might expect a certain degree of interest in measuring the achievement of this objective. Nevertheless, re-appraisals of existing motorways, using actual

traffic counts rather than forecasts and on the same basis of their initial appraisal, are very rare indeed. Furthermore, there is no 'strategic' approach to the task (limited as it is). Apart from very few attempts, research has focused on either a particular scheme/route or a segment of the network. There is no such thing as an evaluation of national transportation policies.

On the bases of both the literature search and the contacts with key figures, a sample of research on the impact of transportation investment was chosen for review. This sample was first chosen at random; there was no criteria for choice except the availability of information. Two studies were then included: The Severn Bridge (Cleary and Thomas, 1973) and The M40 Case Study (Headicar and Bixby, 1992). The reason was the fact that several of the key figures contacted have referred to either, or both, as among the most important in the field. Table (10.2) summarises this sample, classified according to the type(s) of impact each study focused upon and the methodology used.

It is quite clear that the field is dominated by research on the economic impacts of transportation (10 out of 15 studies). The second category to attract attention was traffic patterns and flows. Although the number of cases is half that in economic-related research, this category includes one of the only two 'national' inquiries (SACTRA, 1994). The second national enquiry centred on environmental impacts, though from a narrow focus. The Royal Commission's report was mainly concerned with air pollution. Apart from this report, there

is no explicit consideration of the environmental impacts of trunk roads. Research comes from within the medical profession and centres on the health-related impacts of, mainly, air pollution in urban areas.

Few studies have addressed issues other than or in addition to economic effects. Only one (Briggs, 1981) considered 'social' impacts, in the form of net migration. Two studies (Cleary and Thomas, 1973 and, Mills, 1977) have attempted to re-appraise the project at hand on the basis of the principles used in the initial appraisal (ie ex-post CBA). Only two (Headicar and Bixby, 1992 and, Steptoe and Thornton, 1986) have also addressed the land use impacts of transportation investment. And only two studies (Cleary and Thomas, 1973 and, Steptoe and Thornton, 1986) have addressed more than two types of impact. The latter was, however, very limited in scope compared to the former.

Table (10.3) shows that the proxies used to measure economic impacts are very limited in range. Most of the studies have employed either the level of employment (number of jobs) or the level of income – sometimes both (eg Thompson *et al.*, 1993 and Lombard *et al.*, 1992). Finally, Table (10.4) classifies the sample according to the 'geographical scale' of each study. It is clear that two levels of coverage dominate the field: scheme-based research (5 cases) and state/province- or regional-based research (6 cases).

Table 10.2: Classification of the Sample by Type of Impact and Methodology

Methodology	Impact	Economic	Social	Traffic Pattern and Flows	Re-Appraisal (CBA)	Land-use	Environmental (air pollution)
Cross-Sectional Analysis		Thompson <i>et al.</i> , 1993					
Regression Analysis		- Zografos and Stephanedes, 1992 - Lombard <i>et al.</i> , 1992 - Wilson <i>et al.</i> , 1985 - Stephanedes and Eagle, 1986 - Briggs, 1981	Briggs, 1981				
Survey		Wilson <i>et al.</i> , 1982					
Input-Output Modelling		Allen <i>et al.</i> , 1988					
Before-After Analysis		- Cleary and Thomas, 1973		- Cleary and Thomas, 1973 - Headicar and Bixby, 1992	- Cleary and Thomas, 1973 - Mills, 1977	- Headicar and Bixby, 1992 - Steptoe and Thornton, 1986	Royal Commission, 1994
		- Steptoe and Thornton, 1986		- Steptoe and Thornton, 1986 - SACTRA, 1994 - DoT, 1993b			

Table 10.3: Classification of Economic-Related Research by Type of Proxy

Level of Employment and/or Income	Industrial Location (No. of Firms)	Unemployment	Retail Sales
Thompson <i>et al.</i> , 1993	Wilson <i>et al.</i> , 1982	Stephanedes and Eagle, 1986	Stephanedes and Eagle, 1986
Zografos and Stephanedes, 1992	Cleary and Thomas, 1973		
Lombard <i>et al.</i> , 1992	Stephoe and Thornton, 1986		
Wilson <i>et al.</i> , 1985			
Stephanedes and Eagle, 1986			
Briggs, 1981			

Table 10.4: Classification of the Sample by Geographical Coverage

Study/Ref.:	Coverage: National	State/Province/Regional (or less)	Project/Scheme
Cleary and Thomas, 1973			The Severn Bridge, England
Headicar and Bixby, 1992			The M40, England
Mills, 1977			Wellington by-pass, England
DoT, 1993b			Scheme-based, UK
Stephoe and Thornton, 1986			A local community, Louisiana, USA
Thompson <i>et al.</i> , 1993		Florida, USA	
Zografos and Stephanedes, 1992		Minnesota, USA (87 counties)	
Lombard <i>et al.</i> , 1992		Indiana, USA ((92 counties)	
Stephanedes and Eagle, 1986		Minnesota, USA (30 non-metropolitan counties)	
Wilson <i>et al.</i> , 1985		Brunswick, Canada	
Wilson <i>et al.</i> , 1982		Atlantic Region, Canada	
Briggs, 1981	Non-metropolitan counties, USA		
SACTRA, 1994	Generated traffic, UK		
Royal Commission, 1994	Environmental impacts, UK		

The three tables put together prove beyond doubt the partiality that is a characteristic feature of research on the impacts of transportation investment. More often than not, one type of impact is addressed, usually the economic. The range of proxies used to measure these economic impacts is limited, and often confined to one proxy. Finally, research has focused either on a particular scheme or a segment of the network. There is hardly any consideration of the impacts of transportation investments and policies at the national level.

The main concern here, however, is the methodology adopted in this research. Table (10.2) shows clearly that the field is dominated by two approaches: before-after analysis (7 cases) and regression analysis (5 cases)⁽²⁴⁾. Before-after analysis has a severe drawback. It is virtually impossible to disentangle the impacts of transportation investment from those of other exogenous factors (policies, trends, changes, ...) that may have taken place over the same period of time. Put another way, it is not possible to estimate what might have happened had the scheme(s) at hand not been implemented. This is what is usually referred to as the 'counter-factual' problem. The main advantage of this approach, on the other hand, is its simplicity. It does not require any advanced knowledge of statistics (and that is probably why it appeals to many, particularly decision-makers).

⁽²⁴⁾ It should be borne in mind, however, that these categories are defined in very broad terms. Some of the studies using regression analysis also used other methods as a first stage analysis before specifying and running the model(s) (eg Stephanedes and Eagle, 1986). SACTRA report, on the other hand, is not a 'strict' before-after analysis.

Regression analysis, on the other hand, provides a 'potentially' reliable technique to address this issue explicitly. This reliability, however, hinges on one particular condition. That is, the accurate and complete definition and measurement of *all* other factors that may have influenced the impacts observed on the ground. This, in itself, is the main limitation of regression analysis. The number of 'externalities' is virtually infinite. There is no way of ensuring that all externalities have been identified and accounted for. Furthermore, and subsequently, the technique requires a huge amount of data and information that may not always be available.

Regression analysis has another drawback. It does not gauge the direction of the relationship between the variables. The technique, put rather simply, calculates the statistical significance of the correlation between a number of variables. It does not, however, specify which of the variables really affects the other(s). For instance, highway improvements may stimulate economic growth. At the same time, economic development may create a demand for transport improvement. Causality tests are one way of gauging the direction of this relation. Only two studies among those reviewed had carried out such tests (Zografos and Stephanedes, 1992 and, Stephanedes and Eagle, 1986). Another shortcoming of regression analysis is that the equation that "best fits" — produces the highest correlation coefficient and meets all significance tests — may yet fail to meet the common-sense test in that the coefficient of one or more of the variables may have the wrong sign (Homburger, 1978: 191). This problem as well can be overcome by means of causality tests.

All the studies employing regression analysis have foundered on these difficulties particularly the condition to accurately and completely identify and measure externalities. The other techniques have their pitfalls, and that is probably why they are not as common as before-after or regression analyses. Cross-sectional analysis, as employed by Thompson *et al.* (1993), for instance, was not any different from the before-after approach. Surveys and questionnaires (eg Wilson *et al.*, 1982) raise a multitude of problems that cast considerable shadows on the results they may yield (eg sampling errors, generalisation, ...). Input-output models, by definition, are incomprehensive. These are economic tools that measure only economic impacts. Moreover, the models are neither very useful for specific sub-area nor are they sensitive to changes in travel time (Hirschman and Henderson, 1990: 38).

Finally, a crucial consideration that all studies have failed to consider is the distributional impacts of transportation investment on various groups of the society. This category of impacts raises four important questions: 1) What is an equitable distribution and what is not?, 2) On what bases should these groups be identified?, 3) How to measure the impact of the policy on each of these groups? and, 4) How to assess the overall impact of the policy on the society entire? Unfortunately, there appears to be no satisfactory answer to any of these questions. It is not surprising then that none of the case studies considered distributional impacts.

10.3.2 The Dutch Experience

In contrast to both the British and American, the Dutch experience provides a strikingly different, more comprehensive approach to monitoring and evaluation in the field of transportation. In 1990, the Dutch Government published a Government Decision known as the *Tweede Structuurschema Verkeer en Vervoer (SVV II)* – the Second Transport Structure Plan. This was a statement of the Government's long-term traffic and transport policy.

One of the actions that followed almost immediately after the publication of the SVV was the establishment of a project called '*Meten is Weten*' (To Measure is to Know – M=W), within the Ministry of Transport, Public Works and Water Management. This project aims at defining performance indicators, collecting information on these indicators and, ultimately, carrying out the monitoring of the SVV. *Monitoring of both the implementation of the Plan and the problems it aimed to tackle is actually being conducted on an annual basis.* Another action spawned by the SVV was the publication of the *SVV Aktieboek* (Action Book). Given the broad nature of the SVV, the *Aktieboek* identified all the actions seen as necessary to achieve the goals of the SVV, estimated the completion date of each action, its cost and the annual overall budget. The *Aktieboek* even identified the person(s) responsible for implementing each action. It is updated annually.

In addition to monitoring, the Dutch experience exhibits a considerable degree of political commitment to evaluation. In 1992, the Ministry of

Transport commissioned a study⁽²⁵⁾ that had the following three objectives:

1. To design a methodology that can be used to help explain which actions taken in response to the SVV caused what observed effects.
2. To demonstrate the methodology on a subset of SVV actions.
3. To document the methodology so that it can be applied more broadly at some future time. (Walker *et al.*, 1993: iii).

By addressing the causal relationship between actions and impacts, the main objective of this study, in other words, was to design a methodology that can answer the question whether the policy has achieved its goals or not. In itself, the study was seen as "a first step in carrying out a full-scale evaluation of the SVV" (Walker *et al.*, 1993: 1).

Although, as will be explained later, the Dutch study does not employ any particularly novel technique, it differs from other reported research in at least two important aspects. First, and at the outset, being commissioned by a central government department, the study indicates a certain degree of political commitment to evaluation. According to a senior officer at the Ministry of Transport, the question whether the Plan has achieved its goals or not has always been, and still is, at the top of the Ministry's agenda (May, 1995). Securing this political commitment, according to the senior officer, was not much of a problem. The SVV II, as approved by the Dutch parliament, incorporates explicitly a commitment to both monitoring and evaluation. In more general terms, there are strong indications within the Dutch context of a growing understanding of the importance of policy evaluation, ex-ante and

⁽²⁵⁾ Walker *et al.*, 1993.

ex-post, and hence an increasing commitment within the government circles for such research (Field Trip, 1995).

The second difference is the fact that the study does not aim at a single route or section of the network. Neither does it confine itself to a particular, perceived impact of the Plan. Rather, it aims to examine the causal relationship between the SVV-related activities and observed changes on the ground. In doing so, the counter-factual problem is explicitly being addressed. The results will thus provide an answer to the main question behind the study: Has the SVV II achieved its goals? The existence of a strong relation between activities and impacts is no doubt a positive answer to this question; the reverse is equally true.

Conceptually, the methodology of the study is quite straightforward. Given the main objective of the study, and the broadness of the Plan, the first task was to re-structure the information of the SVV in a way that would make it easier to link the SVV actions (tactics⁽²⁶⁾) to their impacts (proxies⁽²⁷⁾). This restructuring took the form of the "Expected Interaction Matrix - EIM" (see Fig 10.4). In this matrix, the rows identified SVV tactics while the columns identified impact measures (indicators) which were derived from the

⁽²⁶⁾ A tactic is "a single specific action taken to help solve one or more of the policy problems" (Walker *et al.*, 1993: 8).

⁽²⁷⁾ A proxy is "a measurable consequence of policy that is related to a goal." Concentration of NO_x is a proxy of air pollution; however, it is not the same as air pollution (Walker *et al.*, 1993: 6). "Proxies are a subset of impacts." (p. 9).

'M=W' project. Entries to the matrix indicated the way a specific tactic was expected to affect the corresponding impact measure. These entries could be either quantitative or qualitative. The matrix also contained two additional columns that identified the chain linking tactics to their expected impacts.

The proposed methodology then proceeded in three steps:

1. Create an *analysis file* that contains the necessary data, in a form that can be used for the analysis.
2. Use the analysis file for a *descriptive analysis* that shows, in a simple way, the relationships between tactics and outcomes (...), and which shows other factors that might stand in the way of, or support, conclusions about the effectiveness of the various tactics.
3. *Test* the relationships found in the descriptive analysis using formal statistical techniques, some of them simple, some of them more complicated, to verify that relationships suggested by the descriptive analysis are not just the result of chance. (Walker *et al.*, 1993: 31-32, stress in original)

Fig 10.4: A Sample of the EIM

Tactic	Proxy		
	Car km.	Noise (dB(A))	Deaths
Restrict employee parking	++	+	+
Build new roads	-	-	?
Reduce speed limits	0	+	++

A tactic is expected to:

strongly help (++) hinder (-) or, not affect (0)
 help (+) strongly hinder (--)

the achievement of a particular target for each proxy. '?' denotes the possibility of some relationship, but the direction is unknown. A 'blank' indicates that a relationship is unlikely to exist.

Source: modified from Walker *et al.*, 1993: 17.

It was recognised that the analysis may never proceed in such a strictly linear process and that it was very likely to involve several feedback loops (see Fig. 10.5). Descriptive analysis makes use of simple statistical techniques, such as before-after, time-series and cross-sectional analysis. Multivariate regression analysis was used in the demonstration to test the findings of the descriptive analysis.

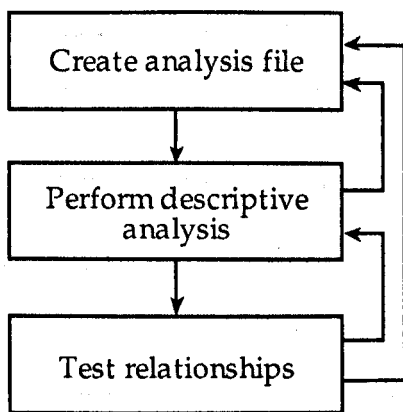


Fig 10.5: Analysis Steps and Their Interrelationships

Source: Walker *et al.*, 1993: 32.

The study faced several practical problems. Perhaps not surprisingly, the major difficulty was the lack of data; "much essential data for policy evaluation are not available or are difficult to obtain" (Walker *et al.*, 1993: xiv). Another barrier was the fact that few SVV tactics had actually been implemented by the time the Study commenced. Most of these tactics had been implemented too recently to have produced any real impact (p. xii). Consequently, many of the study's suggestions were concerned with the creation of data files (on both impacts and externalities) and data collection systems (Walker *et al.*, 1993: xiv-xv). On the other hand, it was evident that

over time tactics implemented at the early stages of the Plan will have produced more observable impacts.

As discussed earlier, the objective of this study was to design a methodology that would be used in a later time to conduct a full-scale evaluation of the SVV. Unfortunately, the idea had to be temporarily "put into the fridge", according to a senior officer within the Ministry of Transport (May, 1995). According to the senior officer, the reason was one particular drawback of regression analysis [which, in our view, has been exacerbated by the sheer broadness of the idea]. At the level of such an evaluation of a national policy, regression analysis requires a huge amount of information, much of which is not yet available (May, 1995).

However, the idea of evaluating the Plan is *not* discarded. Policy-makers within the Ministry [and outside it] are increasingly asking whether the SVV has achieved its goals or not. This question remains at the top of the Ministry's agenda and, in line with the study's recommendations, arrangements are being made to secure the data needed for this exercise (May, 1995). One such arrangement is the continual monitoring of the Plan on an annual basis. Another important arrangement concerns the co-ordination and collaboration between the different Ministries to secure much of the data needed (especially on 'externalities'). A workshop is scheduled next autumn (1995) between the Ministries of Transport, Environment and Public Housing and, Finance. The co-ordination with the Ministry of Environment and Public

Housing is seen to be of great importance in this respect given the close interaction between its policy domain and that of Transport. The Ministry of Finance is involved because of its role in co-ordinating and advising on evaluation research within the central government.

Questions remain in regard to two crucial aspects: qualitative data and the distributional impacts of the policy. Entries to the EIM could be either quantitative or qualitative. It is unclear how the latter can enter into the analysis. One particularly important category in this respect is environmental impacts. Although quantitative proxies can be developed for several of these impacts (eg increase or decrease in noise levels), many others are measurable only on a qualitative scale (eg visual intrusion). It is true that ex-post Environmental Impact Assessment is compulsory in the Netherlands and that it is the responsibility of public authorities. The study may have thus relied on this fact to accommodate for environmental impacts. However, of the ex-post EIAs that should have been carried out, only about 1% have actually been completed (Field Trip, 1995). More important, if the objective of the study is to establish what SVV actions have caused what effects, environmental effects should be no exception. The SVV will definitely have its environmental effects. Furthermore, the question remains in respect to other 'qualitative' impacts.

The second crucial issue relates to the distributional impacts of the SVV. There is no doubt that the Plan will affect different groups of the society to

varying degrees. The study used aggregate data, on the national and provincial levels. There is no mention whatsoever of distributional impacts. The lack of data may have been the reason. Nevertheless, and despite its explicit methodological objective, the study advanced no suggestions as how these impacts can be taken account of in the analysis.

Finally, the study may be criticised for the lack of any explicit attention to the "economic" aspects of the Plan and its implementation. However, it has to be borne in mind that within the Dutch context, economic efficiency and effectiveness is the domain of the 'Court of Audit' which is an entirely independent body, with considerable powers, responsible for conducting such analysis. The work of the 'Court of Audit' is actually being seen as supplementary to the monitoring and evaluation of the Plan. In addition, there is no reason why proxies of 'economic impacts' cannot be developed and then incorporated into the model. That has been the case with several of the reported case studies (see above).

10.4 DISCUSSION

10.4.1 Appraisal Methodology

The current practice of trunk roads and motorways appraisal in the UK is partial and fragmented. The DoT admitted that COBA is a "partial technique; economic appraisal of the sort embodied in COBA does not purport to measure value for money over the whole range of costs and benefits, including those broadly defined as environmental" (DoT, 1981, np). Environmental assessment, in the form of an EIS, seems to have a marginalised role in the whole process⁽²⁸⁾.

The practice of economic appraisal, being based on CBA, lends itself to all the conceptual and practical difficulties the technique faces in public policy-making (see chapter 3). In addition, the practice raises several other crucial concerns. Many of these concerns have been expressed for a long time, though nothing has been done so far to overcome them. First, the practice is narrowly focused; the range of benefits recognised is very limited and confined to transport-user benefits (eg ACTRA, 1978, para 12.5 & 12.16; see also Gwilliam, 1972: 124 and Mills, 1977:21). The application of a broader framework (Seskin, 1990) clearly indicates the inherent danger in the current, narrowly-focused practice. Confining the analysis to road user benefits can lead to either an under- or over-estimation of the NPV of a scheme.

⁽²⁸⁾ Transportation schemes in the Netherlands, on the other hand, are the subject of systematic and consistent multi-criteria appraisal. The indications are that environmental issues are given considerable weight in this appraisal.

Second, and related, COBA does not allow, adequately or otherwise, for generated development and traffic. As early as 1978, ACTRA (para 12.18) received evidence on COBA's inadequate allowance for regional benefits. Headicar and Bixby (1992: 6) asserted that "the issue of generated development and traffic is not included in the assessment of motorways and trunk roads." SACTRA (1994: para 11.16, p. 168) concluded that "induced traffic can and does occur, probably quite extensively". However, COBA is based on the 'fixed-trip matrix' which assumes that only traffic re-assignment may take place as a consequence of a new or improved road. The more complex effects (re-distribution, generation and modal split) are assumed to be insignificant in most cases (DoT, 1981, para 1.3.2). As for generated traffic, specifically, the DoT holds the view that: "There is no evidence of such an effect" (DoT, 1993b, para 0.2).

