

Learning through Evaluation:
the Nordic Experience

Learning through Evaluation: the Nordic Experience

Kaisa Lähteenmäki-Smith (ed.)

Nordregio Report 2007:3
ISSN 1403-2503
ISBN 978-91-89332-64-5

© Nordregio 2007

Nordregio
P.O. Box 1658
SE-111 86 Stockholm, Sweden
nordregio@nordregio.se
www.nordregio.se
www.norden.org

Analyses & text: Kaisa Lähteenmäki-Smith (ed.)

Dtp: Hanna Dubois

Linguistic editing: Chris Smith

Nordic co-operation

takes place among the countries of Denmark, Finland, Iceland, Norway and Sweden, as well as the autonomous territories of the Faroe Islands, Greenland and Åland.

The Nordic Council

is a forum for co-operation between the Nordic parliaments and governments. The Council consists of 87 parliamentarians from the Nordic countries. The Nordic Council takes policy initiatives and monitors Nordic co-operation. Founded in 1952.

The Nordic Council of Ministers

is a forum of co-operation between the Nordic governments. The Nordic Council of Ministers implements Nordic co-operation. The prime ministers have the overall responsibility. Its activities are co-ordinated by the Nordic ministers for co-operation, the Nordic Committee for co-operation and portfolio ministers. Founded in 1971.

Nordregio – Nordic Centre for Spatial Development

works in the field of spatial development, which includes physical planning and regional policies, in particular with a Nordic and European comparative perspective. Nordregio is active in research, education and knowledge dissemination and provides policy-relevant data. Nordregio was established in 1997 by the Nordic Council of Ministers. The centre is owned by the five Nordic countries and builds upon more than 30 years of Nordic cooperation in its field.

Stockholm, Sweden 2007

Contents

Contents	5
Preface	7
About the Authors	9
Introduction: Learning through Evaluation: the Nordic experience by Kaisa Lähteenmäki-Smith	11
Learning from the construction of programme logic: Cognitive mapping as a participatory approach in the evaluation of regional development programmes by Ville Valovirta	21
The Impact Chain Model as an Analytical Tool for Evaluation within Research and Development by Soile Kuitunen	31
Top-down or bottom up? Methods for evaluating employment effects in Nordic Structural Fund programmes by Jon Moxnes Steineke	43
Sitting on both sides of the fence: Meshing theory and practice to construct useful regional programme evaluations by Jörgen Gren	53
Sustainable Development As a Horizontal Theme In European Evaluations by Keith Clement	61
The European Commission's Approach to Evaluation in the 2007-2013 Programming Period by Kaisa Lähteenmäki-Smith	77

Preface

The “Nordregio School of Evaluation” project was originally set up as a collaborative effort between Nordregio’s staff and some of our co-operation partners. Through the analysis of specific research questions stemming from Nordregio’s evaluation experiences it constituted an attempt to create a platform for organisational learning within Nordregio and beyond. Since the institute’s inception Nordregio’s evaluation portfolio has altered considerably, though evaluative research still constitutes one of the main types of applied research that is undertaken. Though the contributions to the volume represent only a limited glimpse into the evaluation activity in which Nordregio is currently involved, they touch upon many more generic questions of relevance to evaluative research in general. These include, for instance, what can be learned from evaluation activity beyond the limited project or programme duration, how to strike a balance between theory and practice, as well as between the interests of the commis-

sioners of evaluation and the researchers and evaluators involved in delivering them. The anthology also touches upon the question of what lessons can be drawn across sectors, organizations and territorial contexts when it comes to evaluation (European influences on national and local contexts, but also influences across policy sectors, e.g. innovation and regional development). Hopefully the volume will provide food for thought and perhaps even inspiration for evaluators and policy makers alike.

The editor wishes to thank both the specific contributors to this volume and the participants in the Nordregio workshops organised in the Autumn of 2006 for the constructive debates that ensued. In addition thanks also go to Hanna Dubois, Johanna Roto and Chris Smith for their help with the layout, illustrations and language editing.

Stockholm, September 2007

About the authors

Keith Clement is a Senior Consultant with ESD Europe Ltd. His main areas of activity include regional environmental integration and sustainable regional development, and he has substantial experience in Structural Funds programme design and evaluation. Formerly Senior Research Associate with Nordregio (2000-2004), he was the lead author of a series of reports on sustainable regional development in the Nordic countries. He has carried out specialist assignments for the European Commission, the World Bank, the OECD and the Nordic Council of Ministers, and he has been an Expert Adviser to the Institute for European Environmental Policy, the Institute of Technology Management, Scottish Natural Heritage, the Scotland Europa Environment Group, and Strathclyde European Partnership, amongst others.

Jörgen Gren holds a PhD in European Studies from the University of Cambridge and has published several books and articles on the subject of regions and regional development. He has worked with evaluation and regional development matters in a variety of posts such as regional development programme evaluator in both the UK and in the Nordic countries while also working with regional projects and programme planning in Sweden. He is currently working in the European Commission's General Directorate for Regional policy. Jörgen Gren is also a lecturer at the University of Paris I (Sorbonne) and at the University of Metz in regional policy and infrastructure finance.

Soile Kuitunen holds a PhD in Political Science from the University of Turku, Finland. She currently works as a director for Net Effect Ltd which specialises in evaluation, research and consultancy on public policies and interventions and where she is a team leader in the innovation and industrial policy team. She has published several articles and books in the field of technology and innovation policy and initiatives. Kuitunen has also been involved in a large number of evaluation studies focusing on the impacts and effectiveness of R&D&I initiatives.

Kaisa Lähteenmäki-Smith holds a PhD in Political Science from University of Turku and has worked as a senior research fellow at Nordregio since 2000. In addition to evaluation theory and practice her research interests include territorial governance, regional innovation and regional technology policies. Her PhD, completed in 1999, was on the topic of regionalisation within the EU and she has continued working on this theme ever since, initially at the Finnish Ministry of the Interior and thereafter at Nordregio. She has primarily been involved with projects dealing with regional development, regional strategies and administrative reforms in the Nordic countries, as well as with evaluation projects on regional development and innovation, both in a national Finnish context and within the EU.

Jon Moxnes Steineke holds a MSC in Economic Geography from the Norwegian School of Economic and Business Administration. Since 2002, he has participated in a range of evaluations both of the Structural Funds and of transnational cooperation programmes (Interreg), with a particular interest in improving on programme indicator systems. He is currently at Nordregio where his main research interests are regional development/urban growth, regional innovation systems, information society issues and cross border cooperation.

Ville Valovirta holds a Masters degree in Political Sciences from the University of Helsinki in Finland and currently works as a Senior Researcher at VTT Technical Research Centre in Finland. His main research interests are related to innovation policy, public policy evaluation, and knowledge-based regional development. Prior to joining VTT, he worked as a public policy consultant, contributing to projects related to technology programmes, regional innovation strategies, innovation intermediaries and technology transfer programmes. He has been engaged in evaluation studies over several policy fields and has carried out consulting and training assignments for public sector and non-profit organisations. His areas of expertise also cover strategic management, performance assessment, and evaluation methods in public policy-making.

Introduction: Learning through Evaluation: the Nordic experience

By Kaisa Lähteenmäki-Smith

Introduction: towards learning in evaluation practice

Human beings are engaged in all kinds of efforts to make the world a better place. These efforts include assessing needs, formulating policies, passing laws, delivering programs, managing people and resources, providing therapy, developing communities, changing organizational culture, educating students, intervening in conflicts, and solving problems. Quinn Patton (1990)

Evaluation is found today in almost every area of policy and public intervention. It is often portrayed as an administrative or managerial necessity, rather than having the potentially nobler goal of enlightenment or of “making the world a better place”, as reflected in Quinn Patton’s quote above. It is safe to say that evaluation has become firmly embedded in the (“New”) public management practice and agenda. For reasons of accountability, evaluation is taken to be one of the main requirements in terms of monitoring and the assessment of policy in its different forms the project and programme levels, as well as encompassing the more systemic perspectives and holistic ambitions involved in management systems such as performance steering and guidance systems - all seeking to enhance the knowledge-base upon which public policies are developed (e.g. Walker and Boyne 2006; Lähteenmäki-Smith, Hyytinen, Kutinlahti & Konttinen 2006). Evaluation as a type of professional activity, as well as applied research with evaluative ambitions and motivations play an important part in this shift towards improving the knowledge base of “what works”, i.e. enlightening and improving the effectiveness of policy, while also making policy-making more transparent and more accountable. The evaluative rationale behind policies has become so widely diffused that reference is now made to notions such as ‘the audit society’, ‘the evaluation society’ and even ‘the impact assessment society’ (e.g. Vedung 2003; Rajavaara 2007).

This collection focuses on some of the key aspects of evaluative research to have emerged in the first ten years of Nordregio’s existence, seen from the perspective of a sam-

ple of evaluation projects undertaken during these years and inspired by debates among evaluators, policy-makers and colleagues in connection with these projects. Nordregio’s researchers have, over this ten year period, been involved in a multitude of evaluative projects in the area of regional development, spatial planning and innovation, within Nordregio’s project portfolio. These projects have ranged from the evaluation of European and national programmes to project and thematic evaluations. In the early days of Nordregio’s existence in particular evaluation was one of the key types of applied research undertaken (Marjussen and Uhlin 2006). The publication is the product of a project undertaken both as an internal competence development project and as an external networking exercise, where Nordregio’s researchers, together with external cooperation partners from the evaluation field analysed some of the normative, methodological and normative challenges associated with evaluation in the area of regional development.

The intention of the project was to analyse the evaluation practice of Nordregio and use project experiences to develop insight into the institute’s methodological practice: what are the common challenges and themes that we are faced with when working on evaluations in the broad field of spatial development? What have we learned and where should one look further, in order to discover methodological inspiration or further knowledge of societal challenges and client needs, knowledge which may not have been central in our previous evaluation projects. Some of the original questions for the project included:

- What constitutes a “good evaluation” (by which standards, but also for whom, and for what purpose)?
- How to draw lessons from specific types of evaluations (e.g. Structural Funds on programme areas or projects) that could be useful for other types of evaluations (on different scales, from national to regional and local)?
- Are there ways of pursuing sector-based programme

evaluations (e.g. in the fields of R&D and innovation, urban and rural policy) that would allow for better comparative lessons to be drawn?

- How to best approach thematic or horizontal themes (environment, sustainability, equal opportunities, information society, Lisbon agenda etc.) without merely addressing them as a “necessary evil” in evaluation exercises? (Methodological, theoretical, but also more practical concerns of inclusion, participation, and learning)?
- What lessons can be learned in terms of the wider philosophical, epistemological and methodological questions related to the evaluation task?

The project has been reported in two ways: in the form of this publication and through two workshops organised in late 2006. The discussion was intended as an instrument in competence development internal to the organisation, especially in connection with the two workshops organised on the topic of evaluation in late 2006. Here some co-operation partners were invited in to make presentations and provide inspiration and knowledge on evaluative practice within the EU and in particular in relation to the Structural Funds, as well as nationally. In connection with the EU theme the contribution of Anna Byrulo from the European Commission was essential and this helped to bring our knowledge of the evaluation of the Structural Funds during the 2007-2013 programming period up to date. The more research-based national insights of some of our Finnish co-operation partners and experts with both practical evaluation experience and a solid research background was also extremely useful and here Soile Kuitunen, Ilari Karppi and Ville Valovirta were inspirational. One of the main motivations for the project was to find room for concrete methodological solutions in relation to the construction of institutional and organisational learning capacities.