Third, the principles underlying the valuation of time savings, in themselves, raised several issues. At the outset, ACTRA (1978, para 21.9) recommended a three-level definition of types of trips: in work, to and from work and pure leisure. Nevertheless, both the DoT's standard practice and the central government guidelines (eg HM Treasury, 1991) still adhere to two types: working and non-working. Furthermore, all time savings are valued at the same rate regardless of their actual length. This assumes that even a small time saving can instantly be translated into productive output, which is not true (eg Heggie, 1979: 59). Non-working time is valued on the basis of inferences from people's preference expressed in their day-to-day choices

between different modes of travel. Whether people actually know the attributes of competing modes is open to dispute. ACTRA has also referred to concerns about valuing leisure time at an equity rate and evidence suggesting that higher income travellers would pay more for the same unit of time saving than would lower income travellers. However, the Committee had the view that concerns about such bias in favour of those with higher income "is a matter for consideration outside the cost benefit analysis" (ACTRA, 1978: para 21.8, p. 102).

Finally, perhaps the most problematic issue in trunk road appraisal, as in other fields, is those costs and benefits broadly defined as environmental. The manual of 'Environmental Assessment' (DoT, 1993a) probably goes as far as current knowledge permits in identifying and measuring environmental impacts. However, there remain several issues of concern. First, the manual falls short of meeting some of the criteria SACTRA (1992) considered essential for any environmental assessment system. It does not relate, either explicitly or otherwise, to stated policy objectives. To an extent, the manual takes a passive attitude towards the environment concentrating in the main on measurement and presentation without any discussion of limiting, let alone preventing, negative impacts. Also related, the manual omits several crucial effects, eg blight and use of non-renewable resources⁽²⁹⁾.

⁽²⁹⁾ The DoT considered the knowledge of the causes of blight, and methods of forecasting its extent, insufficiently developed to enable advice to be given on its assessment (DoT, 1993a, Section 3, Part 6: Annex AII).

Second, many of the environmental effects are measured on a qualitative scale based in the main on expert's judgment. There is not much objection to either of the two. The problem lies in the absence of any systematic measurement technique and the apparent complete reliance on value judgement. It is very likely that different individuals will come up with different assessments of the same effect. Third, the measurement of some of the effects does not reflect their true social cost. The agricultural land taken is lost to the nation for ever; households affected by compulsory purchase of their properties are usually unwilling sellers (Gwilliam and Wilson, 1980: 90).

SACTRA's recommendation to start experimenting with techniques such as revealed valuation (eg property values) does not represent much of a solution. Property values reflect a multitude of factors other than and in addition to environmental concerns. The use of the cost of mitigation measures as an indication of environmental costs may not always be practical. In cases, mitigation costs may not reflect the true 'social' cost of the impact. In others, there may simply be no mitigation measure applicable. Finally, one may question the definition and number of groups under which the assessment is presented and whether this is the best classification applicable or not.

Despite these limitations, or rather because of them, the practice has a considerably significant implication for, and indeed highlights a major shortcoming of, ex-post evaluation research. The drive behind both the road

programme and its appraisal is chiefly an economic one (DoT, 1989, para 26 and DoT, 1991, para 2.2.40). Assessing the achievement of this objective, it follows, should be a main theme in both the appraisal and evaluation of schemes. Yet, the range of benefits and costs taken into account in the appraisal of trunk roads is limited and confined to those of road-users. There is no consideration of any wider impacts of transportation investments, although the whole programme is seen as a "vital further boost for British industry" (DoT, 1989, para 3). In sharp contrast, ex-post evaluation research is primarily concerned with those wider impacts to the extent of almost ignoring benefits to road-users.

Meanwhile, environmental assessments seem to have a marginalised role in the appraisal process. It is unclear how they are actually being incorporated into the final decision. On the other hand, this kind of assessment is almost completely absent in ex-post evaluation. Attention is primarily drawn to pollution, particularly air pollution. And, neither of the two approaches considers the distributional impacts of road improvements on different groups of the society.

In sum, a broader framework that incorporates explicitly all likely impacts (particularly environmental) and distributional effects has to be adopted in both ex-ante and ex-post evaluation. The concern with value for money should be at the heart of appraisal and evaluation since it is the main objective behind the road programme. However, that should not, by any

means, be at the expense of other, equally important, impacts these schemes are likely to have. In turn, the focus of both appraisal and evaluation practice have to be broadened enough to allow for a more comprehensive assessment of the impacts of road programmes.

10.4.2 Ex-post Evaluation

The choice of trunk roads and motorways, rather than intra-urban schemes, as the focus for review had two related reasons behind it. First, there was the long tradition of systematic and consistent appraisal this field enjoys. Second, the scale of road programmes and the amount of resources expended on road networks, on both sides of the Atlantic. One might well expect a similar tradition of ex-post evaluation to assess the 'value for money' of such massive public programmes, especially given the growing emphasis on this notion of 'value for money'.

However, that was not the case, at least neither in the UK nor the USA. The case studies reviewed and the search in the available literature have revealed that there is no such thing as systematic, comprehensive ex-post evaluation of even a single motorway or trunk road scheme, not to mention the Road Programme as a whole.

In support of the research, a number of key figures in the field of transportation were invited to comment on this conclusion. They were also asked, if correct, to provide their explanation why it is the case. Those 'invited

commentators' were: Prof. Phil Goodwin, TSU, University of Oxford; Peter Mackie, ITS, University of Leeds; Prof. Peter Hills, TORG, University of Newcastle and, Peter Headicar, Oxford Brookes University (see Appendix J for a copy of the 'standard' letter sent to each of them). The choice of these commentators was based primarily on their involvement in, and contribution to, the field of transportation planning and research. Response was received from the first two; the others have apparently declined to reply. In addition, a query was placed at the Regional Science Association List-Server on the InterNet. Few responses were received of which that of Prof. Martin Wachs (UCLA) was the most important. Contact was also sought with Prof. Peter Hall (UCL) to provide a wider, slightly different, view of the issue, ie from the viewpoint of town and regional planning in general.

All received responses have agreed on the validity of the above conclusion: there is no tradition of systematic, let alone comprehensive, ex-post evaluation of transportation investment:

... there is no tradition at all of ex post appraisal of roads in the UK.

... my belief is that in the UK very few serious ex post appraisals of road investment have ever been made
(Peter Mackie, personal contact, 1994)

I can confirm that (...) ex post evaluation of road schemes is typically partial
(Prof. Phil Goodwin, personal contact, 1994)

The question that strongly poses itself is, given the scale of the Road Programme, why is comprehensive ex-post evaluation something of a myth than reality? Peter Mackie provided the most detailed answer to this question.

The difficulties facing ex-post evaluation could be classified into two main categories:

1. Technical difficulties

- (a) ambient variability. Traffic varies day to day, with fluctuations around the mean level, which itself varies systematically by time of day and day of week. There is therefore a problem of measuring, with statistical confidence, what is happening on the network, and of measuring statistically significant differences in flows.
- (b) behavioural response. From a scheme appraisal point of view, it is not sufficient to know what is happening on the scheme, it is also necessary to know to what extent flow changes are caused by rerouting, changes in origin, destination, mode, and completely new (generated) traffic. The experimental design to capture these effects separately is difficult and costly.
- (c) dynamic aspects. Road improvements cause changes in travel conditions which cause a mixture of short-term and longer term adjustments. Rerouting may occur almost immediately, but changes in land-use resulting from better travel conditions may take years to manifest themselves. Proper studies would need to be conducted over a lengthy period.
- (d) the counterfactual problem. It is inadequate to compare what happened with what was expected to happen. It is necessary also to consider the extent to which the difference is explained by exogenous factors (GDP growth etc.) and what would have happened in the absence of the scheme. Thus, you need a model; this is expensive and time consuming.

2. Political constraints

there has, ..., been a lack of political will to understand and monitor what happens when roads are improved. If a commercial investment fails, people have to learn lessons, because the profitability of the company is reduced. There is no comparable discipline in the road sector. ..., the appraisal culture is always looking to the next road, never back to the last one (unless there are engineering failures, which are taken seriously). (Peter Mackie, personal contact, 1994)

One can only agree with this view. Support also comes from other sources, even if not as exclusive as the above quotation. Holland and Sherman (1980: 70-1) shared the view that a major difficulty in assessing the impact of a strategic highway is to identify those effects which stemmed directly from the scheme — the counter-factual problem. SACTRA (1994: ii), despite having concluded that induced traffic does occur, found it "remarkably difficult to establish unequivocal quantitative evidence" that proves either way whether improved road networks induce new traffic. Several reasons were given:

1. "New roads have ripple effects on traffic over a wide area",
2. "Induced traffic may build up over time rather than appear at once", especially if it was associated with road-related land use development, and
3. "Evidence from traffic counts and before and after surveys is inherently subject to various sources of error, both in measuring what has happened and in assessing what would otherwise have happened in the absence of the scheme." (SACTRA, 1994: ii)

This finding and the reasons behind it are certainly expandable to other road-related effects. At the outset, it is always difficult to define the area over which the road may have had its impacts. One is always in danger of overlooking an area that either has witnessed a road-related impact or, indeed, influenced the impact of the road elsewhere (eg large conurbations or metropolitan areas). As mentioned above, it is difficult to establish the exact lag period over which impacts may have occurred.

Furthermore, there are those 'problematic' effects known broadly as 'environmental'. To put it rather simply, there seems to be no way of assessing the overall impact of a scheme on the environment except by value judgement. The widely-accepted format of Environmental Impact Statements/Assessments does not provide such a conclusion. Probably no one can claim to do so. Environmental impacts, by their very nature, are measured on a multitude of different scales, some are even non-measurable so far. An Environmental Impact Statement, one way or another, provides an estimate of the different impacts of a scheme. It remains the task of the reader

to judge the overall direction of the document at hand. Is the overall environmental impact of the scheme positive or negative?

Assuming that such a question was answered, one way or another, it remains still to assess the overall impact of the scheme. This means that monetary-valued effects will have to be weighed against environmental effects. This amounts to placing a monetary value on the environment. Apart from any ethical arguments against such attempts, there is simply no knowledge available of how to do so. "We simply do not know how to evaluate environmental effects in monetary terms, or how to translate them into an equivalent NPV. We likewise lack an institutional framework that would enable environmental costs to be recovered through an efficient pricing mechanism" (Heggie, 1979: 62-3, see also Sharp, 1979: 90).

Finally, there is the distributional impacts of transportation investments. As mentioned above, this category of impacts raises four crucial questions: 1) What constitutes an equitable distribution and what does not?, 2) On what bases should the different groups be identified?, 3) How to measure the impact of the policy on each of these groups? and, 4) How to assess the overall impact of the policy on the society as a whole? There appears to be no satisfactory answer to any of these questions.

The Dutch example faced some of these problems, and avoided others. The major two barriers the study faced were the severe lack of data and the

lag period. Being carried out at the national level, the study avoided the problems of defining a study area and, more importantly, the distributional impacts. It is nevertheless unclear how 'qualitative' data, including environmental impacts, can enter the analysis.

On the other hand, the example provides somewhat more positive conclusions. First, and at the outset, it exhibits a certain degree of political commitment and support to evaluation. This commitment has to be considered in its wider context. There is a growing understanding of the role and importance of evaluation, and hence commitment, within government circles in the Netherlands. Second, it represents the only attempt to comprehensively evaluate a policy, in terms of strategic and geographical coverage and types of impacts considered. It aimed to develop a methodology to evaluate a national policy at the national level, without being confined to a particular type of impact (It remains to be seen, however, whether it will be successful or not). Third, and related, the study explicitly addressed the counter-factual problem. The methodology proposed was aimed at explaining which SVV actions have achieved what effects.

10.4.3 Research Implications

We now turn to the question that is at the heart of this chapter: How does the experience of trunk roads and motorways appraisal and evaluation reflect on urban policy evaluation? Does the field, somehow, offer any 'useful' recommendations for urban policy ex-post evaluation? Are there any

transferable 'lessons'? The answer is, perhaps surprisingly, 'Yes'. In the first place, the practical and conceptual difficulties of ex-post evaluation are almost identical in both fields. It is important to have a clear vision of such difficulties, beforehand, prior to commencing on any evaluation exercise. This in itself may ensure a rigorous process that addresses these issues in a systematic way, or at least attempts to. It may also be of use in directing future research on the topic (though apparently that has not been the case in the past). Perhaps the most crucial concern is the fact that any such comprehensive evaluation will be a costly and time consuming task. This raises two problems. First, the resources and time required may simply not be available. Second, and more important, that *may* not be what decision-makers (the end-users of evaluation findings) are looking for. This conflict of interest actually makes matters worse. Decision-makers are asking for a thorough evaluation of complex issues which does not take long, nor costs much, to undertake.

In this respect, the Dutch study appears to be the exception. It was commissioned by a central government department with an explicit methodological, and comprehensive, objective. Arrangements are underway to secure the data needed for a full-scale evaluation of the Plan. Decision-makers (inside and outside the Ministry) are increasingly asking whether the SVV has achieved its goals. As mentioned before, there are strong indications of a growing understanding and commitment to evaluation within government circles in the Netherlands. Why is that so is open to speculation.

The resolution of the apparent conflict, in other contexts, between decision/policy-makers and researchers is considered to be a long-term process (Prof. P. Rietveld, personal contact, 1995). Evaluation research will have to be an integral component of education in all sorts of disciplines (economic, social, ...). Over time, it is hoped that the 'public service' will have a growing number of officials who 'understand' the importance of evaluation and, hence, become more committed to it.

Meanwhile, and as the Dutch experience shows, regular, systematic monitoring represents the first immediate step towards the resolution of this conflict. Monitoring provides a continuous feedback on the performance and outputs of the policy at hand. This information is, no doubt, of great value to all stakeholders involved in the planning and implementation process. If indicators were to be close proxies of policy objectives, monitoring can even provide a 'first indication' on policy outcomes and impacts. A broader view of monitoring will serve to indicate the changes taking place within the wider environment. Overall, the information gained from regular, systematic and broadly-focused monitoring is a prerequisite of any evaluation and implementation analysis.

However, it is not only a matter of demonstrating clearly what the difficulties are. The different broad categories of the case studies – especially the Dutch example – put together, seem to complement each other and point out potential directions for integration. It is clear from 'before and after'

studies that as much detailed a documentation as possible of the situation prior to the introduction of the policy is required. This documentation will have to cover a reasonable period into the 'past'. Past trends have to be fully taken account of in order to avoid spurious conclusions. As argued above, evaluation will obviously have to be preceded by a wide-scope, systematic and regular monitoring which charts (possibly) *all* changes on the ground. It is also clear that a much broader focus of the analysis is needed to take into account all likely factors, not only the policy at hand. Advances in computer technology may then facilitate, to a large extent, the use of regression analysis and causality tests; a means of estimating the statistical significance of the relationship between the policy at hand and the observed effects.

Nevertheless, difficulties remain. Perhaps the major problem is the time and costs such a comprehensive evaluation will require. The question will always remain: How to undertake a thorough evaluation of a policy, programme or project without imposing requirements far beyond the limits of what any political context can, or may be willing to, offer and at the same time, ensure that all questions raised are answered satisfactorily? The wider the policy at hand, the more issues at stake, the more challenging becomes the question and the less likely an answer may be found. This certainly, and under any circumstances, does not mean to abandon evaluation research altogether. Definitely not, if only for one reason: to avoid repeating past failures which cost nations scarce resources.

**CHAPTER 11:
TOWARDS A SYSTEMATIC
APPROACH**

11.1 THE MAIN OBJECTIVE

The main objective of this research was to develop a systematic approach for comprehensive ex-post evaluation of urban policy. 'Comprehensive' is defined as the attempt to answer *all* the questions policy-makers and stakeholders are likely to raise at the various stages of the planning and implementation process. These questions will relate to the output, impact, economy, efficiency, effectiveness, process and equity of the policy at hand. If evaluation is to inform policy-makers and stakeholders, *timely* answers will have to be provided to *each* and *all* of these questions. In order to do so, evaluation will have to draw on several traditions, and employ a range of analytical tools, at once. At the same time, it has to be borne in mind that circumstances differ from one case to the other. This systematic approach will, therefore, have to be flexible and capable of modification.

In moving towards the achievement of this objective a number of different, yet supplementary, sources of knowledge and experience have been examined. We have reviewed the underlying assumptions and strengths and weaknesses of existing appraisal methods and examined their applicability in ex-post evaluation and the choice among them (chapters 3 and 4). We have examined the different views and models of both monitoring and implementation analysis and the role each can play within a comprehensive evaluation approach (chapters 6 and 7).

The research has also benefited from first hand experience in urban policy evaluation within a governmental context (chapter 7). We have also critically reviewed the methodology adopted for evaluation in a number of case studies in the field of urban policy (chapter 8). Evaluation traditions in the fields of regional (economic) policy in Britain and, trunk road and motorway schemes in the UK, USA and the Netherlands have also been reviewed (chapters 9 and 10). The research has also profited considerably from discussions and contacts with a number of researchers and practitioners in different fields in these three countries.

11.2 THE PROPOSED APPROACH

On the basis of these various strands, I believe that the components of the proposed approach should be:

1. Multi-group monitoring;
2. Cost-effectiveness analysis (CEA);
3. Multi-criteria evaluation (MCE);
4. Implementation analysis; and
5. Statistical analysis.

The structure of the proposed approach is shown in Fig 11.1. While we agree that "the decision-making process is seldom sequential and thus a strict order in the evaluation process should be avoided" (Alterman *et al.*, 1984: 382), monitoring appears to represent the most appropriate starting point for the evaluation *process*. We will return to this point later in detail.

Initial Stages

The first task in any evaluation exercise is to establish an understanding

of the policy at hand (*Initial Stage 1*). Evaluators need to determine what is the policy about?, what are its objectives?, what does it aim at? what are the problems it addresses?, how is it envisaged these problems should be tackled and its objectives be achieved?, who is intended to benefit from this policy? and, where do the resources required come from?

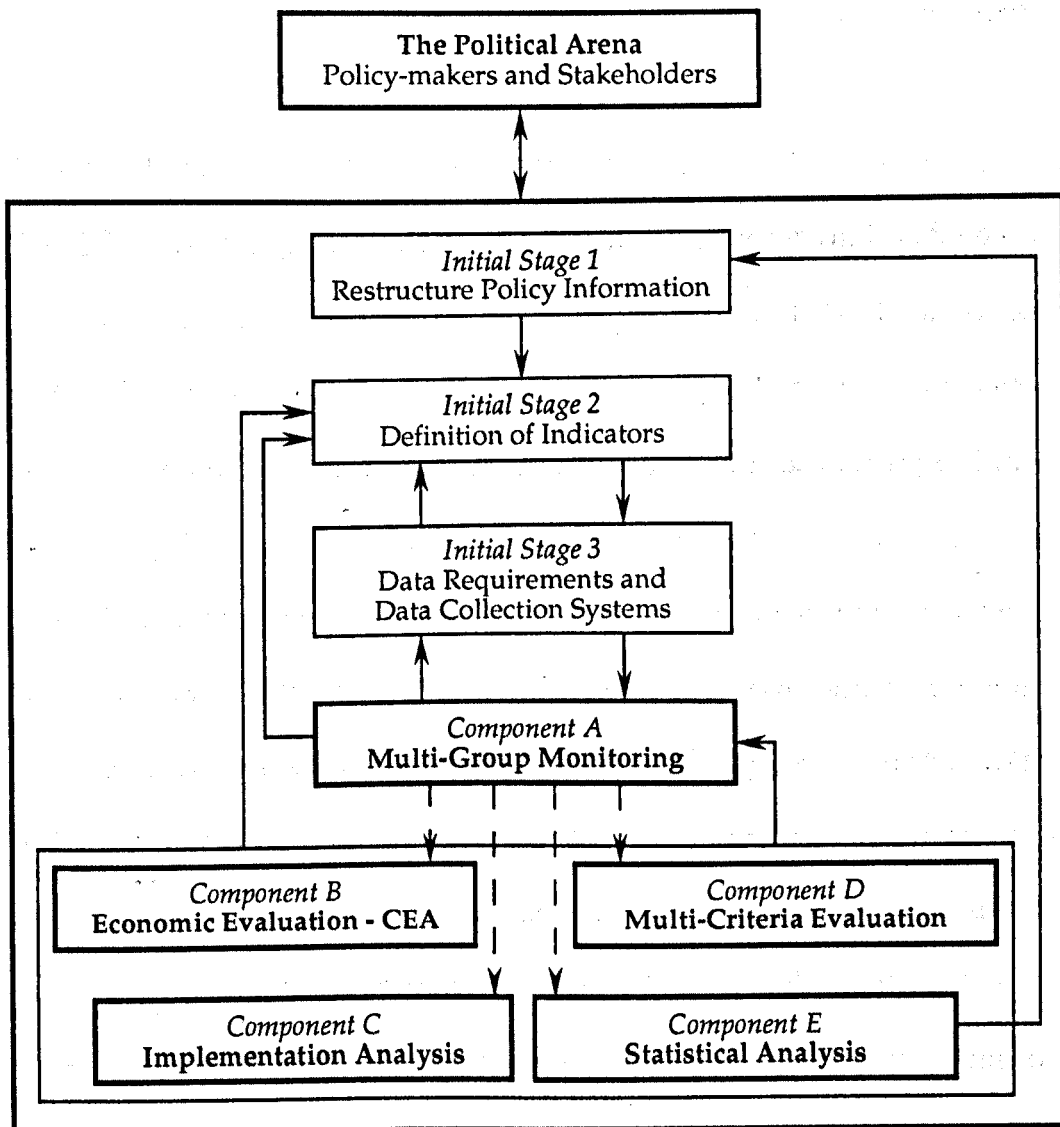


Fig 11.1: The Proposed Approach

This is no trivial task. The answers to these questions are one of the bases upon which performance indicators will be defined (*Initial Stage 2*). Another, and perhaps more important, basis for defining indicators is the questions policy-makers and stakeholders have about the policy at hand. Evaluation has to be more context responsive otherwise it will be an academic exercise (eg Hill, 1985a: 31). Evaluators, therefore, have to initiate, and maintain, a dialogue with all interested parties right from the earliest stage possible in evaluation.