The conclusions drawn from the workshop discussions and presentations made by the experts relate to all stages of evaluative practice, i.e. the motivations and rationales, as well as to concrete implementation of evaluations (e.g. methodological questions) and finally, to drawing conclusions for policy and future evaluative practice. In many cases the goal or intention of evaluation has focused on the need for improved accountability though the desire to legitimise interventions already undertaken has also been a factor here. Both Ilari Karppi's and Soile Kuitunen's presentations at the workshop in December 2006 started from this external legitimating-motivation perspective, while at the same time reminding us that while this is indeed an understandable and legitimate motivation in itself, it often falls short of identifying and fully utilising the learning potential that evaluation practice entails. As was argued by Karppi for instance, learning can in fact take place on many levels, in variable scales and with varying implications for policy practice: evaluation can help us understand the behaviour and motivations of different stakeholders,

while also helping us understand more fully the logics of interventions and the theories of action embedded in policy interventions, programmes and projects (this was also a point elaborated by Ville Valovirta in his presentation).

The legitimacy issue and accountability as a key principal for public sector governance are largely behind the emergence of evaluation as a ‘profession’, or perhaps even, dare we say it, ‘industry’. In this regard it is useful to remind ourselves of what distinguishes evaluation from other types of research and assessment endeavours. In essence, as also noted in Kuitunen's chapter, an evaluative project consists of a process of determining the merit, worth, or value of something, or the product of that process, perhaps best outlined in the description originally forwarded by Scriven (1991, 139): “the systematic collection of information about the activities, characteristics and outcomes of programs, personnel and products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel or products are doing and affecting”. The comparative aspect of assessment is essential: there is then a need to have information on the ‘value’ of something as compared to that of something else (baseline, bench-mark, original objective). The dimensions of comparison can thus vary from temporal, spatial or geographic to sector- or intervention-specific. For its focus on comparison it is not surprising that evaluation has been developed in close connection with applied research (e.g. in regional development the two go hand in hand and are in a constant process of cross-fertilisation).

Evaluation within regional development has gradually become institutionalised to the extent that it can in fact also be seen to be part of the ‘carrier wave’ that constitutes an institutional memory in regional development (e.g. Karppi's presentation elaborated on this perspective in relation to Interreg project and programme activity). This also implies that evaluation builds and accumulates the institutional understanding of what regional development is, how the dynamics work and to what extent public policy can play a role (these rationales are reflected for instance in Gren's contribution to this report). Some of the findings also identified the need for a more systematic utilisation of evaluative information and for what has increasingly become a fashionable perceptible into “regional strategic intelligence”, i.e. the knowledge base that can be accumulated by the closer integration of strategic, evaluative and foresight approaches (Kuitunen's case of R&D evaluation was particularly illustrative of this point).

The concept of, and increased interest in, regional strategic intelligence is illustrative of many of the changes in rationale across the public sector as a whole, and regional development in particular. It is not entirely new in itself, rather reflecting the longer development of the knowledge base and of our understanding of the complex dynamics of regional development as well as the need to identify and utilise more fully different existing and emergent knowl-

edge bases. The “new vocabulary” or discourse of “strategic intelligence may not however simply mask, under a management language, the old ways of using diverse information sources, but instead entail a new systemic view of applied research, evaluation information, foresight activities etc., while highlighting the utilisation of various knowledge sources in an attempt to promote a more systematic and broadly-based approach to the gathering of information relevant to regional development, in itself also reflective of a broader understanding of regional development as something that is not merely measured in quantitative economic development terms (traditionally measured via GDP or employment indicators), but rather reflecting a concern to better understand the underlying dynamics. In recent years this focus of interest has been particularly visible in the emergence of a powerful discourse on regional innovation and regional innovation systems, where the final goal may be the creation of jobs and the accumulation of economic well-being that can then be utilised in the achievement of territorial welfare and cohesion, though the elements upon which such development can be built remain somewhat more varied in nature.

The aim of making better policies and achieving more accountable and transparent public interventions is a common denominator in terms of evaluation activity, though actual utilization and/or learning that results from the evaluative activity varies. It is claimed that accountability as a motivation for evaluative activity is a minimum condition, while utilization or evaluative information brings a certain ‘value added’ to evaluation. In light of the rapidly expanding evaluation literature, we know in this context that this should already be taken into consideration when planning an evaluation, as the involvement of stakeholders is essential to useful and targeted evaluation (Patton 1997; Chen 2006, 36).

The popularity of evaluation as an instrument of steering and management is an essential part of the currently prevailing ideal of evidence-based decision-making, which is often taken to include the more limited type of cognitive base, i.e. research results. Yet it can also be argued that the whole information and knowledge base used in the policy field also involves evaluation information, which though more limited in scope often uses research as a cognitive base in turn (e.g. what we know about causalities and connections between phenomena based on research findings will be reflected in how we formulate our evaluation questions in sector specific cases). What accounts for evidence or evaluation is also an interesting question. In addition to traditional forms of academic research, evidence can be defined (see Davies, 2004) as:

- The informed interests of stakeholders and constituents in the policy area
- Information gathered from literature reviews, including internet searches
- Public attitude evidence including that garnered from

opinion polling and focus groups

- Statistical evidence drawn from government and non-governmental sources (e.g. the OECD).
- Economic data including cost-benefit analysis or modelling.

This is similar to though it provides a more instrumental perspective on what was above referred to as “strategic intelligence”. Needless to say there are a variety of other forms of information, deliberation, rhetoric and interests that impact upon any single policy decision. The evidence, such as that provided by an evaluative process, can also be both appreciated and utilised to varying degrees. As the famous quote from John Maynard Keynes suggests, “There is nothing a government hates more than to be well-informed; for it makes the process of arriving at decisions much more complicated and difficult.” (Keynes cited in *ibid*, 2). This is of course not only a characteristic of the policy process but also reflects an inherent reluctance to change and to undertake organisational or management reforms more broadly. Here the limitations of learning (from evaluation) are likely to be quite similar to other types of organisational learning contexts and situations. Here we can cite Nutley and Webb (2000) who have summarised the enemies of evidence based policy as follows:

- Bureaucratic logic: the logic that says things are right because this is the way they have always been
- The bottom line: the logic of the business environment
- Consensus: extensive consultation to find the solution that “works” (e.g. in most cases settling for the most non-offensive, ‘smallest common denominator’ solution)
- Politics: the art of the possible, “muddling through”
Civil service culture: strong distrust of information generated outside the system
- Cynicism: an attitude of mind that allows us to go along with the “company view” or “conventional wisdom” even when we know it to be false
- Time: “No wonder there is so little room for evidence-based policy: there is scarcely room to think”

Many of these issues also pose challenges to traditional forms of evaluation practice and, as such, are likely to be reflected in the level of ambition, time-scale and objectives of evaluation, and more specifically in the implementation and utilisation of evaluation results. The establishment of evaluation practice as an essential and stable part of good management and policy practice (as indicated by Jørgen Gren’s chapter for instance) has however helped to deal with some of these challenges. It is however still the case that evaluation needs are often identified too late, while evaluations implemented for external legitimating purposes, are not expressly used to develop organisational learning (e.g. Saarinen, Hyytinen and Lähtenmäki-Smith

2007; Van der Meer and Edelenbos 2006). Therein lies a challenge for the evaluators: to be as concise as possible in formulating the evaluation findings and their policy implications. Other means of enabling learning from evaluation are identified in the context of the individual chapters of this report, summarised in the brief section entitled “what implications for evaluation practice” below.

Though our focus here has been on evaluation in the strict sense, evaluative assessment is undertaken in all areas and during all stages of policy elaboration and throughout the policy cycle, from initiative to development, as well as implementation. This is why the motivations for, and the standards by which, evaluation is undertaken (legality, legitimacy, effectiveness, efficiency) are also part and parcel of the normal day-to-day assessment of policy options within policy practice (e.g. Majone 1989, 169-170). Thus it may be justified to refer to evaluation practice as the process of learning, as it stretches beyond the evaluation as a specific type of professional activity, undertaken by professional evaluators within the evaluation community, thus reflecting the numerous roles of evaluators (Chen 2005, 268).

Both evidence-based and knowledge-based policy are undoubtedly valuable and in line with the broader societal rationales necessitating a multi-disciplinary and cross-sector knowledge base. More information is however needed in order to achieve enlightened policy choices in an increasingly interdependent and complex society. In addition to the traditional legitimisation motivation (policy intervention, funding or indeed an organisation's existence), the additional dimensions more implicit in evaluative practice, e.g. learning and organisational development have also become the source of interest in policy evaluation. The rationales and roles of evaluation are in fact various, ranging from evaluation as a management tool, assessing the intervention in order to justify its existence in the sense of addressing the legitimacy function referred to above. The need to justify a policy intervention or to verify its appropriateness has many different aspects, with one of the most central being the need to ensure the accountability of public intervention. As argued by Kuitunen in her chapter, there is within R&D evaluation for instance, something of a gradual shift taking place from efficiency to effectiveness concerns, reflecting the accountability philosophy: the main focus of which encompasses the more societal aspects of R&D policy intervention. All in all, the traditional rationale for policy evaluation is rather more instrumental or even technical, in line with the role of policy intervention, linking evaluation to the social engi-

neering-dimension of policy, though an enlightenment perspective is also implicit, as is an interactive, learning role and motivation (Bryant 1995; see also Hansson 2006, 162).

As referred to by many of the authors of this report, it is the question of how it works, rather than whether it works, i.e. the question of understanding mechanisms and preconditions (e.g. Gren, Valovirta) that is at the heart of the evaluation task. Sometimes the evaluation may be required or have the indirect effect of actually forming the understanding of the interventions' logic and programme theory. The ‘theory of change’ approach to evaluation has an in-built learning process, though naturally even in the ideal-type case of the reflexive understanding of a process of change that the policy intervention seeks to put in motion, many limitations and bottlenecks remain. It is also worth noting that in assessing this question the question of what counts as evidence, and how different types of evidence are prioritised, touched upon earlier in this introduction also remain relevant.

In this report, the main cognitive interest remains firmly focussed on the basic question of “what works”. This also requires a better understanding of the causalities and impact chains through which change and development occurs. Instead of understanding the logic of interventions in terms of a summative assessment, i.e. a concluding assessment based on the accumulated findings of project or programme evaluation, an exposition of the whole rationale underlying the expected causal chains is required. Here ‘theory of change’ approaches are central to the evaluation endeavour (e.g. Chen 1990, 197-198). ‘Theory of change’ approaches were developed as it was seen that many evaluations within societies involved multi-level and multi-dimensional processes of change, linkages and logics, that together constitute a theory of change – a systematic and cumulative approach to assessing policy intervention, in the sense of finding a place for activities, outcomes, results and contexts in a comprehensive manner. Ideally, it has been perceived that this approach could at least be a partial ‘cure’ to the attribution problems familiar from evaluation – asking, for instance, who are the stakeholders, and what is their involvement and role? Theories of change expect that by identifying the relevant stakeholders of the intervention and giving them a place, a role – not only in the intervention, but also in developing and evaluating it, developing better-suited policy interventions in a particular context, as well as allowing for inclusive, participatory and consensual interventions.

‘Theory of Change’ Approach summarised

A theory of change (TOC) is a tool for developing solutions to complex social problems. A basic TOC explains how a group of early and intermediate accomplishments set the stage for producing long-term results. A more complete TOC articulates the assumptions about the process through which change will occur and specifies the ways in which all of the required early and intermediate outcomes related to achieving the desired long-term change will be brought about and documented as they occur.¹

Steps in the Creation of a ‘Theory of Change’

1. Identify a long-term goal.
2. Conduct “backwards mapping” to identify the pre-conditions necessary to achieve that goal.
3. Identify the interventions that your initiative will perform to create these preconditions.
4. Develop indicators for each precondition that will be used to assess the performance of the interventions.
5. Write a narrative that can be used to summarize the various moving parts in your theory.