In order to maximize their utility, there are several considerations that have to be taken into account when defining indicators (eg Jackson, 1988: 12 and Likierman, 1993: 15-21). For instance, the set of indicators should be a comprehensive one, in the sense that it includes all factors central to the policy and important to interested groups. At the same time, the number of indicators should be appropriate both to the diversity of the policy and the audience which indicators aim to inform. Policy performance is not independent from the environment within which it takes place. Therefore, the organizational structure(s) and the uncertainty and complexity of the external environment are all factors that have to be taken account of. Perhaps the most crucial design criterion is to avoid the tendency for easily measurable indicators. In order both to provide clearer answers to stakeholders' questions and to facilitate further stages of the evaluation process, performance indicators have to be defined as the closest proxies feasible of policy objectives, not merely what is readily available or measurable.

The definition of indicators is the basis on which data and information requirements can be determined (*Initial Stage 3*). Having done that, data and information collection system(s) should be designed and set in place. The 'optimal' timing to set up such systems is *prior* to the start of implementation. For one reason, this will ensure a 'more accurate' estimate of the base-line condition which, in turn, will help in estimating policy impacts. Moreover, programme records can be designed in a way that facilitates data collection without increasing the work load on either programme staff or evaluators.

It should go without saying that the information needed will be both quantitative and qualitative. Large quantities of different types of information are likely to be acquired and need to be processed. A computerised information system may represent the best option to handle the task. It should be noted, however, that the primary function of such systems is to provide the users with the *right* information at the *right* time. This should be the principal criterion of choice among available alternatives. It should also be noted that the need may arise to conduct survey(s) and/or questionnaire(s) to assess different groups' perceptions of policy impacts. If that was the case, and it should be, the design of these surveys and/or questionnaires should be given due consideration at this stage.

Main Components

- *Multi-Group Monitoring*: These three initial steps should pave the way for the main components of the approach. The first of these components is what

is termed here 'multi-group monitoring' (*Component A*). Regular systematic monitoring is an indispensable source of information on policy outputs, and possibly outcomes. The clear emphasis on regular monitoring built into the latest urban initiatives (eg CC and SRB) serves only to confirm policy-makers' growing concern about such information (regardless of their reasons). Monitoring, therefore, has a vital role to play in the evaluation process. It represents a quick response to policy-makers' and stakeholders' demands for much needed information. As such, monitoring can overcome the oft-cited criticism of delayed (and, irrelevant) evaluation results. It can help bridge the gap of trust and communication between evaluators and policy-makers. Monitoring information is also a vital input to the management and implementation process. Moreover, and as will be discussed later, monitoring is the source of information needed for other components of the approach.

Public policies affect different groups of the society in different ways. If distributional effects were to be taken into account, it follows, the incidence of policy outputs/outcomes should be incorporated explicitly into the analysis. That is, monitoring, and consequently evaluation, will have to establish who gains and who loses as a result of the policy in question. There are strong indications of growing concern about equity considerations. The guidelines for the monitoring and evaluation of the SRB, for instance, have emphasised the need to ensure that those intended to benefit from a scheme do so (DoE, 1995a and b). These guidelines have also stressed such concepts as 'ethnic monitoring' (DoE, 1995b) and 'equality of opportunity' (DoE, 1994a).

A multi-group monitoring should, however, adopt an *extended view* and move away from the traditional, control view which seems to underline such guidelines. Monitoring should not be confined to merely counting outputs (implementation monitoring). It should take into account the political, administrative and organizational context within which the policy is operating. It should also take into consideration emerging policies, potentials and problems otherwise policy modifications will be out-dated before they are even implemented. This information, in itself, is a vital input to implementation analysis and can aid considerably in disentangling policy impacts from those of other policies and trends. Moreover, if performance indicators were defined in a way that better reflects policy objectives, monitoring can provide (at least) a first indication of policy impacts (impact monitoring).

In brief, an extended view of regular, systematic multi-group monitoring can both provide *timely* answers to several questions and secure data and information required for further analysis. That is why monitoring is believed to be the most appropriate starting point for the evaluation process. It can lead to any, or all, of the remaining components of the approach. However, it has to be stressed here that there is no "best" sequence to follow; indeed, there is no recommended sequence for the analysis once the monitoring process has been set in place. As we shall explain later, the choice and sequence of different components both depend largely on the questions policy-makers and stakeholders ask.

• *Economic Evaluation:* As it focuses, among other things, on outputs/outcomes and expenditure, monitoring is a precursor to economic evaluation (*Component B*). Policy-makers and stakeholders are, understandably, concerned about the economy, efficiency and effectiveness of public policies. This is evidently clear in the growing emphasis on achieving greater 'value for money' and public sector accountability. Cost-effectiveness analysis (CEA) is believed to be a 'more appropriate' method for economic evaluation of urban policies than cost-benefit analysis (CBA) or value for money (VFM). CEA has the relative advantage over CBA in that it avoids the latter's major drawback of measuring all impacts in money units. The only advantage of VFM over CEA is the explicit emphasis on economy and efficiency. This is, arguably, implicit in CBA, and subsequently CEA, being derived from economic theory.

The fact that policy output/outcomes have been measured in a multi-group fashion goes some way towards incorporating equity considerations within economic evaluation. Problems, however, remain. For instance, it is very likely that different groups will have different perceptions of alternative uses of public resources. That is, it is unlikely that a consensus can be reached as to the opportunity costs of those resources expended in the implementation of the policy. This further complicates the analysis in the sense that no agreement may be reached as to the 'real' costs of the policy. Another problem is the discounting of intangibles. There is simply no resolution to this issue but to compare different streams of costs and benefits

rather than aggregate figures. Similarly, the lack of a common denominator among policy outputs, and the fact that some costs may be intangibles, both mean that an 'overall effectiveness index' is almost impossible to arrive at.

- *Implementation Analysis*: Having considered the political and administrative context of implementation, monitoring can also lead to implementation analysis (*Component C*). The definition of the influential factors behind both the outputs/outcomes of the policy and the process through which it is being put into place will facilitate the modifications of a current programme to make it more efficient. Such information will, no doubt, be of great value in planning for future policies, if we were to build on success and avoid past mistakes. Implementation analysis can also facilitate recommendations on the applicability of a programme, or parts of it, in other locations. It can also help disentangle the effects of a policy from those of extraneous factors. Moreover, implementation analysis can reduce the possibilities of making errors type I and II (false positive and false negative).

The choice of the analytical model depends, to a large extent, on the prevailing policy-making mode, or style (see chapter 6). The recent government urban initiatives (eg CC and SRB) exhibit a clear shift towards the lower-levels of the policy-making process. They also exhibit an explicit emphasis on local groups and actors. These are all evident in the significance attached to the idea of 'local partnerships'; a key proposition of both CC and the SRB. However, in both initiatives the central government reserves an

important role to play - the release of funding. If that is the case, it is believed that implementation analysis should focus on the lower-levels of the process - the partnerships. It has, nevertheless, to take into account how the role of the central government influences the decisions and actions of local actors and the commitment of partnerships.

In general terms, the synthesis of the top-down and bottom-up approaches (Sabatier, 1986) represents the 'most appropriate' starting point for analysis. It combines the focus of both approaches and provides the analyst with the freedom to place emphasis on either the Centre or the Periphery in accord with the prevailing policy-making and implementation mode(s).

- *Multi-Criteria Evaluation*: Given that indicators have been defined as the closest proxies feasible to *all* policy objectives, and that policy performance has been measured in respect to each and all of these indicators in a multi-group fashion, monitoring can also facilitate the conduct of a multi-criteria evaluation (MCE) (*Component D*). Performance indicators, as such, represent what is known as 'evaluation criteria' in standard MCE. However, and unlike monitoring, evaluation will require a detached and critical look at policy objectives and how they are being/have been achieved. Such an exercise will benefit from information on different groups' objectives and preferences. This information should have been made available through either representatives of these groups or - when possible - surveys and/or questionnaires, or both. To arrive at an overall picture of policy achievements, the task will depend,

to a large extent, on value judgement if only for the fact that policy objectives are usually vaguely stated, or not quantified. Moreover, at this stage, it is somewhat unknown whether or not outputs/outcomes are solely attributable to the policy at hand, though monitoring and implementation analysis can help reduce this uncertainty. Therefore, the results will have to be treated with caution.

- *Statistical Analysis*: One way of disentangling the effects of the policy from those of others is by means of statistical analysis (*Component E*). One alternative is before-after and/or time series analysis. Again, regular, extended-view monitoring is of considerable aid to such analysis since it provides the data and information needed. It should be noted, however, that the relative simplicity of such analysis comes somewhat at the expense of the results. This analysis will indicate a policy impact; it remains yet to estimate the magnitude of this impact.

Another alternative is more sophisticated statistical analysis such as regression analysis. Once again, monitoring is of considerable value for this analysis as it secures the data and information required, not only in regard to the policy at hand but also in respect to other policies and trends that took place within the same environment. Implementation analysis can further facilitate regression analysis by highlighting the various chains of decisions and actions and thus help define the 'cause-effect' hypotheses underlying such statistical analysis. The major drawback of regression analysis is its data

requirements; it requires large quantities of quantitative data. Moreover, intangibles, self-evidently, cannot be incorporated into the analysis. The treatment of such effects will remain reliant on simpler approaches such as before-after analysis.

Feedback Loops

The approach contains a number of feedback loops. In the design and setting up of data collection systems (*Initial Stage 3*), it may emerge that the data required for one, or more, of the indicators is not available and/or cannot be obtained. Evaluators will, therefore, have either to re-define this indicator in an operational way or discard it and define others. During the monitoring process (*Component A*), evaluators may realise that some information is missing. In this instance, the data collection system (*Initial Stage 3*) will have to be revised and modified accordingly. Indeed, such systems should themselves be the subject of regular monitoring and, when needed, (objective) improvements.

It may also emerge that certain aspects, activities or objectives are not fully covered. In this case, monitoring will lead back to the 'definition of indicators' (*Initial Stage 2*) to define, or re-define, appropriate indicators. This, in itself, may lead to the modifications of the data collection system(s). Any, and all, of the remaining components of the approach can lead back to as early a stage as the 'definition of indicators' if certain aspects are not covered. Finally, statistical analysis, particularly regression analysis, can lead back to the very

beginning of the whole process (*Initial Stage 1*) if the 'cause-effect' hypotheses of the policy were somewhat unclear.

11.3 DIFFERENT MODES OF THE APPROACH

Admittedly, the proposed approach involves some conceptual and practical difficulties, though these are not unique to it. Displacement, distributional and equity effects raise a number of problems. The 'benefits' to the targeted area/population may be at the expense of costs to other areas/groups. Unfortunately, there are no guidelines as to how to define the coverage of evaluation. The only rule to bear in mind is that confining evaluation to the targeted area/population is destined to overlook policy impacts elsewhere (spill-over effects). However, too wide a coverage will complicate the analysis. Similarly, there are no guidelines as to how to best define different groups of the society. The problem is further compounded by the fact that different policy objectives may require different classifications of groups. At the outset, there is likely to be no consensus as to what is an equitable distribution and what is not.

The most critical problem such an approach is likely to face is the sheer amount of resources it requires (time, personnel and financial) which may prohibit its application. This is probably the reason why the 'Integrated Evaluation Approach' (Alterman *et al.*, 1984) is seldom used (Carmon, personal contact, November 1995). In an ideal world, policy-makers and stakeholders are committed to evaluation and, therefore, resources are made

available to apply the approach in its 'full mode' as described above. However, reality is always different. For one reason or another, it may not be possible to carry out evaluation as such. The proposed approach, however, is sufficiently flexible to adapt to different circumstances and respond to different needs; it can adopt a number of different modes (see Fig 11.2).

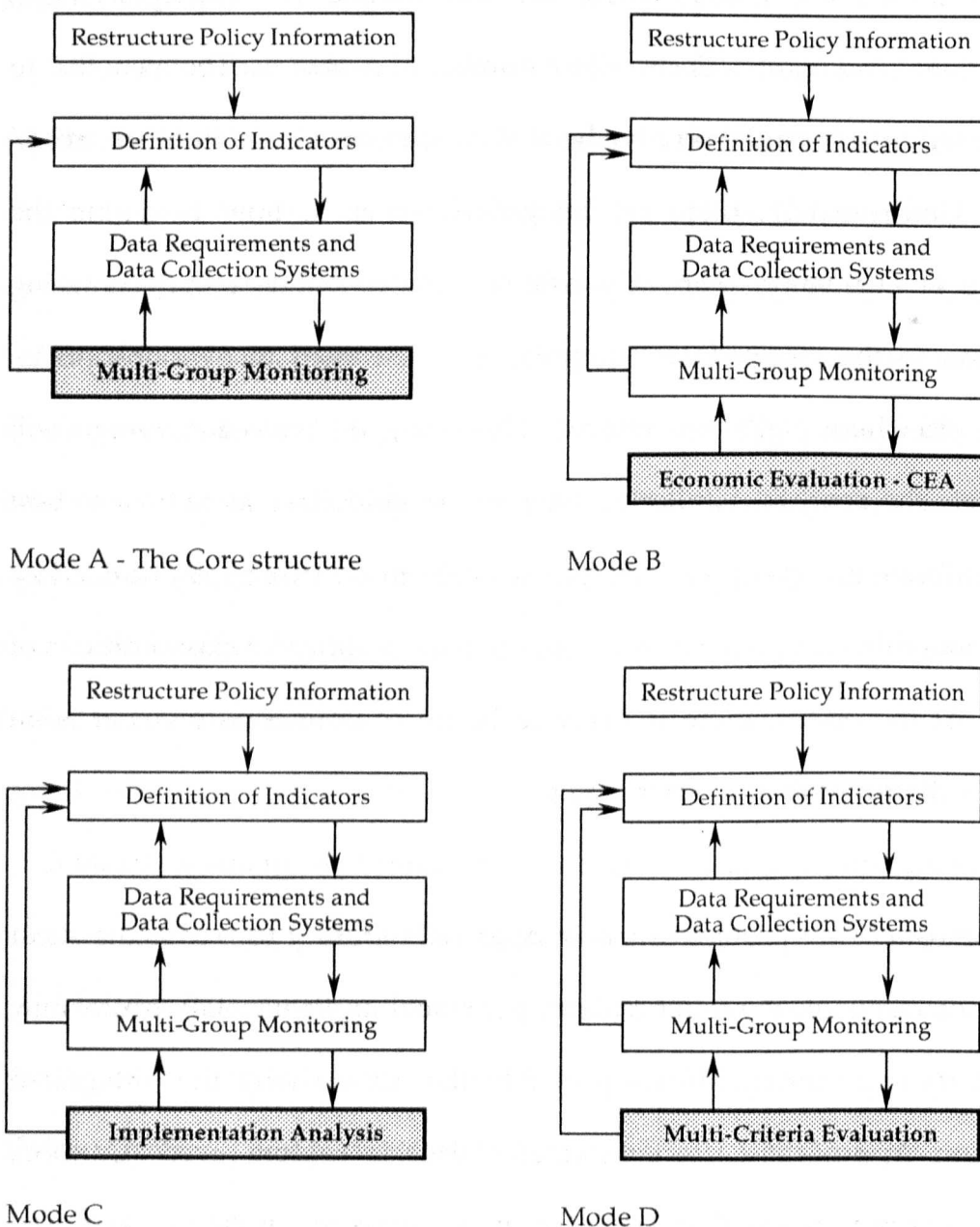


Fig 11.2: Some Alternative Modes of the Proposed Approach

One such mode, we call 'mode A', would involve the initial tasks of the approach and only multi-group monitoring. As such, policy-makers and stakeholders will be provided with the required information on a regular basis - at, or near, the right timing. The definition of performance indicators as the closest proxies feasible of policy objectives can play a significant role in enhancing the quality of information provided and, at the same time, facilitate further analysis when possible, or required. This is seen as the 'core structure' of the proposed approach. It represents a quick response to continuous demands of different types of information while facilitating, or leading, to further modes of the approach.

Alternative modes of the approach will include its 'core structure' and any of the other components in accord with stakeholders demands, or commitment. Mode B, for instance, further includes CEA if the main concern is with the economic aspects of the policy at hand and its performance. If, as was the case in the local evaluation of CC, the main concern is about the 'process' of the policy, mode C will focus on implementation analysis. Mode D, by including MCE, can provide policy-makers with, at least, a 'first indication' of policy achievements without long time requirements. Other modes can also be formulated by 'adding' more than one component to the 'core structure' of the approach. In fact, the approach can be applied on the basis of 'progressive commitment' with further analysis (other components) being conducted when questions arise, or resources are made available.

The major advantage of the approach is its 'context responsiveness'; it can respond to varying demands and commitments on the part of policy-makers and stakeholders. It is a flexible approach; at no time is there a rigid sequence to follow; at no time are the questions to be asked defined beforehand. Evaluators have the main guidelines and the details of the process will be defined in accord with the situation at hand; indeed, that is how it should be.

11.4 THE WAY FORWARD

It is nevertheless crucial that policy-makers are alerted to the need for systematic evaluation; the need to move away from merely counting outputs to disentangling policy *impacts*. This is probably the trickiest task evaluators will have to handle. Policy-makers seem to be concerned primarily with "doing something", or appearing to be so. Each politician seems keen on "leaving his/her mark"; the usual way is to introduce a new policy. The evidence is the wide array of policies introduced one after the other without any re-appraisal of current ones. There is a need to promote an understanding of the significance of evaluation within political circles, and hence a commitment to evaluation research. This is very likely to be a long-term endeavour. It is not an impossible objective, however; the Dutch experience shows clearly and beyond doubt that it can be achieved.

Evaluators will have to alert policy-makers to the inherent dangers in short-termism, and consequently the need to re-consider current policies before new ones are introduced. Evaluators may be aided in this task by

policy-makers' call for greater value for money. It is clear that the duplication of existing programmes or terminating effective ones do not conform with this call. The sheer financial - and, indeed, political - costs of wrong public decisions is an argument that is unlikely to fall on deaf ears. Evaluators may be further aided by the pressure other stakeholders are likely to exert for better informed, more rational decisions.

However, it is simply unrealistic to expect a complete "U turn" on the part of policy-makers. Evaluators have to realise the pace at which changes occur and, consequently, the speed at which decisions may have to be made. Therefore, there is a need to be somewhat pragmatic and find a 'compromise', a way of providing policy-makers with clearer answers on a short notice; a compromise that does not, however, come at the expense of research quality. That is where an extended view of multi-group monitoring - with indicators defined as the closest proxies feasible of policy objectives - comes into play (as we have discussed above).

This extended view of multi-group monitoring, in itself, will call for further research in two directions. Firstly, to define better, more appropriate indicators. Secondly, to secure the data and information needed to make these indicators operational. In our review of the case studies of urban policy evaluation (chapter 8) it has been noted how less appropriate indicators have undermined the results of the research and left the policy's real impacts somewhat unknown. On the other hand, it is not uncommon that more

appropriate indicators are discarded simply because of the absence of required data. Assessing the effects of regional economic policy in Britain is a good example of such situations where unemployment rates were considered an inappropriate indicator of policy effects because of the lack of information on other factors influencing unemployment (see chapter 9). Almost all those who have been contacted during the research cited data availability among the most common and severe problems encountered in their practice, or research.

An extended view of monitoring can largely benefit from - and actually calls for - more co-ordination between the different parties involved in the planning and implementation process. This can facilitate the flow of information and, thus, reduces both the work load and possibly the costs of evaluation. It may also lead to better co-ordinated policies that do not run counter to each other. The SRB and the re-structuring of Government Offices - by bringing some 20 programmes and four central departments together - can certainly be seen as a step forward in this direction. However, it has to be noted that co-ordination requires more than the allocation of different groups in the same premises. It may actually involve some extra work on the part of each group. It certainly calls for a more open organizational culture whereby information is shared among different groups. As such, resistance is likely to surface unless participants shared an understanding of the 'added value' of their co-operation.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ackermann, C and Steinmann, W (1982) Privatized Policy-Making: Administrative and Consociational Types of Implementation in Regional Economic Policy in Switzerland, *European Journal of Political Research*, 10: 173-185.