A similar idea is outlined in the programme theory concept, with the idea being that each programme embodies its own inherent conception of the structure, functions, and procedures that are necessary or appropriate to attain its goals. This conception constitutes the «logic» or plan of the programme.

The original motivation can therefore be seen as quite instrumental in its nature, and gradually the core of our endeavour becomes one of learning, in different guises:

- Policy-learning, through programme logic and different stages of evaluative practice: how do policies work and how can we reflect the policy learning in evaluative practice?
- Theoretical learning: what are the underlying theoretical assumptions and normative commitments in research? How explicit should one be about the theoretical underpinnings of an evaluation exercise?
- Methodological learning – what works (and what does not) in terms of methodologies?
- Thematic and sector-specific learning: what happens ‘out there’ in neighbouring policy sectors? Spatial development is in itself not a single clearly defined field; rather it is a multi-disciplinary meeting place of different sectors and different disciplines. This makes the need for cross-sectoral learning all the more pressing.
- ‘Customer’ and ‘stakeholder’ knowledge – which are

¹ Adapted from Anderson, A. (2005). *The community builder’s approach to theory of change: A practical guide to theory and development*. New York: The Aspen Institute Roundtable on Community Change.

Source: <http://www.gse.harvard.edu/hfrp/eval/issue30/expert3.html>

the networks and linkages of relevance for evaluative practice and how do we best serve them? How can evaluators be better in tune with the different needs, interests and even tensions that exist between different types of clients, commissioners, end users, as well as the academic world, especially as the boundaries between these spheres are ambiguous or unclear in themselves?

In addition to the legitimization function of evaluation, learning thus implies that evaluations (in the EU as elsewhere) have an important systematization function, as they contribute to rendering information more comparable and commensurate. In this respect evaluation has important system effects in the development of policy influence and diffusion and in seeking to manage the social complexity of policy contexts (e.g. Gren’s chapter highlights these motivations and rationales underlying evaluation). This has also allowed for a more active informational function for evaluations: one of the main functions of EU evaluations is thus to inform the EU institutions and (in the best case scenario, EU citizens), of EU policies and their effects. These functions are, in part, also reflected in the perceived ‘professionalisation’ of the field: as the need for comparable and easily accessible information increases, there is an increasing need for persons who have the skills to produce and deliver such information.

The European Union, with its shared methodologies and approaches to evaluation has contributed greatly to the professionalisation and institutionalisation of evaluation activities. Though the information produced is often quite ‘managerial’ in nature, with the EU having little, if any, aspirations in respect of radical changes in policies and programmes, such information nevertheless provides a necessary starting point for both the profession and the wider community more generally to take further.

This report seeks to address the nature, extent and impact of learning within the programme cycle associated with European and national policy programmes. Our main research question is: what is there to be learned from Northern European experiences with programme evaluation in light of the ongoing policy changes within the European Union? This question can be organized into a number of sub-questions, which provide us with a better idea of what kinds of questions are relevant in this policy context. Some of the sub-questions are summarised below.

• Quality of evaluation:

What constitutes a ‘good’ evaluation (from the point of view of policymakers, researchers, and the wider public)? This relates both to the evaluation criteria as a methodological issue, as well as to the motivations for evaluation.

Different actors related to evaluation activities (policy makers, evaluators, practitioners, researchers) have different expectations and interests when it comes to policy evaluation. The question then becomes how could these be better integrated and indeed better served.

In terms of quality, we can also ask what constitutes a 'good' evaluation in a more normative sense. What are the ethics of evaluation? Are there ideal types or practical guidelines to pursuing ethically sound policy evaluation, and similarly, are there potential bottlenecks in this regard?

- **Goal attainment:**

An integral part of the programming cycle is the assessment of goal attainment. This is in fact a wider concern than that of merely assessing the 'results' or 'impact' of programmes, as important as these are. In assessing the goal attainment of an intervention one needs to ask, did the programme do what it set out to do, was the strategy implemented as planned, and were the results and impacts such that they could have been forecast or planned in the programming stages? Here the utilisation of indicators helps to measure the objectives to be achieved. Indicators are in many cases quantitative, as described in the European Commission's MEANS collection.

The information provided by an indicator is of a quantitative nature and is used to measure facts or opinions (e.g. percentage of regional enterprises which have been assisted by public intervention; percentage of trainees who claim to be satisfied or highly satisfied). An indicator must, among other things, produce simple information, which is easy to communicate, negotiate and decide. For that purpose, it should preferably be linked to a criterion on the success of the intervention. It must reflect precisely whatever it is meant to measure (validity of construction). The indicator and its measurement unit must be sensitive, that is to say, the quantity measured must vary significantly when a change occurs in the variable to be measured. Indicators may be specifically constructed by the evaluation team and quantified by means of surveys or statistical data. They are often borrowed from the monitoring system or statistical series. An indicator may be elementary or derived from several other indicators in the form of ratios or indexes. (Means Collection Vol. 6,67)

Goal attainment is then based on indicators, though it goes beyond merely verifying them. It also includes an assessment of the results of the programme/intervention (advantages addressees obtain at the end of their participation in a public intervention), as well as the impact assessment (assessing the consequences of a programme/intervention for addressees following their participation in an

intervention, or an indirect consequence affecting other addressees). In many cases there are important temporal or strategic dimensions to the goal attainment aspects of a programming cycle, as impacts may (at best) be long-term ones (sustainable results) and they may be positive, negative, expected or unexpected (thereby also influencing the next stages of the programming cycle and the strategic outlook of a programme during the next cycle). As the indicators are only able to provide us with a limited picture, i.e. of the gross impact of the intervention, one of the main themes of interest here will be how to measure net impacts (deducing from the gross impacts the deadweight and displacement impact) and how to achieve reliable measurement of impacts.