Adamowicz, W L, Bhardwaj, V and Macnab, B (1993) Experiments on the Difference Between Willingness to Pay and Willingness to Accept, *Land Economics*, 69(4): 416-427.

Advisory Committee on Trunk Road Assessment - ACTRA (Chairman Sir G Leitch) (1978) Report of The Advisory Committee on Trunk Road Assessment, HMSO, London.

Alexander, E R (1989) Improbable Implementation: The Pressman-Wildavsky Paradox Revisited, *Journal of Public Policy*, 9(4): 451-465.

Alexander, I (1978) The Planning Balance Sheet: An Appraisal, in: McMaster, J C and Webb, G R (eds.) Australian Project Evaluation: Selected Readings, Australia and New Zealand Book Co., Sydney, pp. 46-68.

Allen, B L, Butterfield, D W, Kazakov, A, Kliman, M L, Kubursi, A A and Welland, J D (1988) Measuring Economic Stimulation from Capital Investment in Transportation, *Transportation Research Record*, (1197): 49-55.

Allen, K, Begg, H, McDowall, S and Walker, G (1986) Regional Incentives and the Investment Decision of the Firm, HMSO, London.

Alterman, R (1980) Decision-Making in Urban Plan Implementation: Does the Dog Wag the Tail or the Tail Wag the Dog?, *Urban Law and Policy*, 3: 41-58.

Alterman, R (1981) Implementation Analysis in Urban and Regional Planning: Toward A Research Agenda, paper presented at the Conference on Planning Theory in the 1980's, Oxford Polytechnic, April 2-4.

Alterman, R (1983) Implementation Analysis: The Contours of an Emerging Debate: Review Article, *Journal of Planning Education and Research*, 2(3): 63-65.

Alterman, R (1987) Opening Up the "Black Box" in Evaluating Neighborhood Problems: The Implementation Process in Israel's Project Renewal, *Policy Studies Journal*, 16(2): 347-361.

Alterman, R (1988) Implementing Decentralization for Neighborhood Regeneration: Factors Promoting or Inhibiting Success, *Journal of the American Planning Association*, 54(4): 454-469.

Alterman, R (1991a) Planning and Implementation of Israel's Project Renewal: A Retrospective View, Chap. 10, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 147-169.

Alterman, R (1991b) Dilemmas About Cross-National Transferability of Neighbourhood Regeneration Programmes, Chap. 15, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 209-221.

Alterman, R, Carmon, N and Hill, M (1984) Integrated Evaluation: A Synthesis of Approaches to the Evaluation of Broad-Aim Social Programs, *Socio-Economic Planning Sciences*, 18(6): 381-389.

Amos, F J C (1980) Development Control Reappraised, in: Value for Money in Development Control: A Series of Essays, Institute of Local Government Studies (INGOLOV), University of Birmingham, pp. 1-4.

Anderson, R J and Crocker, T D (1971) Air Pollution and Property Values, *Urban Studies*, 8(3): 171-180.

Andresen, S J, Mahmassani, H S, Helaakoski, R, Euritt, M A, Walton, C M and Harrison, R (1993) Economic Impact of Highway Bypasses, *Transportation Research Record*, (1395): 144-152.

Armour, A (1990) Integrating Impact Assessment in the Planning Process: From Rhetoric to Reality, *Impact Assessment Bulletin*, 8(1): 3-14.

Armstrong, H (1986) The Assignment of Regional Industrial Policy Powers, *Regional Studies*, 20(3): 258-261.

Armstrong, H and Taylor, J (1985) *Regional Economics and Policy*, Philip Allen, Oxford.

Armstrong, H and Taylor, J (1993) *Regional Economics and Policy*, 2nd ed., Harvester Wheatsheaf, Hertfordshire.

Ashcroft, B (1978) The Evaluation of Regional Economic Policy: The Case of the United Kingdom, Studies in Public Policy no. 12, Centre for the Study of Public Policy, University of Strathclyde, Glasgow.

Ashcroft, B and Taylor, J (1977) The Movement of Manufacturing Industry and the Effect of Regional Policy, *Oxford Economic Papers*, 29(1): 84-101.

Ashcroft, B and Taylor, J (1979) The Effect of Regional Policy on the Movement of Industry in Great Britain, Chap. 2, in: Maclennan, D and Parr, J B (eds.) Regional Policy: Past Experience and New Directions, Martin Robertson, Oxford, pp. 43-64.

Bakr, A H (1994) The Single Regeneration Budget: Reflections on the First Bidding Round, WP 52, Working Paper Series, Department of Civic Design, University of Liverpool, December, 1994.

- Ballard, S C, Devine, M D, James, T E and Chartock, M A (1982)** Integrated Regional Environmental Assessment: Purposes, Scope, and Products, *Impact Assessment Bulletin*, 2(1): 5-13.
- Banks, T (1990)** Performance Measurement: The Needs of Managers and Policy Makers, *Public Money & Management*, 10(2): 47-49.
- Barlow Report (1940)** Report of the Royal Commission on the Distribution of the Industrial Population, HMSO, London (only as quoted in Diamond and Spence, 1983).
- Barnekov, T and Hart, D (1990)** The Changing Nature of US Urban Policy Evaluation: The Example of the Urban Development Action Grant, paper presented at the Annual Conference of the British Section of the Regional Science Association, University of Liverpool, 6 September 1990.
- Barnekov, T, Hart, D and Benfer, W (1990)** U.S. Experience in Evaluating Urban Regeneration, (A study commissioned by the DoE), HMSO, London.
- Barrett, S and Fudge, C (1981)** Examining the Policy-Action Relationship, in: Barrett, S and Fudge, C (eds.) Policy and Action, Methuen, London, pp. 3-32.
- Bartels, C P A, Nicol, W R and van Duijn, J J (1982)** Estimating the Impact of Regional Policy: A Review of Applied Research Methods, *Regional Science and Urban Economics*, 12: 3-41.
- Batey, P (1983)** Strategic Methods In Strategic Land Use Planning: Some Reflections on Recent British Experience, in: Batty, M and Hutchinson, B (eds.) Systems Analysis in Urban Policy-Making and Planning, Plenum Press, New York, pp. 69-87.
- Batty, M (1988)** Review of *Multicriteria Evaluation for Urban and Regional Planning* by H. Voogd, *Regional Studies*, 18: 433-435.
- Beacham, A and Osborn, W T (1970)** The Movement of Manufacturing Industry, *Regional Studies*, 4(1): 41-47.
- Begg, H M, Lythe, C M and Macdonald, D R (1976)** The Impact of Regional Policy on Investment in Manufacturing Industry: Scotland, 1960-71, *Urban Studies*, 13(2): 171-179.
- Begg, H M and McDowall, S (1987)** The Effect of Regional Investment Incentives on Company Decisions, *Regional Studies*, 21(5): 459-470.
- Bennett, R J (1978)** Regional Monitoring in the U.K.: Imperatives and Implications for Research, *Regional Studies*, 12: 311-321.
- Bentley, P, Collins, D J and Rutledge, D J S (1979)** The Distributional Impact of Road Expenditure and Finance in Australia, *Journal of Transport Economics and Policy*, 13(2): 182-208.

- Berg, R S** (1967) Armed Forces' Use of Cost-Effectiveness Analysis, Chap. 6, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Praeger, Washington, DC, pp. 91-103.
- Berk, R A and Rossi, P H** (1990) Thinking About Program Evaluation, Sage, Newbury Park.
- Blake, C** (1972) The Effectiveness of Investment Grants as a Regional Subsidy, *Scottish Journal of Political Economy*, 19: 63-71.
- Blumstein, A** (1967) The Choice of Analytic Technique, Chap. 3, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Praeger, Washington, DC, pp. 33-43.
- Boukaert, G** (1994) A Review of 'Managing Value for Money in the Public Sector' by Bates, Jonathan G (Chapman & Hall, London, 1993), *Public Money & Management*, 14(1): 60-61.
- Bovaird, T, Gregory, D and Martin, S** (1988) Performance Measurement in Urban Economic Development, *Public Money & Management*, 8(4): 17-22.
- Bowen, E R** (1982) The Pressman-Wildavsky Paradox: Four Addenda or Why Models Based on Probability Theory Can Predict Implementation Success and Suggest Useful Technical Advice for Implementers, *Journal of Public Policy*, 2(1): 1-22.
- Bracken, I** (1981) Urban Planning Methods: Research and Policy Analysis, Methuen, London.
- Bradford, M and Robson, B** (1995) An Evaluation of Urban Policy, Chap. 3, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 37-54.
- Briggs, R** (1981) Interstate Highway System and Development in Nonmetropolitan Areas (Abridgment), *Transportation Research Record*, (812): 9-12.
- Brignall, S** (1993) Performance Measurement and Change in Local Government: A General Case and a Childcare Application, *Public Money & Management*, 13(4): 23-30.
- Brokenshire, D P** (1977) Monitoring and the Planning Process, proceedings of seminar B: Policy Analysis for Urban and Regional planning, PTRC Summer Annual Meeting, June 1977, pp. 127-136.
- Brook, C A** (1980) Value for Money in Development Control: Current Procedures and Options in a Metropolitan District, in: Value for Money in Development Control: A Series of Essays, Institute of Local Government Studies (INGOLOV), University of Birmingham, pp. 15-26.
- Brown, P J B** (1984) Monitoring and Regional Information Systems under Uncertainty, Chap. 5, in: Nijkamp, P and Rietveld, P (eds.) Information Systems for Integrated Regional Planning, North-Holland, Amsterdam, pp. 81-100.

- Browne, A and Wildavsky, A (1984a)** What Should Evaluation Mean to Implementation, Chap. 9, in: Pressman, J L and Wildavsky, A (eds.) Implementation, 3rd ed., University of California Press, Berkeley, pp. 181-205.
- Browne, A and Wildavsky, A (1984b)** Implementation as Mutual Adaptation, Chap. 10, in: Pressman, J L and Wildavsky, A (eds.) Implementation, 3rd ed., University of California Press, Berkeley, pp. 206-231.
- Browning, R P, Marshall, D R and Tabb, D H (1981)** Implementation and Political Change: Sources of Local Variations in Federal Social Programs, Chap. 5, in: Mazmanian, D A and Sabatier, P A (eds.) Effective Policy Implementation, Lexington Books, Massachusetts, pp. 127-146.
- Buck, T and Atkins, M (1983)** Regional Policies in Retrospect: An Application of Analysis of Variance, *Regional Studies*, 17(3): 181-189.
- Buck, T W (1970)** Shift and Share Analysis - A Guide to Regional Policy?, *Regional Studies*, 4(4): 445-450.
- Buckley, M (1988)** Multicriteria Evaluation: Measures, Manipulation and Meaning, *Environment and Planning B*, 15(1): 55-64.
- Buit, J (1985)** Explicit Evaluation in a Period of Decline and Scarcity, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 67-77.
- Burdge, R J (1991)** A Brief History and Major Trends in the Field of Impact Assessment, *Impact Assessment Bulletin*, 9(4): 93-104.
- Burton, P and Boddy, M (1995)** The Changing Context for British Urban Policy, Chap. 2, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 23-36.
- Butt, H and Palmer, B (1985)** Value for Money in the Public Sector: The Decision-Maker's Guide, Blackwell, Oxford.
- Button, K (1995)** What Can Meta-analysis Tell Us about the Implications of Transport, *Regional Studies*, 29(6): 507-517.
- Bynner, J, McCormick, R and Nuttall, D (1982)** Organization and Use of Evaluation, Curriculum Evaluation and Assessment in Educational Institutions 6, The Open University, Milton Keynes.
- Cameron, G C (1990)** First Steps in Urban Policy Evaluation in the United Kingdom, *Urban Studies*, 27(4): 475-495.
- Carlén, G and Cars, G (1991)** Renewal of Large-Scale Post-War Housing Estates in Sweden: Effects and Efficiency, Chap. 9, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 130-146.

- Carley, M** (1988) Beyond Performance Measurement in a Professional Public Service, *Public Money & Management*, 8(4): 23-27.
- Carmon, N** (1987) A Neighborhood Program That Works: Israel's Project Renewal, *Policy Studies Journal*, 16(2): 262-276.
- Carmon, N and Hill, M** (1988) Neighborhood Rehabilitation Without Relocation or Gentrification, *Journal of the American Planning Association*, 54(4): 470-481.
- Carmon, N, Hill, M and Alterman, R** (1980) Multi-Group Integrated Evaluation: Synthesis of Approaches to the Evaluation of Broad-Aim Programs, Technion, Haifa, Israel.
- Carter, T** (1991) Neighbourhood Improvement: The Canadian Experience, Chap. 2, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 9-27.
- Chadwick, G** (1971) A Systems View of Planning, (only as quoted in Shefer and Tsubari, 1990 and, Voogd, 1983a), Pergamon Press, Oxford.
- Chamberlain, K** (1976) Conference 1976: 'Information Systems for Policy Planning' - Overview, *BURISA*, 25: 3-4.
- Chelimsky, E** (1985a) Comparing and Contrasting Auditing and Evaluation: Some Notes on Their Relationship, *Evaluation Review*, 9(4): 483-503.
- Chelimsky, E** (1985b) (ed.) Program Evaluation: Patterns and Directions, (only as quoted in the previous reference), Classics VI, ASPA PAR, Washington, DC.
- Chelimsky, E** (1987a) Linking Program Evaluation to User Needs, Chap. 3, in: Palumbo, D J (ed.) The Politics of Program Evaluation, Sage, Newbury Park, CA, pp. 72-99.
- Chelimsky, E** (1987b) The Politics of program Evaluation, *Society*, 25(1): 24-32.
- Chisholm, M** (1976) Regional Policies in an Era of Slow Population Growth and Higher Unemployment, *Regional Studies*, 10(2): 201-213.
- Churchman, A** (1987) Issues in Residential Participation: Lessons from the Israeli Experience, *Policy Studies Journal*, 16(2): 290-299.
- Cleary, E J and Thomas, R E** (1973) The Economic Consequences of the Severn Bridge and Its Associated Motorways, Bath University Press, Bath.
- Cohon, J L** (1978) Multiobjective Programming and Planning, Academic Press, New York.
- Command Paper no. 7845 (cmd. 7845)** (1980) The Government Green Paper on the Role of the Comptroller and Auditor General, HMSO, London.

Connor, A (1993) *Monitoring and Evaluation Made Easy: A Handbook for Voluntary Organizations*, HMSO, Edinburgh.

Cook, T D and Leviton, L C (1980) Reviewing the Literature: A Comparison of Traditional Methods with Meta-analysis, *Journal of Personality*, 48(4): 449-472.

Copp, D (1987) The Justice and Rationale of Cost-Benefit Analysis, *Theory and Decision*, 23(1): 65-87.

Cowling, T M and Steeley, G C (1973) *Sub-Regional Planning Studies: An Evaluation*, (Chapter 8: Monitoring and Evaluation), Pergamon Press, Oxford.

Daru, R (1985) Evaluation by Graphics, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 89-107.

Davies, T and Mason, C (1982) Gazing Up at the Bottoms: Problems of Minimal Response in the Implementation of Manpower Policy, *European Journal of Political Research*, 10: 145-157.

Davoudi, S and Healey, P (1995) City Challenge - A Sustainable Mechanism or Temporary Gesture?, Chap. 11, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 158-174.

Dello, P (1985) Strategic Choice and Evaluation: Some Methodological Considerations, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 109-123.

Del Monte, A and De Luzenberger, R (1989) The Effect of Regional Policy on New Firm Foundation in Southern Italy, *Regional Studies*, 23(3): 219-230.

Department of the Environment and Department of Transport (1994) *Planning Policy Guidance: Transport*. PPG 13, HMSO, London.

Department of the Environment - DoE (1972) *Getting the Best Roads for Our Money - The COBA Method of Appraisal*, HMSO, London.

Department of the Environment - DoE (1974) *Structure Plans, Circular 98/74*, DoE, London.

Department of the Environment - DoE (1975) *Principles of Monitoring Development Plans*, Structure Plan Note 1/75, DoE, London.

Department of the Environment - DoE (1992) *Policy Evaluation: The Role of Social Research*, HMSO, London.

Department of the Environment - DoE (1994a) *Bidding Guidance: A Guide to Funding from the Single Regeneration Budget*, DoE, London (April, 1994).

Department of the Environment - DoE (1994b) *Index of Local Conditions: An Analysis Based on 1991 Census Data*, DoE, London.

Department of the Environment - DoE (1994c) News Release, 29 November 1994, DoE, London.

Department of the Environment - DoE (1995a) Single Regeneration Budget: Partnership Delivery Plans, SRB Guidance Note No. 1, February, 1995, DoE, London.

Department of the Environment - DoE (1995b) Single Regeneration Budget Bidding Round: Monitoring and Periodic Review, SRB Guidance Note No. 2, February, 1995, DoE, London.

Department of the Environment - DoE (1995c) Single Regeneration Budget Bidding Round: Project Appraisal and Approval Guidance, SRB Guidance Note No. 3, February, 1995, DoE, London.

Department of the Environment - DoE (1995d) Single Regeneration Budget: Financial Guidance, SRB Guidance Note No. 4, February, 1995, DoE, London.

Department of the Environment - DoE (1995e) Bidding Guidance: A Guide to Bidding for Resources from the Government's Single Regeneration Budget Challenge Fund, DoE, London (April, 1995).

Department of the Environment - Inner Cities Grants Division (1993) City Challenge Guidance Note: Appraisal, Monitoring and Evaluation, DoE-ICGD, London (December 1993).

Department of Transport - DoT (1981) COBA 9 Manual, DoT, London (revisions up to January 1992).

Department of Transport - DoT (1983) MEA: Manual of Environmental Appraisal, DoT, London.

Department of Transport - DoT (1989) Roads for Prosperity (cmdnd 693), HMSO, London.

Department of Transport - DoT (1991) TAM: Traffic Appraisal Manual, DoT, London.

Department of Transport - DoT (1992) Assessing the Environmental Impact of Road Schemes - Response by the DoT to the Report by SACTRA, 1992, HMSO, London.

Department of Transport - DoT (1993a) Design Manual for Roads and Bridges - Vol 11: Environmental Assessment, DoT, London.

Department of Transport - DoT (1993b) Comparison of Forecast and Observed Traffic on Trunk Road Schemes, DoT, London.

Dery, D (1984) Evaluation and Termination of the Policy Cycle, *Policy Sciences*, 17(1): 13-26.

Diamond, D and Spence, N (1983) Regional Policy Evaluation: A Methodological Review and the Scottish Example, Gower, Aldershot.

- Dodgson, J S (1974) Motorway Investment, Industrial Transport Costs, and Sub-Regional Growth: A Case Study of the M62, *Regional Studies*, 8(1): 75-91 (ex-ante CBA model).
- Doeleman, J A (1985) Historical Perspectives and Environmental Cost-Benefit Analysis, *Futures*, 17(2): 149-163.
- Drummond, H (1991a) Effective Decision Making: A Practical Guide for Management, Kogan Page, London.
- Drummond, H (1991b) Power: Creating it, Using it, Kogan Page, London.
- Edwards, G C (1980) Implementing Public Policy, (only as quoted in Gurnack and Harty, 1987), Congressional Quarterly Press, Washington, DC.
- Edwards, J A (1976) Industrial Structure and Regional Change: A Shift-Share Analysis of the British Columbia Economy, *Regional Studies*, 10(3): 307-317.
- Eisenschitz, A and North, D (1986) The London Industrial Strategy: Socialist Transformation or Modernising Capitalism, *International Journal of Urban and Regional Research*, 10(3): 419-440.
- Etzioni, A (1967) Mixed-Scanning: A "Third" Approach to Decision-Making, *Public Administration Review*, 27: 385-392.
- Faludi, A and Voogd, H (1985) (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands.
- Fielden, J (1984) A Consultants Experience in Undertaking Value for Money Review, in: Richardson, J J (ed.) Value for Money and Effectiveness Auditing in Public Sector: A Symposium, Strathclyde Papers on Government and Politics 30, Department of Politics, University of Strathclyde, Glasgow, pp. 36-46.
- Finsterbusch, K (1984) Social Impact Assessment as a Policy Science Methodology, *Impact Assessment Bulletin*, 3(2): 37-43.
- Floyd, M (1978) Structure Plan Monitoring: Looking to the Future, *Town Planning Review*, 49: 476-485.
- Flynn, A, Gray, A, Jenkins, W and Ruherford, B (1988) Making Indicators Perform, *Public Money & Management*, 8(4): 34-41.
- Folmer, H (1980) Measurement of the Effects of Regional Policy Instruments, *Environment and Planning A*, 12(10): 1191-1202.
- Fothergill, S and Gudgin, G (1979) In Defence of Shift-Share, *Urban Studies*, 16: 309-319.
- Francis, K (1981) The Monitoring of Structure Plans in the 1980s, in: Voogd, H (ed.) Strategic Planning in a Dynamic Society, DUM, Delft, The Netherlands, pp. 181-192.