Evaluation as a learning instrument:

How can we draw conclusions from specific types of evaluations (e.g. the Structural Funds) that could both be readily available and useful for other types of evaluations (national, regional or local)? How do the Structural Funds as an object of evaluation differ from other evaluations, and what are the lessons to be drawn here? The lessons here relate to issues such as the characteristics of administration and management structures, 'additionality' and partnership. Also to be addressed in this respect is the potential 'value added' of programming, specific problems with trans-national or cross-border programmes, as well as the role of different actors in the management process of programmes (e.g. the Commission, national authorities, but also the local level, businesses etc.).

Are there ways of pursuing evaluations of different types of national programmes (e.g. in the fields of technology, expertise and competence development, urban and rural policy) that would allow for better comparative lessons to be drawn? While a lot has been learned through comparative evaluations a significant amount remains to be done in methodological terms, as well as in respect of indicators etc., in order to improve the commensurability of national approaches to programming regional development. The picture that emerges from the contributions tends to support the view of a certain level of scepticism when it comes to quantitative macro-economic impacts, as well as promoting an approach that is both more context-sensitive and regionally responsive. Thus issues such as the micro-macro level discussion and the need to address multi-level governance impacts can be seen in the light of this need for evaluation to somehow... 'make a difference'. In fact this is increasingly seen even by the Commission itself, who recently stated to this effect in the Third Cohesion Report (CEC 2004) that, while [the quality of] evaluation has improved over time, it varies considerably between Member States in the ways it is implemented. Interestingly enough, the report also states in this context that more involvement of regions and Member States in the process might make it more useful and relevant (ibid, 16).

Relating the research questions to the policy challenges

The above-mentioned research questions need to be addressed both in terms of policy content, such as work with cohesion policy and reforming the Structural Funds after 2006, and in terms of policy scope, i.e. with the changes that the European Convention and EU enlargement entail, where should the initiative and responsibility for the different aspects of policy programming lie? The approach is both empirical and theoretical and the cases upon which the individual chapters are based are all from European and Nordic countries' national or European policy programmes that, in some way or other, aim to promote regional development, growth and welfare. The policy dimensions covered by the cases, moreover, range from the environmental and sustainability fields to innovation policy and industrial renewal.

There are two main organizing themes running through the book providing it with a common structure beyond the wider issue of addressing a broader shared theme, namely that of evaluation. On the one hand the project or programming cycle that places the contributions into a wider thematic and temporal structure, and on the other the policy challenges that the various contributions seek to address. Together these place the different contributions into a shared policy framework that seeks to provide an interesting overview of regional and spatial planning evaluation for practitioners dealing with evaluation in their ever-day work, as well as for the more academically-orientated readers interested in the methodology or theoretical underpinnings of evaluation.

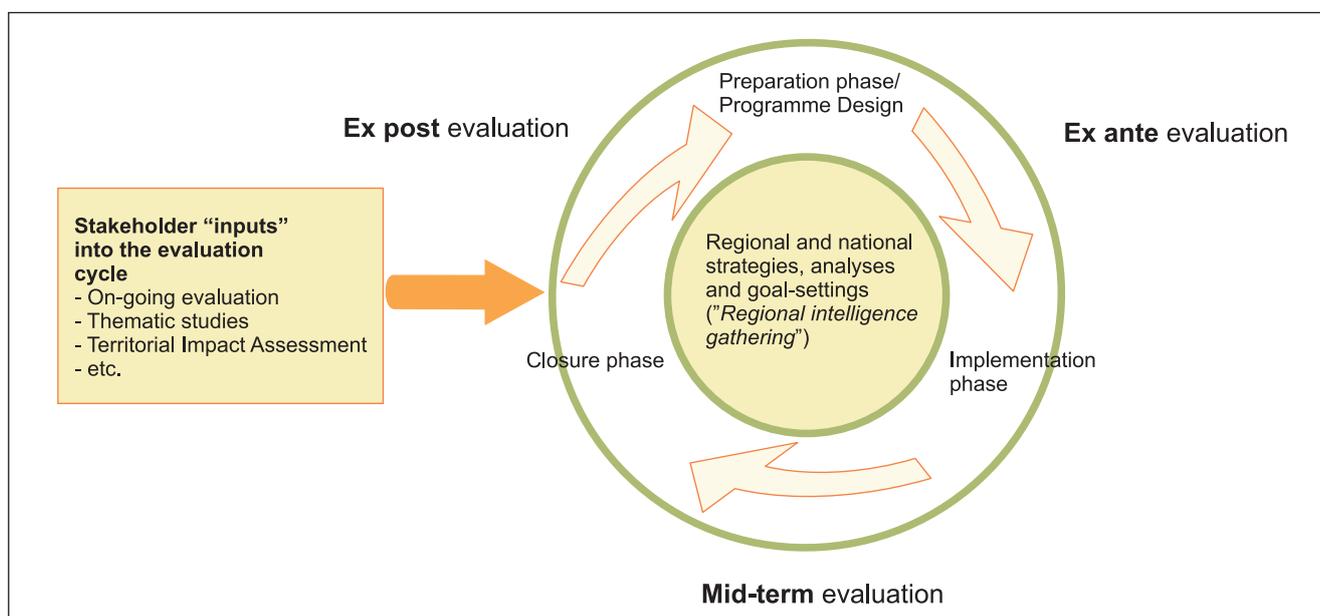
The programme cycle and the evaluation practice that has been built upon it provide a shared regime within the policy field in question. This means that the actors and 'stakeholders' involved in different aspects of the programming cycle have a shared understanding of what to expect:

the evaluation regime that the European Union has set out provides a shared set of implicit and explicit principles, norms, rules, and decision making procedures in the policy sector in question and thus ideally enables the actors to relate to the programming activities of their peers in other countries and therefore makes the learning process potentially international or trans-national. Many of the contributions to this volume in fact address this very issue and provide an interesting picture of the experiences gained in respect of organizational learning beyond national borders. The programming structure also ensures that stakeholders are consulted and relevant information is available, such that informed decisions can be made at key stages in the life-cycle of a project. This is supposed to contribute to the long-term nature of programming, as well as contributing to the transparency of the programming process. A simple picture of the programming cycle as it exists in respect of EU policies, as well as increasingly also in national policy-making, is provided below.