Friend, J (1980) Planning in a Multi-Organizational Context, *Town Planning Review*, 51(3): 261-269.

Friend, J and Hickling, A (1988) *Planning Under Pressure: The Strategic Choice Approach*, (revised print of 1st edition), Pergamon Press, Oxford.

Friend, J K and Jessop, W N (1977) *Local Government and Strategic Choice*, 2nd ed., Pergamon Press, Oxford.

Friend, J K, Power, J M and Yewlett, C J L (1974) *Public Planning: The Inter-Corporate Dimension*, Tavistock, London.

Garner, M R (1984) Managerial Audit in the Nationalized Industries: More or Less?, in: Richardson, J J (ed.) Value for Money and Effectiveness Auditing in Public Sector: A Symposium, Strathclyde Papers on Government and Politics 30, Department of Politics, University of Strathclyde, Glasgow, pp. 1-19.

Gáspár, A and Birghoffer, P (1991) The Upgrading of Large-Scale Housing Estates in Hungary, Chap. 12, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 179-186.

Gilfoyle, I and Bocking, T (1980) New Constructive Relations between County and Districts Post the Local Government Planning and Land Bill, in: Value for Money in Development Control: A Series of Essays, Institute of Local Government Studies (INLOGOV), University of Birmingham, pp. 34-40.

Gillis, J D S, Brazier, S, Chamberlain, K J, Harris, R J P and Scott, D J (1974) Monitoring and The Planning Process, The Notts./Derbys. Monitoring and Advisory Committee, Institute of Local Government Studies (INLOGOV), University of Birmingham.

Gilsinan, J F and Volpe, L C (1984) Don't Cry Wolf Until You are Sure: The Manufactured Crisis in Evaluation Research, *Policy Sciences*, 17(2): 179-191.

Goldman, T A (1967) (ed.) *Cost-Effectiveness Analysis: New Approaches in Decision-Making*, Washington Operations Research Council, Washington, DC.

Good, D A (1971) *Cost-Benefit and Cost-Effectiveness Analysis: Their Application to Urban Public Services and Facilities*, RSRI Discussion Paper no. 47, Regional Science Research Institute, Philadelphia.

Grieco, M (1994) *The Impact of Transport Investment upon the Inner City*, Avebury, London.

Grosse, R N (1967) Preface, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Approaches in Decision-Making, Washington Operations Research Council, Washington, DC, pp. v-vii.

Gurnack, A M and Harty, S S (1987) The Emergence of Policy Implementation Analysis: The Wyatt v. Stickney Decision, *International Journal of Public Administration*, 9(4): 363-396.

- Gwilliam, K M (1972) Economic Evaluation of Urban Transport Projects: The State of The Art, *Transportation Planning and Technology*, 1: 123-142.
- Gwilliam, K M and Wilson, R L (1980) Social and Economic Effects of Motorways, in: 20 Years of British Motorways, proceedings of the Conference held in London, 27-28 Feb., 1980, ICE, London, pp. 89-105.
- Hambleton, R (1983) Planning Systems and Policy Implementation, *Journal of Public Policy*, 3(4): 397-418.
- Hambleton, R and Thomas, H (1995) Urban Policy Evaluation - The Contours of the Debate, Chap. 1, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 1-20.
- Hanf, K (1982) Regulatory Structures: Enforcement as Implementation, *European Journal of Political Research*, 10: 159-172.
- Hanf, K, Hjern, B and Porter, D O (1978) Local Networks of Manpower Training in the Federal Republic of Germany and Sweden, Chap. 12, in: Hanf, K and Scharpf, F W (eds.) Interorganizational Policy Making, Sage, London, pp. 303-341.
- Hart, D (1976) Strategic Planning in London: The Rise and Fall of the Primary Road Network, Pergamon Press, Oxford.
- Hart, D (1991) US Urban Policy Evaluation in the 1980s: Lessons from Practice, *Regional Studies*, 25(3): 255-261.
- Hartgen, D T, Stuart, A W, Walcott, W A and Clay, J W (1990) Role of Transportation in Manufacturers' Satisfaction with Locations, *Transportation Research Record*, (1274): 12-23.
- Hatry, H B (1967) The Use of Cost Estimates, Chap. 4, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Praeger, Washington, DC, pp. 44-68.
- Haveman, R H (1976) Evaluating the Impact of Public Policies on Regional Welfare, *Regional Studies*, 10(4): 449-463.
- Haynes, P A (1974) Towards a Concept of Monitoring, *Town Planning Review*, 45: 5-29.
- Headicar, P and Bixby, B (1992) Concrete and Tyres: Local Development Effects of Major Roads - the M40 Case Study, CPRE (Council for the Protection of Rural England), London.
- Heggie, I G (1979) Economics and the Road Programme, *Journal of Transport Economics and Policy*, 13(1): 52-67.
- Henderson, R A (1980) The Location of Immigrant Industry within a U.K. Assisted Area: The Scottish Experience, *Progress in Planning*, 14(2): 105-226.

- Henerson, M E, Morris, L L and Fitz-Gibbon, C T (1987)** How to Measure Attitudes, 2nd ed., Program Evaluation Kit 6, Sage, Newbury Park.
- Hickling, A (1985)** Evaluation is a Five-Finger Exercise, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 125-134.
- Hill, M (1968)** A Goals Achievement Matrix for Evaluating Alternative Plans, *Journal of the American Institute of Planners*, 34(1): 19-29.
- Hill, M (1973)** Planning for Multiple Objectives: An Approach to the Evaluation of Transportation Plans (chapter 3), Monograph Series no. 5, Regional Science Research Institute, Philadelphia.
- Hill, M (1985a)** Decision Making Contexts and Strategies for Evaluation, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 9-34.
- Hill, M (1985b)** Can Multiple-Objective Evaluation Enhance Rationality in Planning?, Chap. 11, in: Breheny, M and Hooper, A (eds.) Rationality in Planning, Pion, London, pp. 166-182.
- Hill, M (1986)** Comprehensive Evaluation Study of Israel's "Project Renewal", summary of the major findings as presented in a lecture at a conference in Jerusalem, March, 1986. Technion, Haifa, Israel.
- Hill, M (1990)** A Goals-Achievement Matrix for Evaluating Alternative Plans, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, (originally published in the *Journal of the American Institute of Planners*, 1968), Pion, London, pp. 3-29.
- Hill, M and Alterman, R (1974)** Power Plant Site evaluation: The Case of the Sharon Plant in Israel, *Journal of Environmental Management*, 2(2): 179-196.
- Hill, M, Alterman, R, Carmon, N, Churchman, A and Shechter, M (1990)** Integrated Evaluation of Israel's Neighborhood Revitalization Project, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 43-62.
- Hill, M and Shechter, M (1971)** Optimal Goal Achievement in the Development of Outdoor Recreation Facilities, in: Wilson, A G (ed.) Urban and Regional Planning, London Papers in Regional Science 2, Pion, London, pp. 110-120.
- Hill, M and Tzamir, Y (1972)** Multidimensional Evaluation of Regional Plans Serving Multiple Objectives, *Papers of the Regional Science Association*, 29: 139-165.
- Hill, M and Werczberger, E (1978)** Goals Programming and the Goals-Achievement Matrix, *International Regional Science Review*, 3(2): 165-181.
- Hirschman, I and Henderson, M (1990)** Methodology for Assessing Local Land Use Impacts of Highways, *Transportation Research Record*, (1274): 35-40.

Hjern, B (1982) Implementation Research: The Link Gone Missing, *Journal of Public Policy*, 2(3): 301-308.

Hjern, B and Porter, D O (1981) Implementation Structures: A New Unit of Administrative Analysis, *Organization Studies*, 2(3): 211-227.

HM Treasury (1988) Policy Evaluation: A Guide for Managers, HMSO, London.

HM Treasury (1991) Economic Appraisal in Central Government: A Technical Guide for Government Departments (the Green Book), HMSO, London.

Holland, C J and Sherman, P M S (1981) The Link Between Transport and Economic Development, proceedings of Seminar D, PTRC Summer Annual Meeting, University of Warwick, 13-16 July 1981, pp. 67-80.

Holmes, J C (1972) An Ordinal Method of Evaluation, *Urban Studies*, 9(1): 179-191 (only as quoted in Voogd, 1983a).

Holtham, C (1988) Developing a System for Measuring Departmental Performance, *Public Money & Management*, 8(4): 29-33.

Homburger, W S (1978) The Impact of a New Rapid Transit System on Traffic on Parallel Highway Facilities, *Transportation Planning and Technology*, 4(3): 187-201.

House of Commons (1973) Regional Economic Incentives: Minutes of Evidence - Expenditure Committee (Trade and Industry Sub-Committee), Session 1972-1973, Paper 327, HMSO, London.

House of Commons (1974) Regional Development Incentives: Minutes of Evidence, Expenditure Committee (Trade and Industry Sub-Committee), Session 1973-74, Paper 85, HMSO, London.

Howells, C and Smith, M (1977) Monitoring and Policy Review in the Context of Continuous Strategic Planning in a Metropolitan County, proceedings of seminar B: Policy Analysis for Urban and Regional Planning, PTRC Summer Annual Meeting, June 1977, pp. 150-162.

Hull, C and Hiern, B (1982) Helping Small Firms Grow: An Implementation Analysis of Small Firms Assistance Structures, *European Journal of Political Research*, 10: 187-198.

Inger, R (1984) Effective Auditing: The Swedish Experience, in: Richardson, J J (ed.) Value for Money and Effectiveness Auditing in Public Sector: A Symposium, Strathclyde Papers on Government and Politics 30, Department of Politics, University of Strathclyde, Glasgow, pp. 47-67.

Institute of Local Government Studies (INLOGOV) (1980) Value for Money in Development Control: A Series of Essays, INLOGOV, University of Birmingham.

Jackson, J N (1972) The Urban Future: The Choice Between Alternatives, George Allen & Unwin Ltd., London.

Jackson, P M (1988) The management of Performance in the Public Sector, *Public Money & Management*, 8(4): 11-16.

Jackson, P M (1993) Public Service Performance Evaluation: A Strategic Perspective, *Public Money & Management*, 13(4): 9-14.

Johnes, G (1987) Regional Policy and Industrial Strategy in the Welsh Economy, *Regional Studies*, 21(6): 555-564.

Jones, J (1986) An Examination of the Thinking behind Government Regional Policy in the UK since 1945, *Regional Studies*, 20(3): 261-266.

Jordan, G (1982) The Moray Firth Working Party: 'Performance' Without 'Conformance', *European Journal of Political Research*, 10: 117-129 (related to Pressman-Wildavsky Paradox).

Judd, C M (1987) Combining Process and Outcome Evaluation, *New Directions for Program Evaluation*, 35: 23-41.

Judge, E J and Button, K J (1974) Inter-Urban Roads in Great Britain: Perspectives and Prospects, *Transportation Planning and Technology*, 2(2): 185-194.

Kaplan, M (1991) American Neighbourhood Policies: Mixed Results and Uneven Evaluations, Chap. 3, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 28-34.

Keeble, D E (1970) The Movement of Manufacturing Industry - Comments, *Regional Studies*, 4(3): 395-397.

King, J A, Morris, L L and Fitz-Gibbon, C T (1987) How to Assess Program Implementation, 2nd ed., Program Evaluation Kit 5, Sage, Newbury Park, CA.

Kingston, M E (1981) Monitoring in Town and Country Planning, unpublished Master dissertation, Department of Planning and Landscape, University of Manchester.

Knight, R L and Trygg, L L (1977) Evidence of Land Use Impacts of Rapid Transit Systems, *Transportation*, 6: 231-247 (USA experience).

Kuhn, T S (1970) The Structure of Scientific Revolution, 2nd ed., University of Chicago Press, .

Lavery, P (1977) Monitoring and Review of Structure Plans in a Metropolitan County, proceedings of seminar B: Policy Analysis for Urban and Regional Planning, PTRC Summer Annual Meeting, June 1977, pp. 163-169.

Law-Yone, H, Meyer-Bronditz, M and Wilkansky, R (1990) Impact Assessment of Regional Plans, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 128-144.

Lesley, L (1993) Value for Money in Urban Transport Public Expenditure: The Case of Light Rail, *Public Money & Management*, 13(1): 27-33.

Levin, H M (1975) Cost-Effectiveness Analysis in Evaluation Research, in: Guttentag, M and Struening, E (eds.) Handbook of Evaluation Research, Vol. 2, Sage, Beverly Hills, CA, pp. 89-122.

Levin, H M (1981) Applying Well-Established Methods to a Current Dilemma, Chap. 1, in: Smith, N K (ed.) New Techniques for Evaluation, Sage, Beverly Hills, CA, pp. 13-70.

Lichfield, N (1956) Economics of Planned Development, (5th print, 1974. pp. 253-280), The Estate Gazette, London.

Lichfield, N (1960) Cost Benefit Analysis in City Planning, *Journal of the American Institute of Planners*, 26(4): 273-279.

Lichfield, N (1964) Cost Benefit Analysis in Plan Evaluation, *Town Planning Review*, 35: 159-169.

Lichfield, N (1966a) Cost Benefit Analysis in Town Planning: A Case Study of Cambridge, The Cambridgeshire and Isle of Ely County Council.

Lichfield, N (1966b) Cost-Benefit Analysis in Town Planning: A Case Study - Swanley, *Urban Studies*, 3: 215-249.

Lichfield, N (1966c) Cost Benefit Analysis in Urban Development: A Case Study - Swanley, *Papers of the Regional Science Association*, 16: 129-153.

Lichfield, N (1968) Economics in Town Planning: A Basis for Decision Making, *Town Planning Review*, 39(1): 5-20.

Lichfield, N (1969) Cost Benefit Analysis in Urban Expansion: A Case Study - Peterborough, *Regional Studies*, 3: 123-155.

Lichfield, N (1970) Evaluation Methodology of Urban and Regional Plans: A Review, *Regional Studies*, 4: 151-165.

Lichfield, N (1985) From Impact Assessment to Impact Evaluation, in: Faludi, A and Voogd, H (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 51-66.

Lichfield, N (1990) Plan Evaluation Methodology: Comprehending the Conclusions, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 79-97.

Lichfield, N and Chapman, H (1968) Cost Benefit Analysis and Road Proposals for a Shopping Centre - A Case Study: Edgware, *Journal of Transport Economics and Policy*, 2(3): 280-320.

Lichfield, N, Kettle, P and Whitbread, M (1975) *Evaluation in the Planning Process*, Pergamon Press, Oxford.

Likierman, A (1993) Performance Indicators: 20 Early Lessons from Managerial Use, *Public Money & Management*, 13(4): 15-22.

Ling, M (1990) Making Sense of Planning Evaluations: A Critique of Some Methods, *Planning and Development*, 6(1): 14-20.

Lipton, D S (1992) How to Maximize Utilization of Evaluation Research by Policymakers, *The Annals of the American Academy of Political and Social Science*, 521: 175-188.

Local Government Chronicle - LGC (1994) Urban regeneration bids face the axe, 9 September, 1994.

Local Transport Today - LTT (1994) Do roads really bring prosperity? a new report reopens the debate, 7 July 1994, pp. 12-13.

Lombard, P C, Sinha, K C and Brown, D J (1992) Investigation of the Relationship between Highway Infrastructure and Economic Development in Indiana, *Transportation Research Record*, (1359): 76-81.

Long, J (1980a) Editorial, Value for Money in Development Control: A Series of Essays, Institute of Local Government Studies (INLOGOV), University of Birmingham.

Long, J (1980b) Value for Money and the Publication of Information, Value for Money in Development Control: A Series of Essays, Institute of Local Government Studies (INLOGOV), University of Birmingham, pp. 41-49.

Mackay, P R (1976) The Impact of Regional Employment Premium, in: Whiting, A (ed.) *The Economics of Industrial Subsidies*, HMSO, London, pp. 225-242.

Mackett, R L (1980) The Relationship Between Transport and the Viability of Central and Inner Urban Areas, *Journal of Transport Economics and Policy*, 14(3): 267-284.

Majone, G and Wildavsky, A (1979) Implementation as Evolution, Chap. 8, in: Pressman, J L and Wildavsky, A (eds.) *Implementation*, 3rd ed., University of California Press, Berkeley, CA, pp. 163-180.

Massam, B H (1988) Multi-Criteria Decision Making (MCDM) Techniques in Planning, *Progress in Planning*, 30(1): 1-84.

Masser, I (1983) Strategic Land Use Planning: An Evaluation of Procedural Methodology, in: Batty, M and Hutchinson, B (eds.) *Systems Analysis in Urban Policy-Making and Planning*, Plenum Press, New York, pp. 89-106.

Masser, I (1984) Strategic Monitoring for Urban Planning in Developing Countries: Some Guidelines from British and Dutch Experience, TRP 50, Department of Town and Regional Planning, University of Sheffield.

- Mayer, H M** (1979) *Transport and Regional Policy in the United States of America*, Chap. 17, in: **Blonk, W A G** (ed.) Transport and Regional Development: An International Handbook, Saxon House, Farnborough, England, pp. 266-295.
- McAllister, D M** (1980) *Evaluation in Environmental Planning: Assessing Environmental, Social, Economic, and Political Trade-Offs*, MIT Press, Massachusetts.
- McCullough, J D** (1967) *Estimating Systems Costs*, Chap. 5, in: **Goldman, T A** (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Washington, DC, pp. 69-90.
- McFarland, W F and Memmott, J L** (1988) *Ranking Highway Construction Projects: Comparison of Benefit-Cost Analysis with Other Techniques*, *Transportation Research Record*, (1116): 1-9.
- McLoughlin, J B** (1973) *Control and Urban Planning*, (Chapter 9: Requisite Variety and Intelligence Systems, pp. 218-247), Faber, London.
- McLoughlin, J B** (1975) *Monitoring, Research and Intelligence in Structure Planning*, in: **Drake, M, McLoughlin, B J, Thompson, R and Thornley, J** (eds.) Aspects of Structure Planning in Britain, Research Paper 20, CES, London, pp. 157-171.
- Miller, D H** (1985) *Equity and Efficiency Effects of Investment Decisions: Multicriteria Methods for Assessing Distributional Implication*, in: **Faludi, A and Voogd, H** (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 35-50.
- Miller, D H** (1990) *The Organizational and Political Environment of Planning Evaluation*, in: **Shefer, D and Voogd, H** (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 117-128.
- Mills, G** (1977) *Economic Appraisal and Reappraisal of an Inter-Urban Road in Great Britain*, *Journal of Transport Economics and Policy*, 11(1): 3-23.
- Mishan, E J** (1988) *Cost-Benefit Analysis*, 4th ed., Unwin Hyman, London.
- Moore, B and Rhodes, J** (1973a) *Evaluating the Effects of British Regional Economic Policy*, *Economic Journal*, 83: 87-110.
- Moore, B and Rhodes, J** (1973b) *The Economic and Exchequer Implications of Regional Policy*, memorandum submitted to the House of Commons Expenditure Committee, Trade and Industry Sub-Committee - Regional Development Incentives: Minutes of Evidence, HC Paper no. 237, HMSO, London, pp. 448-462.
- Moore, B and Rhodes, J** (1974) *Regional Policy and the Scottish Economy*, *Scottish Journal of Political Economy*, 21: 215-235.
- Moore, B and Rhodes, J** (1975) *The Economic and Exchequer Implications of British Regional Economic Policy*, in: **Vaizey, J** (ed.) Economic Sovereignty and Regional Policy, Gill and Macmillan, Dublin, pp. 80-102.

- Moore, B and Rhodes, J (1976a) Regional Economic Policy and the Movement of Manufacturing Firms to Development Areas, *Economica*, 43: 17-31.
- Moore, B and Rhodes, J (1976b) A Quantitative Analysis of the Effects of Regional Employment Premium and other Regional Policy Instruments, in: Whiting, A (ed.) The Economics of Industrial Subsidies, HMSO, London, pp. 191-219.
- Moore, B and Rhodes, J (1977) Evaluating the Economic Effects of Regional Policy, in: Methods of Measuring the Effects of Regional Policy, OECD, Paris, pp. 11-82.
- Moore, B, Rhodes, J and Tyler, P (1977) The Impact of Regional Policy in the 1970s, *Centre for Environmental Studies Review*, 1(1): 67-77.
- Moore, B, Rhodes, J and Tyler, P (1982) A Methodology with which to Assess the Impact of Regional Economic Policies on Unemployment and Migration, Discussion Paper no. 5, Department of Land Economy, University of Cambridge.
- Moore, B, Rhodes, J and Tyler, P (1986) The Effects of Government Regional Economic Policy (a study commissioned by the DTI), HMSO, London.
- Morgan, K (1995) Reviving the Valleys? Urban Renewal and Governance Structures in Wales, Chap. 13, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 197-214.
- Murdock, S H, Leistritz, F L and Hamm, R R (1986) The State of Socioeconomic Impact Analysis in the United States: An Examination of Existing Evidence, Limitations and Opportunities for Alternative Futures, *Impact Assessment Bulletin*, 4(3-4): 101-132.
- Nachmias, D (1980) The Role of Evaluation in Public Policy, *Policy Studies Journal*, 8(7): 1163-1169.
- Nagel, S S (1986) Efficiency, Effectiveness and Equity in Public Policy Evaluation, *Policy Studies Review*, 16(1): 99-120.
- Nagel, S S (1987) Evaluating Public Policy Evaluation, *Policy Studies Journal*, 16(2): 219-232.
- Nakamura, R T and Smallwood, F (1980) The Politics of Policy Implementation, (only as quoted in Gurnack and Harty, 1987), St Martins Press, New York.
- Nash, C, Pearce, D and Stanley, J (1975) An Evaluation of Cost-Benefit Analysis Criteria, *Scottish Journal of Political Economy*, 22(2): 121-134.
- Nientied, P and Schevz, L (1988) Evaluation of Indirect Impacts of Urban Development Projects, *Third World Planning Review*, 10(4): 389-403.
- Nijkamp, P (1979) *Multidimensional Spatial Data and Decision Making*, John Wiley & Sons, Chichester.

- Nijkamp, P (1985) Evaluation and Information: A Spatial Perspective, in: Faludi, A and Voogd, D (eds.) Evaluation of Complex Policy Problems, DUM, Delft, The Netherlands, pp. 79-87.
- Nijkamp, P and Rietveld, P (1989) Information Systems for Regional Policy Evaluation, *Ekistics*, (338/339): 231-238.
- Nijkamp, P, Rietveld, P and Voogd, H (1990a) Multicriteria Evaluation in Physical Planning, North-Holland, Amsterdam.
- Nijkamp, P, Rietveld, P and Voogd, H (1990b) Multiple Criteria Evaluation: Issues and Perspectives, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 147-158.
- Nijkamp, P and Voogd, H (1980) New Multicriteria Methods for Physical Planning by Means of Multidimensional Scaling Techniques, paper presented at the IFAC-Symposium on Water and Related Land Resource Systems, Cleveland, Ohio, May, 1980.
- Niskanen, W A (1967) Measures of Effectiveness, Chap. 2, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Washington, DC, pp. 17-32.
- Oatley, N and Lambert, C (1995) Evaluating Competitive Urban Policy: The City Challenge Initiative, Chap. 10, in: Hambleton, R and Thomas, H (eds.) Urban Policy Evaluation: Challenge and Change, Paul Chapman, London, pp. 141-157.
- O'Leary, T J (1979) The Cost/Benefit Analysis: It Can Be a Misleading Tool for Transportation Planners, *Transportation Planning and Technology*, 5(4): 189-193.
- Oosterhaven, J (1983) Evaluating Land Reclamation Plans for Northern Friesland: An Interregional Cost-Benefit Analysis and Input-Output Analysis, *Papers of the Regional Science Association*, 52: 125-137.
- Openshaw, S and Whitehead, P (1985) A Monte Carlo Simulation Approach to Solving Multicriteria Optimisation Problems Related to Planmaking, Evaluation and Monitoring in Local Planning, *Environment and Planning B*, 12: 321-334.
- Owen, D W, Gillespie, A E and Coombes, M G (1984) 'Job Shortfall' in British Local Labour Market Areas: A Classification of Labour and Demand Trends, *Regional Studies*, 18(6): 469-488.
- Owen, D W and Green, A E (1989) Labour Market Accounts for Travel-To-Work Areas, 1981-1984, *Regional Studies*, 23(1): 69-72.
- Palmer, A J (1993) Performance Measurement in Local Government, *Public Money & Management*, 13(4): 31-36.
- Palumbo, D J (1987a) Politics and Evaluation, Chap. 1, in: Palumbo, D J (ed.) The Politics of Program Evaluation, Sage, Newbury Park, CA, pp. 12-46.

Palumbo, D J (1987b) (ed.) *The Politics of Program Evaluation*, Sage Yearbooks in Politics and Public Policy 15, Sage, Newbury Park, CA.

Palumbo, D J and Nachmias, D (1983) Preconditions for Successful Evaluation: Is There an Ideal Paradigm?, *Policy Sciences*, 16(1): 67-79.

Palumbo, D J and Oliverio, A (1989) Implementation Theory and the Theory-Driven Approach to Validity, *Evaluation and Program Planning*, 12: 337-344.

Paris, J D (1970) Comments on "Further Thoughts on the Shift and Share Approach" by F.J.B. Stilwell, *Regional Studies*, 4(4): 491.

Parkinson, M (1981) *The Effect of Road Investment on Economic Development in the UK*, Government Economic Services Working Papers 43, DoT, London.

Patton, M Q (1987) Evaluation's Political Inherency: Practical Implications for Design and Use, Chap. 4, in: Palumbo, D J (ed.) *The Politics of Program Evaluation*, Sage, Newbury Park, CA, pp. 100-145.

Patton, M Q (1990) *Qualitative Evaluation and Research Methods*, 2nd ed., (1st edition entitled "Qualitative Evaluation Methods", 1980), Sage, Newbury Park, CA.

Peaker, A (1976) New Primary Roads and Sub-Regional Economic Growth: Further Results - A Comment on J.S. Dodgson's Paper, *Regional Studies*, 10(1): 11-13.

Pearce, D W (1983) *Cost-Benefit Analysis*, Macmillan, London.

Pearce, D W and Nash, C A (1981) *The Social Appraisal of Projects: A Text in Cost-Benefit Analysis*, Macmillan, London.

Pearman, A D (1979) Approaches to Multiple-Objective Decision-Making with Ranked Criteria, in: Cullen, I G (ed.) *Analysis and Decision in Regional Policy*, Pion, London, pp. 136-152.

Pearman, A D (1984) Review of *Multicriteria Evaluation for Urban and Regional Planning* by H. Voogd, *Environment and Planning A*, 16: 1118-1119.

Pearman, A D (1985) Uncertainty in Planning: Characterisation, Evaluation and Feedback, *Environment and Planning B*, 12: 313-320.

Perera, M H (1990) Framework for Classifying and Evaluating Economic Impacts Caused by a Transportation Improvement, *Transportation Research Record*, (1274): 41-52.

Perry, H A and Chamberlain, K J (1977) Hertfordshire: Monitoring and the On-Going Review Process, proceedings of seminar B: Policy Analysis for Urban and Regional Planning, PRTC Summer Annual Meeting, June 1977, pp. 137-149.

Persky, J, Felsenstein, D and Wiewel, W (1994) A Cost-Benefit Framework for Assessing the Impact of Urban Economic Development Initiatives, paper presented at the 34th European Congress of the Regional Science Association, 23-26 August, Groningen, The Netherlands.

Planner (1994) Gummer secures further round of bidding for renewal budget, no. 1097, 2 December, p. 32.

Posavac, E J and Carey, R G (1989) Program Evaluation: Methods and Case Studies, 3rd ed., Prentice Hall, New Jersey.

Pressman, J L and Wildavsky, A (1984) Implementation, 3rd ed., University of California Press, Berkeley.

Priemus, H (1991) The Upgrading of Large-Scale Housing Estates in The Netherlands, Chap. 8, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 110-129.

Quade, E S (1967) Introduction and Overview, Chap. 1, in: Goldman, T A (ed.) Cost-Effectiveness Analysis: New Directions for Decision-Making, Washington Operations Research Council, Washington, DC, pp. 1-16.

Rees, R D and Miall, R H C (1979) The Effects of Regional Policy on Manufacturing Investment and Capital Stock within the UK, Government Economic Service Working Papers 26, Department of Industry, London.

Reid, M F (1990) Linking Implementation and Evaluation Through Program Monitoring, *Journal of Urban Affairs*, 12(3): 253-266.

Reiner, T A (1990) Choices and Choice Theory Revisited, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 65-78.

Rhodes, J (1986) Regional Economics and Policy - Armstrong, H. and Taylor, J. (Book Review), *Regional Studies*, 20(4): 379.

Richardson, H W (1978) The State of Regional Economics: A Survey Article, *International Regional Science Review*, 3(1): 1-48.

Richardson, H W (1979) Regional Economics, (pp. 202-206, 253-256), University of Illinois Press, Urbana.

Richardson, J J (1982) Programme Evaluation in Britain and Sweden, *Parliamentary Affairs*, 35(2): 160-180.

Richardson, J J (1984) Preface, in: Richardson, J J (ed.) Value for Money and Effectiveness Auditing in Public Sector: A Symposium, Strathclyde Papers on Government and Politics 30, Department of Politics, University of Strathclyde, Glasgow, pp. i-v.

Riera, B and Jackson, M R (1971) The Design of a Monitoring and Advisory System for Sub-Regional Planning, Notts./Derbys. Sub-Regional Management Committee.

Rietveld, P (1994) Spatial Economic Impacts of Transport Infrastructure Supply, *Transportation Research A*, 28(4): 329-341.

Robson, B T, Bradford, M G, Deas, I, Hall, E, Harrison, E, Parkinson, M, Evans, R, Garside, P and Harding, A (1994) Assessing the Impact of Urban Policy (commissioned by the DoE - Inner Cities Research Programme), HMSO, London.

Rose, E A (1979) Monitoring and Review in the Planning Process: Some Practical Problems, in: Soen, D (ed.) New Trends in Urban Planning, Pergamon Press, Oxford, pp. 22-37.

Rosenbaum, N (1981) Statutory Structure and Policy Implementation: The Case of Wetlands Regulation, Chap. 3, in: Mazmanian, D A and Sabatier, P A (eds.) Effective Policy Implementation, Lexington Books, Massachusetts, pp. 63-85.

Rossi, P H and Freeman, H E (1989) Evaluation: A Systematic Approach, 4th ed., Sage, Newbury Park, CA.

Rossi, P H and Freeman, H E (1993) Evaluation: A Systematic Approach, 5th ed., Sage, Newbury Park, CA.

Rossi, P H, Freeman, H E and Wright, S R (1979) Evaluation: A Systematic Approach, 1st ed., Sage, Beverly Hills.

Rothenberg, J (1975) Cost-Benefit Analysis: A Methodological Exposition, in: Guttentag, M and Struening, E (eds.) Handbook of Evaluation Research, Vol. 2, Sage, Beverly Hills, CA, pp. 55-88.

Royal Commission on Environmental Pollution (1994) Eighteenth Report: Transport and the Environment (Cm 2674), HMSO, London.

Russell, H (1994a) Liverpool City Challenge: Report of the Independent Evaluation Team - Year 2, European Institute for Urban Affairs, Liverpool John Moores University, Liverpool.

Russell, H (1994b) City Challenge Interim Evaluation - Phase I: Progress Report to the Steering Committee, European Institute for Urban Affairs, Liverpool John Moores University, Liverpool.

Sabatier, P A (1986) Top-Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis, *Journal of Public Policy*, 6(1): 21-48.

Sabatier, P A and Klosterman, B J (1981) A Comparative Analysis of Policy Implementation under Different Statutory Regimes: The San Francisco Bay Conservation and Development Commission, 1965-1972, Chap. 7, in: Mazmanian, D A and Sabatier, P A (eds.) Effective Policy Implementation, Lexington Books, Massachusetts, pp. 169-206.

Sabatier, P A and Mazmanian, D A (1981) The Implementation of Public Policy: A Framework of Analysis, Chap. 1, in: Mazmanian, D A and Sabatier, P A (eds.) Effective Policy Implementation, Lexington Books, Massachusetts, pp. 2-35.

Sabatier, P A and Mazmanian, D A (1983) Policy Implementation, Chap. 7, in: Nagel, S S (ed.) Encyclopedia of Policy Studies, Marcel Dekker Inc., New York, pp. 143-169.

Samuelson, G D (1968) Motorways and Industry: The West Midland Conurbation, British Road Federation, London.

Sassone, P G and Schaffer, W A (1978) Cost-Benefit Analysis: A Handbook, Academic Press, New York.

Scheurwater, J and Masser, I (1981) Monitoring Spatial Planning in The Netherlands: An Outline of an Information Analysis System, in: Voogd, H (ed.) Strategic Planning in a Dynamic Society, DUM, Delft, The Netherlands, pp. 193-204.

Schmoll, F (1991) Post-War Outer Housing Estates in the Federal Republic of Germany: Problems, Strategies, Prospects, Chap. 7, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 91-109.

Schofield, J A (1976) Economic Efficiency and Regional Policy, *Urban Studies*, 13(2): 181-192.

Schofield, J A (1979) Macro Evaluations of the Impact of Regional Policy in Britain: A Review of Recent Research, *Urban Studies*, 16: 251-271.

Schofield, J A (1987) Cost-Benefit Analysis in Urban and Regional Planning, Allen & Unwin, London.

Seigneur, C and Roth, P H (1987) A Methodology for the Cost/Health-Benefit Analysis of Photochemical Smog Control, *Socio-Economic Planning Sciences*, 21(5): 311-319.

Semke, J I and Nurius, P S (1991) Information Structure, Information Technology, and the Human Services Organizational Environment, *Social Work*, 36(4): 353-358.

Seskin, S N (1990) Comprehensive Framework for Highway Economic Impact Assessment: Methods and Results, *Transportation Research Record*, (1274): 24-34.

Sharp, C (1979) The Environmental Impact of Transport and the Public Interest, *Journal of Transport Economics and Policy*, 13(1): 88-101.

Sharp, C H and Gibson, M (1979) Transport and Regional Policy in Great Britain, Chap. 11, in: Blonk, W A G (ed.) Transport and Regional Development: An International Handbook, Saxon House, Farnborough, England, pp. 171-189.

Shefer, D and Kaess, L (1990) Evaluation Methods in Urban and Regional Planning: Theory and Practice, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 99-115.

Shefer, D and Tsubari, V (1990) The Evolution of Evaluation Methodology in the Works of Morris Hill, in: Shefer, D and Voogd, H (eds.) Evaluation Methods for Urban and Regional Plans, Pion, London, pp. 31-42.

Simmons, M (1985) Motorways and Development: The Case of the M25 London Orbital, *The Planner*, 71(2): 55-58.

Spiro, S E (1991a) Physical and Social Changes Achieved by Israel's Project Renewal, Chap. 11, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 170-178.

Spiro, S E (1991b) The Evaluation of Neighbourhood Rehabilitation Programmes: Lessons from Israel's Project Renewal, Chap. 14, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 197-208.

Standing Advisory Committee on Trunk Road Assessment - SACTRA (1980) Trunk Road Proposals - A Comprehensive Framework for Appraisal, HMSO, London.

Standing Advisory Committee on Trunk Road Assessment - SACTRA (1992) Assessing the Environmental Impact of Road Schemes, HMSO, London.

Standing Advisory Committee on Trunk Road Assessment - SACTRA (1994) Trunk Roads and the Generation of Traffic, HMSO, London.

Starkie, D (1982) The Motorway Age: Road and Traffic Policies in Post-War Britain, Pergamon Press, Oxford.

Steeley, G C (1976) Strategic Policy Monitoring: 'Meta-Information' and Decision-Making, *BURISA*, 25: 11.

Stein, M M (1975) Regional Impacts of National Transport Systems on Population and Travel, *Journal of Transport Economics and Policy*, 9(3): 255-267.

Stephanedes, Y J (1990) Distributional Effects of State Highway Investment on Local and Regional Development, *Transportation Research Record*, (1274): 156-164.

Stephanedes, Y J and Eagle, D M (1986) Time-Series Analysis of Interactions between Transportation and Manufacturing and Retail Employment, *Transportation Research Record*, (1074): 16-24.

Stephoe, R and Thornton, C (1986) Differential Influence of an Interstate Highway on the Growth and Development of Low-Income Minority Communities, *Transportation Research Record*, (1074): 60-68.

Stilwell, F J B (1970) Further Thoughts on The Shift and Share Approach, *Regional Studies*, 4(4): 451-458.

- Sugden, R and Williams, A (1978)** The Principles of Practical Cost-Benefit Analysis, Oxford University Press, Oxford.
- Tang, J C S and Phataraloaha, Y (1987)** Project Appraisal for Developing Countries: A Systematic Approach, *Socio-Economic Planning Sciences*, 21(6): 377-387.
- Thompson, G L, Weller, B and Terrie, E W (1993)** New Perspectives on Highway Investment and Economic Growth, *Transportation Research Record*, (1395): 81-87.
- Thompson, M S (1980)** Benefit-Cost Analysis for Program Evaluation, Sage, Beverly Hills, CA.
- Thrasher, M and Dunkerley, D (1982)** A Social Exchange Approach to Implementation Analysis, *Social Science Information*, 21(3): 349-382.
- Tricart, J P (1991a)** Neighbourhood Social Development Policy in France, Chap. 6, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 84-90.
- Tricart, J P (1991b)** Evaluation of Neighbourhood Social Development Policy, Chap. 13, in: Alterman, R and Cars, G (eds.) Neighbourhood Regeneration: An International Evaluation, Mansell, London, pp. 189-196.
- Troman, J W (1984)** Efficiency Auditing: The Scottish Experience, in: Richardson, J J (ed.) Value for Money and Effectiveness Auditing in Public Sector: A Symposium, Strathclyde Papers on Government and Politics 30, Department of Politics, University of Strathclyde, Glasgow, pp. 20-35.
- Turok, I (1991)** Policy Evaluation as Science: A Critical Assessment, *Applied Economics*, 23(9): 1543-1550.
- Tyler, P (1980)** The Impact of Regional Policy on a Prosperous Region: The Experience of the West Midlands, *Oxford Economic Papers*, 32(1): 151-162.
- Tyler, P and Rhodes, J (1989)** A Model with which to Forecast Employment and Population Change at the Regional and Sub-Regional Level, Chap. 8, in: Congdon, P and Batey, P (eds.) Advances in Regional Demography: Information, Forecasts, Models, Belhaven Press, London, pp. 124-149.
- Van Meter, D S and Van Horn, C E (1975)** The Policy Implementation Process: A Conceptual Framework, *Administration and Society*, 6(4): 445-488.
- Victor Hausner & Associates (1993)** The Programme for the Valleys - An Evaluation (Volume 1), Victor Hausner & Associates, London.
- Voogd, H (1983a)** Multicriteria Evaluation for Urban and Regional Planning, Pion, London.
- Voogd, H (1983b)** Qualitative Calculus and Regional Impact Analysis, Planologisch Memorandum 1983-7, Department of Civil Engineering, Delft University of Technology, Delft.

- Voogd, H (1985) Prescriptive Analysis in Planning, *Environment and Planning B*, 12: 303-312.
- Voogd, H (1988) Multicriteria Evaluation: Measures, Manipulation and Meaning - A Reply, *Environment and Planning B*, 15: 65-72.
- Walker, W, Abrahamse, A and Kleijn, H (1993) A Demonstration Policy Evaluation of the Dutch Second Transport Structure Plan (SVV), RAND, Santa Monica, CA.
- Watterson, W T (1985) Estimating Economic and Development Impacts of Transit Investments, *Transportation Research Record*, (1046): 1-9.
- Wedgewood-Oppenheim, F, Hart, D and Cobley, B (1975) An Explanatory Study in Strategic Monitoring, *Progress in Planning*, 5(1): 1-58.
- Weiss, C H (1987) Where Politics and Evaluation Research Meet, Chap. 2, in: Palumbo, D J (ed.) The Politics of Program Evaluation, Sage, Newbury Park, CA, pp. 47-70.
- Welsh Office (1988) The Valleys: A Programme for the People, Welsh Office, Cardiff.
- Williams, A (1973) Cost-Benefit Analysis: Bastard Science? And/Or Insidious Poison in the Body Politick, Chap. 3, in: Wolfe, J N (ed.) Cost-Benefit and Cost-Effectiveness: Studies and Analysis, George Allen & Unwin, London, pp. 30-60.
- Wilson, F R, Graham, G M and Aboul-Ela, M (1985) Highway Investment as a Regional Development Policy Tool, *Transportation Research Record*, (1046): 10-14.
- Wilson, F R, Stevens, A M and Holyoke, T R (1982) Impact of Transportation on Regional Development (Abridgment), *Transportation Research Record*, (851): 13-16.
- Wittrock, B, Lindstrom, S and Zitterberg, K (1982) Implementation Beyond Hierarchy: Swedish Energy Research Policy, *European Journal of Political Research*, 10: 131-143.
- Won, J (1990) Multicriteria Evaluation Approaches to Urban Transportation Projects, *Urban Studies*, 27(1): 119-138.
- Worrall, L (1990) Information Systems for Urban and Regional Planning in the United Kingdom: A Review, *Environment and Planning B*, 17(4): 451-462.
- Wren, C (1988) Closure Rates among Assisted and Non-Assisted Establishments, *Regional Studies*, 22(2): 107-119.
- Wren, C (1989) The Revised Regional Development Grant Scheme: A Case Study in Cleveland County of a Marginal Employment Subsidy, *Regional Studies*, 23(2): 127-137.
- Wyatt, R (1989) Intelligent Planning: Meaningful Methods for Sensitive Situations, (Chapter 6: Evaluate), Unwin Hyman, London.

Yewlett, C J L (1986) Review of Multicriteria Evaluation for Urban and Regional Planning by H. Voogd, *The Planner*, 72(6): 48.