While the European-level regulatory system provides a context for the way in which policies and interventions are planned, implemented and administered within the EU Member States, significant room remains for national variation. This issue will also be addressed in the contributions to this volume. While remaining sensitive to such national variations, we do however also seek to address those issues that go beyond national Nordic perspectives, namely, what can be learned from the specific national experiences with the programming cycle and evaluations in the wider European context?

The challenges are, in many cases, very similar and relate in particular to the nature and quality of the monitoring and overall coherence of the whole programming cycle. In this context, we only get out what we put in so to say, while the attention given to financial issues and quantitative indicators is likely to influence the strategic over-

Figure 1. Evaluation / policy cycle (Developed for instance in Lähteenmäki-Smith et al 2006.)



view and set-up of the programme as a whole. As was argued recently by the Commission (CEC 2004, 16), particular attention needs to be given here to the monitoring system as a whole. It is easier to monitor quantitative indicators (though even this is not that easy, one might argue), but the focus on financial issues rather than strategic ones tends to lead to (Structural Funds) financing being spent where it is most easily absorbed, not necessarily where it might be most effective (or most needed). As argued by the Commission, “although improvements have been made in the present period by identifying indicators and targets, the former are often not well defined” (ibid.) The decentralisation of management systems over time is also seen as a positive factor, as it is this that has made the systems more responsive to regional needs (ibid). Thus we could conclude here that the ‘universalisation’ of the programme methodology and the programming cycle has contributed to more comparable and responsive programmes, though only sufficient regionalisation and regional responsiveness can ensure the further improvement in terms of the effectiveness of public policy. There are considerable advantages (not least in terms of learning purposes) to working with the same methodology, but this should not be applied in a uniform fashion and evaluation, as with the other steps in the programming cycle, needs to be regionally sensitive.

The various contributions to this book thus address the different phases of the cycle, reflecting the different considerations that come into play within the context of this cycle. The picture sketched out above puts the themes addressed into the programming cycle perspective. There are also however a number of contributions that concentrate on the planning stages, but in most cases the main focus is with the impact side of the equation, something that is probably not surprising if one considers the emphasis that is put on achieving and improving policy efficiency and policy delivery within nation-states, the EU, and specifically, within the administrative levels involved therein.

A programme is defined by the European Commission’s guidelines for evaluation activities as an “organized set of financial, organizational and human resources mobilized to achieve an objective or a set of objectives in a given lapse of time” (MEANS Collection Volume 6, European Commission 1999, 33). Our interest here thus necessarily focuses on the different aspects of this process from financial aspects to organisational and human ones, with the main focus being on the learning that can result from each of these. The lifecycle of a programme is here defined as “design, implementation and ex post evaluation” (ibid.). When the programming cycle is analysed further, one can discern the following dimensions in the MEANS approach:

- The analysis stage: analysing the regional characteristics (SWOT etc.) and relating these to the identification of policy options, objectives and challenges, relevant actors and policy fora, regulatory constraints and

opportunities, policy dimensions of different activities;

- The strategy stage, i.e. the formulation of objectives, strategies and visions within the environment in question;
- Decision-making regarding the way in which the means available and the ends formulated are to be related to each other: who in other words is to do what, with what means available and within which wider policy framework (including the timetable for the intervention), which priorities and measures are most suitable in order to achieve the set goals etc.;
- Implementation of the measures and interventions, the practical organisation of activities around a particular programme and setting it into a wider regional/national framework;
- Monitoring of the priorities, measures and interventions;
- Evaluation of the results, effectiveness and impact of the programme/intervention/measures

After the evaluation stage one then returns back to the beginning of the cycle, to the analysis and identification of the options available etc., thus making the cycle complete. The cycle is familiar from the well-established and influential evaluation models developed by the European Commission (MEANS Collections and their successors, described in more detail in the chapter below presenting evaluation practice in the Structural Funds context). What is important here however is to bear in mind that the systematic utilisation of different knowledge bases and information sources, highlighted in connection to “regional strategic intelligence” is also relevant for better understanding the logic of Structural Funds interventions. Thus also the cyclical picture is increasingly complemented by stakeholder information and by various types of feedback and information stemming from, and relating to, dynamics of change in the environment.

This basic programming cycle picture outlined above thus provides the book’s structure, holding the different contributions together, though there is also an additional organising mechanism underlying the chapters: moving from more specific local and regional context to thematic and EU-wide examples or discussions. The first section after the introductory chapter deals with the quality of evaluation, this is because it is an overarching concern of the whole project and vital to all stages of the programming cycle. The notion of “what constitutes a good evaluation” has both an instrumental and a purely normative dimension to it, and both are addressed in the three chapters dealing with the quality of evaluation. As one of the main questions of evaluation practice is necessarily linked to the notion of goal attainment, this theme however remains present throughout the report. The ‘objective’ and ‘subjective’, as well as the qualitative and quantitative aspects of goal attainment are considered. Needless to say, goal at-

tainment may at times remain “in the eye of the beholder” and therefore different types of goal attainment scaling techniques are often used in evaluations in order to quantitatively measure stakeholder perceptions of project results, outputs and outcomes. The differences and indeed the contradictions between the different methods of assessing goal attainment will thus be of particular interest.