Zografos, K G and Stephanedes, Y J (1992) Impact of State Highway Investment on Employment Along Major Highway Corridors, *Transportation Research Record*, (1359): 151-155.

APPENDICES

A major limitation of CBA is its deterministic approach. It assumes that all variables are identified precisely not allowing for dynamic uncertainties such as wage settlements or inflation rates. Another limitation is that no evaluation of financial risk or range of possible benefits is made. Such limitations can be overcome by evaluating the net revenue as a stochastic variable. Using Monte Carlo (M/C) Analysis, a probabilistic distribution for net revenue can be generated, which can then be used to analyze risk and return by probability statements. Such an approach involves five steps (see Fig A.1):

1. Identify the significant stochastic variables employed in the cost-benefit analysis (the M/C variables). These are the variables which have the potential of significantly altering the account balances.
2. Specify the probability distribution of those M/C variables. This can be done on the basis of either historical data or subjective judgement.
3. Re-formulate each cost-benefit equation by replacing each M/C variable by its appropriate probability distribution and all other variables by constants equal to their exact values.
4. For each trial in the analysis, each M/C variable will be assigned a randomly selected value from its encoded probability distribution. These values are then used in the re-formulated CB equations to calculate one net revenue figure.
5. By undertaking a large number of trials, a large number of net revenue figures can be generated. "The expected value, variance, and relative frequency distribution for this data will describe the true stochastic nature of the net revenue random variable" (p. 190).

⁽³⁰⁾ see O'Leary, T J (1979).

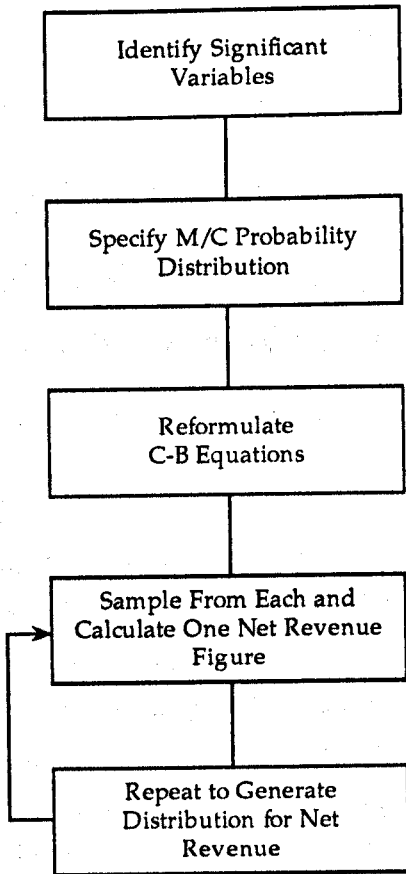


Fig A.1: Monte Carlo Extension of CBA

Appendix B: Output Measures - The SRB

The 'initial' bidding guidance of the SRB (DoE, 1994a) provided a list of output measures grouped under relevant objectives. Partners were required to list the foreseen impacts of their projects in respect to these measures, in addition to any other possible outputs that related to their particular bid. The idea was apparently to ensure that outputs were clearly defined and could be monitored. The 'Monitoring and Periodic Review' guidance note expanded on those measures. Two more measures were included; the whole set, however, was no longer explicitly grouped under relevant objectives. Nonetheless, the numbering system was essentially the same which enables a classification of these measures by objective.

Objective	Output Measure
Enhance employment prospects, education, skills of local people	1A No. of jobs created/safeguarded
	1B No. of pupils benefiting from projects designed to enhance/improve attainment
	1C No. of people trained obtaining qualification
	1D No. of residents of target areas accessing employment through training, advice or specifically targeted assistance.
	1E No. of training weeks
	1F No. of people trained obtaining jobs (of which number who were formerly unemployed)
	1G No. of unemployed people entering self employment
	1H No. of ethnic minority pupils improving ESOL attainment
	1I No. from disadvantaged groups being targeted who obtain a job
	1J No. of young people benefiting from projects to promote personal and social development
	1K No. of employers involved in collaborative projects with education to improve student performance
	1L No. of teachers on placement into business ¹

Objective	Output Measure
Encourage sustainable economic growth and wealth creation	<p>2A No. of new business start ups</p> <p>2B m² new/improved business/commercial floorspace</p> <p>2C Survival rate of new businesses</p> <p>*2D No. of businesses advised</p>
Improve housing	<p>3A No. of dwellings completed/improved</p> <p>3B No. of dwellings included in newly-formed tenant management schemes</p> <p>3C <i>Increase in proportion of total rent</i></p> <p>3D <i>Increase in proportion of responsive repairs</i></p> <p>3E <i>No. of dwellings transferred out of public sector* to owner occupation/rent</i></p>
Promote initiatives of benefit to ethnic minorities	<p>4A <i>No. of ethnic minority business start ups</i></p> <p>4B <i>No. of unemployed ethnic minority people obtaining jobs</i></p> <p>4C <i>No. of ethnic minority pupils enabled to achieve higher levels of education</i></p> <p>4D <i>No. of ethnic minority people entering vocational training</i></p> <p>4E <i>No. of ethnic minority people enabled access to information or advice</i></p>
Tackle crime and promote community safety	<p>5A No. of elderly, women or all people who benefit from community safety initiatives</p> <p>5B No. of dwellings and commercial buildings where security is upgraded</p> <p>5C No. of community safety initiatives</p> <p>5D No. of youth crime prevention initiatives</p>
Protect and improve the local environment	<p>6A Ha. land improved/reclaimed for open space</p> <p>6B Ha. land improved/reclaimed/serviced for development</p> <p>6B No. of buildings improved and brought back into use</p> <p>6D Km road built/improved</p> <p>6E No. of traffic calming schemes</p> <p>6F Waste management/recycling schemes</p>
Enhance the quality of life of local people	<p>7A No. of local people given access to new community facilities and no. of new facilities</p> <p>7B No. of community facilities improved/established and use of those facilities</p> <p>*10A No. of child-care places provided</p>
Harness the talents and resources of the voluntary sector and the community	<p>8A No. of voluntary organizations supported</p> <p>8B <i>No. and % of dwellings included in tenant management organisations</i></p> <p>8C No. of individuals employed in voluntary work</p> <p>8D No. of local employers with employee volunteering schemes</p> <p>8E No. of community enterprise start ups</p>
Lever in private sector resources	<p>9A Value of other funding (public and private) attracted</p>

Notes to the above table:

¹ was initially % of teachers

* Measures added in the Guidance Note, Feb. 1995.

Measures in Italics are no longer applicable.

Source: modified from Annex A, Guidance Note No. 2 (DoE, 1995b), pp. 13-14.

Appendix C: Outcome Measures - The SRB

Objectives	Output Measures	Outcome Measures
Enhance employment prospects, education, skills of local people	1A to 1L	<ul style="list-style-type: none"> - Reduction in unemployment rate - Increase in skill levels - Increase in attainment in learning (eg GCSEs, NVQs) - Increased job centre placings - Level and growth of employment - Schools staying on rates - Levels of truancy - Employer attitude to students' preparation for the world of work
Encourage sustainable economic growth and wealth creation	2A to 2C	<ul style="list-style-type: none"> - Increase in profit/turnover/assets of firms - Reduction in property vacancy - Increase in value added per employee - Increase in GDP per head - Increase in property rentals/prices
Improve housing	3A to 3E	<ul style="list-style-type: none"> - Reduction in voids - Reduction in unfit dwellings - Tenant satisfaction as measured by survey - Reduction in no./% tenants in arrears - Reduced levels of overcrowding - Reduction in no. of houses lacking amenities
Promote initiatives of benefit to ethnic minorities	4A to 4E	<ul style="list-style-type: none"> - Relative reduction in ethnic minority unemployment rates - Increase in attainment in learning (eg GCSEs, NVQs) - Increased growth and survival of ethnic minority businesses
Tackle crime and promote community safety	5A to 5D	<ul style="list-style-type: none"> - Increase in proportion of residents saying that their fear of crime is reduced - Reduction in reported crime (by type)/1,000 households - Reduction in house contents/car area insurance premium
Protect and improve the local environment	6A to 6F	<ul style="list-style-type: none"> - Reduction in vacant land in target area - Reduction in derelict land in target area - Land values - Surveys of business and residents - Reduction in complaints about/incidence in litter, noise, graffiti, vandalism

Objectives	Output Measures	Outcome Measures
Enhance the quality of life of local people	7A & 7B	<ul style="list-style-type: none"> - Increased attendance/usage of facilities - Survey evidence on local peoples' views on facilities - Reduction perinatal/infant mortality - Increase in proportion of people saying that they perceive an improvement in their quality of life - Standard mortality - Reduced LA environmental health officer case loads - Attendance/usage of sports and recreational facilities
Harness the talents and resources of the voluntary sector and the community	8A to 8E	<ul style="list-style-type: none"> - Increase in voluntary sector and community involvement in local decision-making - Effectiveness of community organisations - Tenant satisfaction surveys
Lever in private sector resources	9A	<ul style="list-style-type: none"> - Increase in net additional investment to target area - Diversification of housing and local economy (by employment)

Source: modified from DoE (1994a), Annex C, pp. 27-30.

1. Guidance Note No. 1: Partnership Delivery Plan (DoE, 1995a)

This note provides guidance on how to translate approved bids into practical Delivery Plans (which form the basis for the formal agreement between the government and partnerships). Within the context of government urban regeneration initiatives, this idea is indeed unprecedented. This is the first incident where the central government takes an interest in, and issues guidance on, transferring approved bids (strategies) into operational plans.

A Delivery Plan "must set out, for the scheme as a whole, the partnership arrangements; strategic objectives; descriptions of existing and planned local conditions; the quantifiable and non-quantifiable results partnerships propose to deliver; and the funding required" (DoE, 1995a: 1). A Plan contains:

1. **Introductory Statement:** A description of the purpose of the scheme and its target group/area and the partnership and how it will operate;
2. **Schedule A - Strategic Objectives:** Statement of strategic objectives, description of baseline local conditions and indicators of progress (see Table D.1). "The purpose of requiring a partnership to describe progress against baseline is to go beyond counting outputs and to identify intended changes in local circumstances brought about by the scheme (ie. outcomes)" (DoE, 1995a: 4).
3. **Schedule B - Funding profile:** Forecasts of expenditure showing SRB and other elements separately (see Table D.2).
4. **Schedule C - Milestones:** These are "intermediate totals for quantifiable outputs and the key events or stages towards the achievement of non-quantifiable outputs (including the setting up of the partnership and its delivery mechanisms)" (DoE, 1995a: 5, see Table D.3). This schedule will be used to underpin funding requirements and assess the progress of the scheme.
5. **Schedule C - Quantifiable Outputs:** A summary of all the quantifiable outputs the partnership intends to deliver within each claim period in the first year and the totals for the life time of the scheme (see Table D.4).

2. Guidance Note No. 2: Monitoring and Periodic Review (DoE, 1995b)

This note provides guidance on the monitoring and periodic review of SRB-assisted activities. It described the responsibilities of both the GOs and those delivering the approved bids. Arrangements are also set for assessing progress against forecast targets through Key Indicators of performance. The primary intention of these arrangements was "to ensure that approved schemes are making satisfactory progress towards the targets set out in Delivery Plans" (DoE, 1995b: 2).

Monitoring was seen as a necessary activity to ensure that:

- * the intended outputs and other benefits are being delivered to time and in a cost-effective way;
- * payments can be related to performance;
- * managers (within GORs, Partnerships and projects) can respond to the performance of funded activities;
- * value-for-money assessments of programme activities can be made;
- * the basis for evaluation is consistent across projects, schemes and regions.

(DoE, 1995b: 2)

The guidance, on several occasions, emphasised the need to collect and report accurate information on regular basis and at the various levels of the process (projects, partnerships and region-wide). It has also further detailed both the initial list of output measures and their definition, with an emphasis on 'ethnic monitoring' (wherever applicable, benefits accruing to ethnic minorities are to be spelled out separately) (DoE, 1995b: 12).

Table D.1: An Illustration of Schedule A - Strategic Objectives

Strategic Objectives ^a	Baseline Position	Position at end of Lifetime of bid ^b
Enhance the employment prospects, education and skills of local people SO1	<ul style="list-style-type: none"> • School leavers into jobs with structured training - 35% • School leavers into further education - 17% • Rate of adult registered unemployment - 15% • Proportion of unemployed adults registered for six months or more - 55% • • • 	<ul style="list-style-type: none"> • Increase the proportion of leavers entering employment with structured training or further education to 45% and 25% respectively • Reduction in the rate of adult registered unemployed and long term unemployment to 8% and 48% respectively • Progress made in reducing the rate of unemployed ethnic minorities •
Encourage sustainable economic growth and wealth creation SO2	<ul style="list-style-type: none"> • Below national average business starts and survival rates • Low participation among ethnic minorities in business start up activity • 	<ul style="list-style-type: none"> • Increased number of new business start-ups and improvement in survival rates • Increased proportion of ethnic minority business start-ups •
Improve housing SO3	700 unfit dwellings within the area	Reduce this number by 200 through targeted refurbishment

^a From the list in Annex C, Bidding Guidance (DoE, 1994a) or, Annex A, the revised Bidding Guidance (DoE, 1995e); see Appendix 7.1 above.

^b Improvements over baseline conditions will be identified from local evaluation

Source: modified from Annex 1, DoE, 1995a.

Table D.2: An Illustration of Schedule B - Funding Profile

Breakdown of Expenditure	£000		£000		£000		£000		£000		
	1st period April/June	PLAN	2nd Period July/Sept.	PLAN	3rd Period Oct/Dec	PLAN	4th Period Jan/March	PLAN	Total Year1	PLAN	Total All Years
<u>SRB Contribution</u>											
Capital	70	140	213	320	743	6,816					
Revenue	16	61	51	16	144	1,704					
Sub-Total SRB	86	201	264	336	887	8,520					
Contribution towards admin. and mang. costs											
SRB 1st year (included in figures above)											
Private Sector (included in figures below)											
<u>Private Sector Contribution (cash or in kind)</u>											
Developers, local businesses, ...											
Sub-Total Private Leverage											
<u>Other Public Sector Contributors</u>											
Local Authority											
TEC (public)											
European ERDF											
ESF											
Other Gov. Depts (from outside SRB)											
DoT											
FE											
Other expenditure from the wider SRB											
eg City Challenge											
Sub-Total Other Public Sector											
GRAND TOTAL											

Source: modified from Annex 1, Doe, 1995a.

Table D.3: An Illustration of Schedule C - Milestones

Project Title & Link with Strategic Objective	1st Period: April - June	2nd Period: July - Sept.	3rd Period: Oct. - Nov.	4th Period: Jan - March	Other Milestones
1 Management & Admin. SO1	<ul style="list-style-type: none"> - Formation of Partnership completed APRIL 1995 - Steering Groups members appointed Feb 1995 - Appoint manager 1 job MAR 1995 	<ul style="list-style-type: none"> - First of six-weekly meeting with Voluntary Sector Aug 1995 	<ul style="list-style-type: none"> - Community Forum established SEPT 1995 	<ul style="list-style-type: none"> - Review first year and prepare 96/97 Delivery Plan 	<ul style="list-style-type: none"> - SRB funding ends 2002/03
2 Family Literacy programme of 15 projects SO1, SO4	<ul style="list-style-type: none"> - Info & bid guidance to schools by end APRIL 1995 - Bids submitted JUNE 1995 	<ul style="list-style-type: none"> - Bids approved JUL 1995 - Appoint 18 Staff (1A1) SEPT 1995 	<ul style="list-style-type: none"> - Adaptation complete DEC 1995 	<ul style="list-style-type: none"> - 300 pupils benefiting (1B) FEB 1996 	<ul style="list-style-type: none"> - End date of last project 97/98
3 Language Support SO1	<ul style="list-style-type: none"> - Schools based contracts agreed and signed JUN 1995 	<ul style="list-style-type: none"> - Private sector support confirmed JUL 1995 - Appoint 40 staff (A1i) SEPT 1995 	<ul style="list-style-type: none"> - 700 pupils begin to benefit (1B) OCT 1995 		<ul style="list-style-type: none"> - End date 2002/03
4 Extension Learning/Care SO1		<ul style="list-style-type: none"> - Appoint 4 people (A1i) AUG 1995 	<ul style="list-style-type: none"> - 10 unemployed into work (1Fi & 1Fii) DEC 1995 	<ul style="list-style-type: none"> - 10 unemployed into work (1Fi & 1Fii) MAR 1996 	<ul style="list-style-type: none"> - End date 99/2000
5					
6					
7					
8					

Source: modified from Annex 1, DoE, 1995a.

Table D.4: AN Illustration of Schedule D - Quantifiable Outputs

Code	Outputs Totals (ethnic minority)	1st Period	2nd Period	3rd Period	4th Period	Total 95/96	Total all Years						
A1	Number of Jobs (1) created (ii) safeguarded (iii) construction jobs	15	(0)	81	(6)	16	(8)	15	(8)	127	(22)	392	(2000)
1B	Number of pupils benefiting from projects												
1C	Number of people trained obtaining qualifications												
1D	Number of residents accessing employment through training												
1E	Number of training weeks												
1G	Number of unemployed entering self-employment												
2A	Number of new business start-ups												
2B	(i) area of new business/commercial floorspace (ii) area of improved business/commercial floorspace												
3B	Number of dwellings included in tenant management schemes												
6A	Hectares of land improved/reclaimed/serviced for development												
	Other partnership outputs												
		15	(12)	43	(28)	4	(4)	95	(57)	157	(101)	800	(450)
						10,000				10,000		30,000	

Source: modified from Annex 1, DoE, 1995a.

3. Guidance Note No. 3: Project Appraisal and Approval (DoE, 1995c)

This guidance sets out the requirements for the appraisal and approval of individual projects before they can go ahead. The division of responsibilities between the GOs and partnerships is linked to the nature and funding requirements of the project.

4. Guidance Note No. 4: Financial Guidance (DoE, 1995d)

This note provides guidance for the financial arrangements for SRB grant from 1 April 1995. It sets out the general principles of financial assistance to partnerships and includes guidance on EU state aid limitations and assistance to the private sector.

Appendix E: Core Impact Indicators - City Challenge

ISSUE	IMPACT INDICATOR	COMMENT
Employment/Income	<ol style="list-style-type: none"> 1. Number of Unemployed* (total/male/female/long-term (more than 12 months)/youth (under 21)/ethnic minority) 2. Number of residents in the CC area obtaining permanent employment (full- and part-time) (inside or outside the CC area) as a consequence of the CC actions. 3. Heads of households in receipt of income support* 	<p>Heads of household of working age only. Where this information is not available, proxy measures may prove suitable, for example, number of children registered for free school meals, number of households receiving housing benefit.</p>
Education	Number of pupils entered for GCSEs achieving grades A to C in 5 or more subjects <u>AND</u> Number of pupils leaving school with no GCSEs*	From 1994, GCSEs in National Curriculum subjects will be marked in terms of 10 levels rather than the current A-G grades.
Housing	<ol style="list-style-type: none"> 5. Number of households in owner occupation* 6. Number of unfit dwellings* 	<p>Include households in shared ownership</p> <p>Assess against the statutory housings standard (1985 Housing Act as amended)</p>
Environment/Economy	<ol style="list-style-type: none"> 7. Hectares of: (i) derelict land, (ii) vacant land 8. Floorspace (sq.m) in: (i) derelict buildings, (ii) vacant buildings 	
Crime	<ol style="list-style-type: none"> 9. Reported crime/1000 households: (i) residential burglary and (ii) theft of or from cars* 10. No standard core impact indicator is proposed but people's attitudes towards a range of issues should be tracked over five years. 	
Quality of Life/perception of CC area		

* Partnerships should: (i) provide the number of people/households/dwellings in each category; (ii) provide the total relevant population/households/dwellings, and (iii) show (i) as a proportion/percentage of (ii).

Source: modified from Annex B, DoE-ICGD, 1993.

Appendix F: Baseline Data and Impact Indicators - City Challenge

	BASELINE	IMPACT INDICATOR
DEMOGRAPHY	Size of population, age structure, household structure (single parent, pensioners, etc), ethnic composition.	<ul style="list-style-type: none"> • Change in population structure (age, household) • Migration levels - net gain/loss of population. Also migration of key groups (25034 year olds).
ECONOMY/ENTERPRISE	Level of business investment; number of firms, size, type; initiate a register of firms. Number of jobs (workforce). Land supply/release. Property development. Household incomes.	<ul style="list-style-type: none"> • Net inward investment (£m) • Increase in number of firms in area/new business survival rate. • Increase in profit/turnover of firms in the area. • Reduction in vacancy of commercial property (sq ft) of take-up of commercial floorspace. • Firms' improved perception of business conditions and area. • VAT registrations. • Increase in house prices relative to wider external area (see also housing).
EMPLOYMENT	Economic status of residents; level of unemployment; long-term unemployment, youth unemployment; access to employment. Take-up of benefit.	<ul style="list-style-type: none"> • Change in resident unemployment rate (total/long-term for different groups/youth). • Narrowing of the unemployment gap relative to the City/Borough/Region. • Increase in number of jobs - full-time/part-time (jobs per capita). • Increase in number of women with children under 10 in paid employment. • Decrease in households in receipt of benefit.
TRAINING	Adult literacy; skills/training audits. Training completion rates.	<ul style="list-style-type: none"> • Increase in placement or residents in jobs through Job Centres. • Improvement in skill levels (skills audit) - youth/women/ethnic groups - employed and unemployed. • CC trainees placed in full-time employment.

	BASELINE	IMPACT INDICATOR
HEALTH	Infant mortality; immunisation rates; levels of substance abuse. Health problems/needs (local data or survey).	<ul style="list-style-type: none"> • Awareness and satisfaction levels with facilities available. • Mortality rates for population aged under 75. Infant mortality - ratio of under 1 deaths to birth p.a. <ul style="list-style-type: none"> • Improvement in average birth-weight of babies. • Decrease in visits by LA environmental health officers/ reported incidents of vermin & pest control.
CRIME	Crime rates. Community safety assessment.	<ul style="list-style-type: none"> • Reduction in reported crime rates by type, eg car thefts, burglary, assaults. • Re-classification of the area with Insurance Companies in terms of home insurance premiums. • Reduction in the number of Young Offenders -re-offending. • Reduced fear of crime amongst residents.
SOCIAL	Number of social/leisure facilities. Leisure needs.	<ul style="list-style-type: none"> • Improvement in the sense of community pride; satisfaction levels of particular groups. • Increased attendance at leisure/sports activities. • Increase in the number of facilities available.
OTHER	Attitude surveys, residents and businesses.	

Source: modified from Annex C, DoE-ICGD, 1993.

Appendix G: A Chronology of British Regional Policy

1928	The Industrial Transference Board	Direct (minor) grant and loan assistance for unemployed migrants.	
1934	The Special Areas Act	Establishment of four 'Special Areas' (S. Wales, Scotland, N. East and Cumberland). Direct government assistance to industries.	The Act marked a significant shift in government policy. In contrast to the Industrial Transference Board, the initiative was based on the principle of 'taking work to workers'. This became the main underlying concept of policy ever since.
1937	Special Areas (Amendment) Act	Extension of loan powers in Special Areas. Tax, rent and rates subsidies.	
1940	The Barlow Report	Recommendations to redistribute population and industry	An influential report that had a major effect on post-war regional policy.
1944	The White Paper on Employment Policy	The turning point in Britain's regional policy. Development Areas were set up. A variety of policy instruments were enacted: grants and loans to firms, powers to build factories and establish industrial estates and provision of basic services for industry	
1947	The Town and Country Planning Act	The introduction of the most powerful policy instrument, namely, the Industrial Development Certificate - IDC.	Ironically, the government's commitment began to wane as early as 1947 in the face of a balance of payment crisis. The lack of commitment accelerated in the 1950s. It was not until the 1958 recession that regional policy assumed a central role again.

1958	Distribution of Industry (Industry Finance) Act	Extension of loan and grant assistance to high unemployment areas outside the DAs. Tightening of IDC and increased expenditure.
1960	Local Employment Act	Development Areas were replaced with Development Districts (the coverage of which varied over time).
1963	Local Employment Act	Introduction of standard investment and building grants
1963	Budget	Introduction of free depreciation for firms in assisted areas. Strengthening of the financial advantage of investment in these areas
1965	Control of Office and Industrial Development Act	Control of office development in London and Birmingham in the form of Office Development Permits.
1965	Highlands and Islands Development (Scotland) Act.	Establishment of the Highlands and Islands Development Board with extensive powers of loans, grants, equity participation and factory building.
1966	Industrial Development Act	Development Districts (a total of 165) replaced with five large Development Areas (covering 40% of the total British land area). Free depreciation replaced with a system of 40% investment grants in DAs (20% elsewhere). Other policies retained.
1967	Finance Act	Introduction of the Selective Employment Premium (SEP) and the Regional Employment Premium (REP).
1967	Special Development Areas	Establishment of Special Development Areas (mainly in pre-war Special Areas) with additional incentives and rent-free premises.

In the 1960s, regional policy became to be seen as having a central role in both achieving a faster national economic growth and alleviating the social costs associated with population and industrial concentration. The period is usually considered one of most active regional policy

1970	Local Employment Act	Establishment of Intermediate Areas. Government-built factories, building grants and derelict land clearance grants. Withdrawal of SEP from DAs.
1970	October Mini-Budget	Investment grants replaced by accelerated depreciation in DAs.
1972	Industry Act	Several major revisions: return to investment grants (the Regional Development Grant); proposed phasing out of REP; end of location controls in DAs and SDAs; increase in other forms of assistance; selective assistance to industry; grants and low-interest loans available for services.
1973	Entry to the EEC	This gave assisted areas access to loans, grants and other assistance from EEC funds (eg European Coal & Steel Community, the European Investment Bank, the European Social Fund, ...).
1974		Regional Employment Premium doubled.
1975	European Regional Development Fund	Provision of investment grants and interest rebates on other EC loans for industrial, craft, service and infrastructure investment.
1985/76	Development Agencies	Establishment of the Scottish and Welsh Development Agencies.
1976	Service industry and office location subsidies	Strengthening of grants and rent relief for firms in services and office sector locating in assisted area

1977		Regional Employment Premium abolished.	
1977	The Inner City Programme	Devolving resources to improve the economic fabric of inner city areas.	Increasing concern with the decline of inner city areas resulted in total relaxation of the IDC controls in these areas. The IDC eventually went into virtual abeyance with the growing unemployment in all regions.
1979	Phased reform of GB regional policy	A major package of reforms to be phased in during 1979-83: planned cuts in the regional budget; gradual redrawing of assisted areas boundaries; reduction in available financial incentives.	
1980	Enterprise Zones		
1982		Industrial Development Certificate abolished.	
1984	Reform of British regional policy	Cuts in annual expenditure; SDAs abolished (most downgraded to DAs); redrawing of assisted areas boundaries; extension of Regional development Grants to certain service industries; an imposed cost-per-job limit on RDGs; shift from automatic grants to selective assistance.	
1988	Reform of British regional policy	Expenditure remained unchanged. RDGs abolished and the money saved switched to Regional Selective Assistance (RSA). Introduction of Regional Enterprise Grants. Selection based on need.	

Source: modified from Armstrong and Taylor, 1993, Appendix, pp. 363-374.

Appendix H: An Abstract Example of the 'Framework' Approach

	Options				Comments
	A	B	C	Do-Minimum	
Group 1: Travellers					
Sub-Group	Effect				Units
All vehicle travellers	Time savings				£m (PVB)
	Vehicle operating cost savings				£m (PVB)
	Value of accident savings				£m (PVB)
	Reduction in casualties:				
	Fatal				Number
	Serious				Number
	Slight				Number
	Traffic delays during construction				£m(PVB) or (PVC)
	Driver stress				3 point scale: low, moderate or high
	View from road				Type of landscape
Pedestrian (number of people affected)¹	Change in amenity				Descriptive statement on increased or reduced amenity
	Change in journey time				Included in economic appraisal
	Safety				Descriptive statement on improved or decreased safety
	Severance (New)				4 point scale: none, slight, moderate and severe and, numbers affected
	Properties demolished				Number ²
Group 2: Occupiers	Properties demolished				Number ²
Residential	Noise effects adjacent to new road				Number of houses experiencing an increase of:
					More than 15dB(A) _{L₁₀}
					10-15dB(A)
				5-10dB(A)	
				3-5dB(A)	

Appendix H - continued

Sub-Group	
Residential - continued	
Noise decrease	Number of houses experiencing decrease of: More than 15dB(A) _{L₁₀} 10-15dB(A) 5-10dB(A) 3-5dB(A)
Visual intrusion	Qualitative 'before and after' comparison of the quality of landscape
Visual Obstruction	Number of properties within 300m of centre-line. Subject to: Severe Significant Slight
Severance: a. Relief from existing severance b. Imposition of new severance	Both on a 4 point scale: None, Slight, Moderate and Severe, and the number of people affected in each category.
Disruption during construction	Number of properties within 100m of centre-line and periods of exposure ³
Noise increase	Number of premises subject to an increase of more than 5dB(A) _{L₁₀} ⁴
Noise decrease	Number of premises subject to decrease of more than 5dB(A) _{L₁₀}
Visual obstruction	(see above)
Severance	(see above)
Disruption during construction	(see above)
Commercial Premises	
a) Office Buildings	
Noise increase	(see above) ⁴
Noise decrease	(see above)
Visual obstruction	(see above)
Severance	(see above)

Appendix H - continued

Sub-Group	
Commercial Premises - contd.	
a) Office Buildings - contd.	Disruption during construction (see above)
b) Shops	Noise effects adjacent to new road
	Noise effects adjacent to existing road (see above)
	Visual obstruction
	Severance
	Disruption during construction
<i>Schools and Hospitals</i>	
	Change in noise levels
	Visual obstruction
	Severance (see above)
	Disruption during construction
<i>Farming</i>	Land take
	Number of Farms affected
	Hectares of land (by Grade)
<i>Open Space</i>	Land take
	Hectares
<i>Public Buildings</i>	Land take
	Hectares
Group 3: Users of Facilities	
Sub-Group	
eg Shoppers, Users of community facilities, etc.	Reduction of vehicle/pedestrian conflict
	Change in traffic
	Reduction in amenity (land take, visual intrusion, ...)
	Severance
	Noise

Appendix H - continued

Group 4: Policies for Conserving and Enhancing the Area		
Policy	Authority	Interest
Group 5: Transport, Development and Economic Policies		
Policy	Authority	Interest
<i>Transport</i>		
<i>Development & Economic</i>		
Group 6: Financial Effects		
Sub-Group	Interest	Units
<i>DoT</i>	Construction costs ³	£m (PVC)
	Land costs	£m (PVC)
	Total cost	£m(PVC)
<i>District CC</i>	Construction costs ³	£m (PVC)
	Land costs	£m(PVC)
	Total cost	£m(PVC)

Total Quantified Monetary Benefit		
NPV compared to 'Do-Minimum'		

Notes

1. If the effect on cyclists is thought to be important, a separate sub-group should be included.
 2. Compensation costs of demolished properties are included in the economic appraisal as part of the construction costs.
 3. Costs of traffic delays due to construction is included in Group 1.
 4. In built-up areas where substantial screening is likely it will be sufficient to list only those properties fronting or backing on to the route.
 5. Construction costs should include preparation and supervision costs.
- Source: modified form MEA: Manual of Environmental Assessment, Department of Transport, 1983.

Appendix I: Illustrative Example of Environmental Impact Tables - Stage 3

GROUP 1: LOCAL PEOPLE AND THEIR COMMUNITIES						
SUB GROUP	EFFECTS	UNITS	PREFERRED ROUTE	DO-MINIMUM	COMMENTS	
Commercial buildings used by people:	Properties demolished	Number	0	0		
Retail Premises	Noise dB $L_{A(0,18hr)}$	Number of properties experiencing an increase of more than			The changes in noise are the difference between the forecast for the Published Route for 2010 and the existing levels. The units are d (A) L10 18hr 6am - midnight	
		1 < 3	0	12		
		3 < 5	0	0		
		5 < 10	0	0		
		10 < 15	3	9		
		≥ 15				
		Number of properties experiencing a decrease of				
		1 < 3	5	No Change		
		3 < 5	7	No Change		
		5 < 10	0	No Change		
Visual Impact	Number of properties subject to:					
	Substantial		0	No Change		
	Moderate		1	No Change		
	Slight		0	No Change		
	No Change		0	No Change		
Severance:						
a) Relief to existing severance		-	a) None	a) None		
b) Implications of new severance		-	b) Slight for DIY City	b) None		
Disruption during construction		-	3 premises within 100m of site slightly affected	None		

Appendix I - continued

GROUP 2: TRAVELLERS					
SUB GROUP	EFFECT	UNITS/INTEREST	PREFERRED ROUTE	DO-MINIMUM	COMMENTS
Travellers Amenity: Vehicle Users' Amenity	Driver Stress	-	Low	Eastbound high Westbound Moderate	
View from the Road					
Pedestrians' and Equestrians' Amenity	Severance (New)	Number	Moderate relief of severance	Agricultural Upland Agricultural Upland	Footpaths and Bridleways will be diverted to suit the published route. No designated Footpaths or Bridleways will be served
Change in Amenity					
		Bridleways A7	Reduction in amenity for equestrians and ramblers. Estimated usage 50 and 100 journeys/week respectively.	Existing good amenity unchanged	
Travellers Safety: Pedestrians' and Equestrians Safety		-	Segregation of pedestrians and vehicles will improve safety	No improvement in road safety	
All Vehicle Travellers' safety					
	Reduction in casualties		High Growth	Low Growth	The figures indicate the probable total reduction in casualties over the whole of the 30 years assessment period if the national average rates and distribution between groups apply to each alternative. They take no account of the safety implications of the detailed design of the new route.
	Fatal	Number	47	38	0
	Serious	Number	350	280	0
	Slight	Number	997	801	0

Appendix I - continued

GROUP 3: THE CULTURAL AND NATURAL ENVIRONMENT

SUB GROUP	EFFECT	UNITS/VALUE	PREFERRED ROUTE	DO-MINIMUM	COMMENTS
Heritage: eg listed building	Noise		Decrease of 1-3dB(A) at 3 facades; increase on 3dB(A) at fourth	No change	For note on noise see Group 1.
	Severance		Substantial reduction	No change	
	Visual Impact		Slight increase	No change	
Nature and Landscape: eg AONB	Landlake	ha	0.7 ha	No change	Existing route forms boundary of AONB
	Landscape effect		Some views of and from the AONB will be moderately worsened, although planting will mitigate adverse effects in longer term.		

GROUP 4: POLICIES AND PLANS

POLICY	AUTHORITY	INTEREST	PREFERRED ROUTE	DO-MINIMUM	COMMENTS
Structure Plan Policy number 17(a)	Identify the Authority	To generate the town centre	Removed estimated 50% of through traffic, will enable the introduction of parking and traffic calming measures	The existing poor conditions will deteriorate	
Structure Plan Policy number () and Local Plan Policy number ()	Identify the Authority	Playing fields, parks	Route requires 10% of Field but re-orientation of soccer pitches will enable all to be retained	No Change	

Illustrative Example of Land Use Table - Stage 3

EXISTING LAND USE	AREA REQUIRED PERMANENTLY		AREA REQUIRED TEMPORARILY DURING CONSTRUCTION (ha)	COMMENTS
	Area for carriageway, footways and other hard surfaces (ha)	Area for verges, embankments, cuttings and other landscaping (ha)		
Derelict Land	-	4.0 (4.0)	4.0(4.0)	() Designated development land in plans
Agricultural Land				
Grade 2	2.0	8.0	10.0	
Grade 3	8.0 ¹	19.0 (7.0)	27.0 (7.0)	
Grade 4	-	1.5	1.5	2.0 ²
Community Land				
- Public Open Space	1.0	2.0	3.0	
Other				
- Industrial	1.0	3.0	4.0	
TOTAL	12.0	37.5	49.5	2.0

¹ Includes 3ha exchange land for loss of Public Open Space

² Area required for construction will be returned to agriculture

Appendix I - continued

Illustrative Example of Mitigation Table - Stage 3

MITIGATION MEASURE	LOCATION, PURPOSE AND FORECAST BENEFIT	CAPITAL COST	FORECAST MAINTENANCE REQUIREMENT, METHOD AND COST	COMMENTS
Noise Band	CH 100 - 160 Westbound Noise Reduction of 2-3dB(A) for 20 residential properties	£10,000	No specific maintenance requirement	
Noise Barrier	CH 500 - 700 Eastbound Noise Reduction of 2-3dB(A) for 50 residential properties	£30,000	Regular maintenance required for painting and replacing panels. Maintenance from verges. Estimated cost £1500/yr	Barrier results on increased visual intrusion for 10 properties
Over deepening of balancing pond	Junction 15. Ecological enhancement will allow introduction of wetland species	£5,000	Additional maintenance required to preserve developed wetland species whilst maintaining primary function of balancing pond. Maintenance from pond side. Additional cost estimate of £2,000/yr	
Provision of interceptors	All outfalls to river to meet requirement of NRA for run-off quality. Permit draining to river	£100,000	Routine maintenance cleaning required. Maintenance from highway land. Estimate cost £4,000/yr	
Revised horizontal alignment	East of Fox Wood Nature Conservation. Avoid SSSI	£200,000	No specific extra maintenance requirement	

Source: Adapted from 'Environmental Assessment', DoT, 1993a, Section 4, Part 4, Annex AI and AII.

Appendix J: The Letter Sent to 'Invited Commentators'

Name, Address & Date

*Name & Address of
Commentator*

Dear ...

My name is Ashraf Bakr. I am a Ph.D. research student in the Department of Civic Design at the University of Liverpool. I am working under the supervision of Prof. Peter Batey on a research topic provisionally entitled: 'Impact Evaluation: Towards A Systematic Approach'. The main objective of the research is to develop a systematic, integrated approach for use in the ex-post evaluation of urban policy.

I am currently assessing reported case studies of ex-post evaluation in several fields, one of which is transport. The idea is to find out how evaluation is being conducted in practice and to identify its strengths and weaknesses. These findings will be of value in shaping the framework that is the aim of my research.

Within the field of transport, I am focusing on trunk roads and motorways/freeways construction and improvement. I am seeking case studies of 'comprehensive', retrospective evaluation in this field. By comprehensive I mean the assessment of all likely impacts of road construction/improvement, i.e. economic, social and environmental, in addition of course to its effects in traffic terms.

However, the material I have come across so far is partial, in the sense that each case study looks at only one particular type of impact. I have found no studies that evaluate *all* impacts of even a single motorway or trunk road. My main question is: has there been any such comprehensive evaluation, of even a single motorway that I have failed to locate? Or, is it the case that such practice is a myth and has never really existed?

If there is no tradition for such practice, what are the reasons? In a field where computer modelling has been common practice for decades, how can there be such a severe lack of systematic procedures for ex-post evaluation? How can such an apparent gap be bridged?

I would be grateful should you have the time to consider my questions and provide me with your comments. I would be grateful as well should you kindly provide me with references you see of relevance to my research.

Hoping to hear from you soon,

Yours sincerely

Appendix K: The Choice of A Demonstration Area

Another aspect of the Dutch example that is of wider relevance is the choice of the demonstration area. A full-scale analysis of an entire policy may not always be possible. Complexity of the policy, lack of resources or time and political pressures may all be among the reasons why it may become necessary to focus down the scope of an evaluation research. Even if a full-scale evaluation is possible, a decision will have to be made on where to start the exercise. It may as well be the case, as was with that study, that a 'pilot' project is to be initiated as a first stage of a full-scale evaluation. A choice will have to be made on which policy area is to be covered. In other words, the situation may arise where a 'prioritisation' task between the various policy areas will have to be undertaken.

The SVV II has too broad a scope for the Study team to attempt to analyze its effects in their entirety, and a choice had to be made about the policy area to be used for the demonstration (p. 25). The SVV II placed an emphasis on passenger-related goals. Three policy areas spanned these goals: safety, mobility and accessibility. The choice among the three areas was made using the "scorecard" (see Fig 10.6). The criteria are listed in Table (10.5).

Despite political interest, accessibility was rejected because the goals were not clearly defined in the SVV. Mobility was then chosen "primarily because safety was seen to be more straightforward from the viewpoint of the demonstration and less problematic from a policy perspective" (p. 26).

Fig K.1: Scorecard for Choosing Demonstration Area

Criterion	Safety	Mobility	Accessibility
1. Political interest	+	++	++
2. Generalizability	+	++	++(?)
3. Amount of effort	++	+/-	-
4. Simplicity	-	+/-	+
5. Difficulty	+	-	--
6. Definitions	++	+/-	--
7. Existing knowledge	+	+	+/-
8. Regionalization	+	+	+/-
9. Data availability	++	+	+/-

Source: Walker et al., 1993: 26.

Table K.1: Criteria for Choosing Demonstration Area

Criterion	Description
1. Political interest	Is there interest in the area among policy-makers in the Ministry? (more interest is better)
2. Generalizability	Will the methodology used in this area be applicable to others? (wider applicability is better)
3. Amount of effort	Can the evaluation of the area be carried out in the time available? (less effort is better)
4. Simplicity	Are the methodological problems so easy to solve that not much would be learned? (less simplicity is better)
5. Difficulty	Are the methodological problems for the area so difficult to solve that not much could be accomplished? (less difficulty is better)
6. Definitions	Are the proxies and response indicators related to the area's goals clearly defined? (more clarity is better)
7. Existing knowledge	Are there known relationships between the tactics in the area and the relevant proxies and response indicators? (more existing knowledge is better)
8. Regionalization	Is there differentiation in the implementation of the area's tactics among provinces? (differentiation in implementation is better)
9. Data availability	Are there data available with which to perform the evaluation? (more data availability is better)

Source: Walker et al., 1993: 27.