Normative goals considered part of the programming cycle range from the ‘usual’ horizontal goals of equality of opportunity/gender, and sustainability/environment to latecomers to the horizontal approach such as the Information Society and territorial cohesion. The focus on moving towards mainstreaming in these issue areas is clearly evident, though major differences remain in adopted approaches to addressing such horizontal themes, both among the various programmes and between the different Member States. While the 1996 Commission guidelines for the adjustment of Structural Funds programmes for the programming period 1997-1999 for the first time made specific reference to equal opportunities as a ‘horizontal priority’, the Information Society theme has only really made its mark during the current programming period. Today it is however now addressed both in national and European programmes, in thematic and horizontal terms and with full incorporation into the evaluation field (e.g. EPRC 2001, 7). This however raises the question, how does the addressing of such horizontal themes of strategic importance impact on the programming cycle, and is the quality of programming impacted by the inclusion of such themes? On the other hand, how has evaluation been able to incorporate these horizontal themes into its methodological arsenal? Questions thus relate to both the nature of the horizontal themes as policy priorities and political choices, as well as to the impact of their inclusion into evaluation methodology in terms of quality and the nature of evaluation practice.

Horizontal themes are an important element in evaluating the effects and results of the Structural Funds, as well as other types of regional development activities. Equality of opportunity and sustainable development are those perhaps most often addressed in this regard. It is worth noting here however that ‘sustainable development’ implies different things to different stakeholders, and thus evokes different cognitive interests. As Keith Clement reminds us in his chapter, “For policy-makers, the concerns include identifying the efficiency and effectiveness of policy implementation, whether economic measures have delivered increased sustainability, and the extent to which policy in-

tegration has been achieved. For researchers or evaluators, the focus is wider, ranging from the conceptualisation of sustainable development (SD), the strengths and weaknesses of the methodologies utilised, and whether SD has been carried through to project implementation. For the wider public, the priorities relate more to understanding the concept and its implications, for example whether job-creation has associated costs in terms of loss of heritage, whether the community benefits in broad terms, and whether electoral promises have been delivered.” Each of these understandings necessarily leads to different expectations. This ‘contextuality’ and constructivist starting point is thus, naturally, not only relevant for sustainability but also for evaluation and evaluative research as a whole.

One of the key findings in this report is identification of the need to identify different sources of methodological inspiration and to test various tools and methods. The fact that there is a firmly established evaluation community of good experience and comparative grounds for testing and assessing the usefulness and validity of various evaluation methods is an important element in understanding better the policies and their impacts, as well as understanding how they can better be assessed. This is one of the benefits of developing more formalised evaluation practice within the regional development arena, as illustrated by many of the chapters in this book, which in turn is a prerequisite for more systematic comparisons across time and scale. In some areas of intervention evaluation practice is only now being developed more systematically, as illustrated for instance in Clement’s analysis, where he points out that “Nordic regional initiatives approximating to sustainable regional development have been characterised by diversity with no common frame of reference, and with evaluation seeking comparable data on projects that differ in substance, sector, time frame and funding sources, encompassing bottom-up, top-down, and even middle-down initiatives. In addition, Nordic sustainable development activity has been concentrated at national and local levels, with an absence of specific regional guidance against which to measure progress.” As compared to the national interventions, European interventions through the SF framework are potentially more structured, following a “pre-determined operational framework”. The achievement of a more formalised evaluation practice allows for better comparability and thus also enhances accountability and transparency, thereby fulfilling some of the main underlying motivations for evaluation in the first instance.

And finally...

The motivations for and rationales of evaluation are versatile, as are the methods and practices associated with them. The intention of the project reported here was to bring together some of the experiences gained from the evaluation field where Nordregio and its researchers have been active over the last ten years and to identify some examples of institutional and organisational learning through evaluation, as well as providing a forum for learning in itself. The report can of course only provide a partial assessment of the lessons learned and the perspectives that the contributors have identified as illustrative of evaluation practice today. With the new programming period new opportunities have however emerged and the interest in various evaluation methods and in learning across policy fields themselves now bring the possibility of pursuing evaluation processes as integrative learning processes in themselves much closer. Accountability and transparency within the public sector remain among the main external drivers of evaluation, but there is also a much more pronounced interest and need to use evaluation for internal development and learning purposes. This debate has only just begun.

References

- Bryant, C.G.A. (1995) *Practical Sociology: Post-Empiricism and the Reconstruction of Theory and Application*. Cambridge: Polity Press.
- Davies, P. (2004) "Is Evidence-Based Government Possible?", Presentation at "Campbell Collaboration Colloquium", Washington D.C. 19th of February 2004. Electronic document available at: <http://www.policyhub.gov.uk/downloads/JerryLeeLecture1202041.pdf#page=1>, 2004.
- Chen, H-T (2005) *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness*. Thousand Oaks, CA: Sage.
- Hansson, F. (2005) "Organizational Use of Evaluations", *Evaluation*, Vol. 12, No. 2, 159-178.
- Lähteenmäki-Smith, K., Hyytinen, K., Kutinlahti, P. & Konttinen, J. (2006). *Research with an impact. Evaluation practises in public research organisations*. Espoo: VTT.
- Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy process*. New Haven and London: Yale University Press.
- Mariussen, Å and Uhlin, Å. (2006) *Trans-national Practices. Systems Thinking in Policy Making*. Stockholm: Nordregio.
- Davies, H.T.O., Nutley, S.M. and Smith, P.C. (2000) *What Works? Evidence-based policy and practice in public services*. Bristol: Policy Press.
- Quinn Patton, M. (1997) *Utilization-focused evaluation: The New Century Text*. (3rd ed.) Thousand Oaks, CA: Sage.
- Quinn Patton, M. (2000) "Utilization-focused evaluation" in Stufflebeam, D.J., Madaus, G.F and Kellaghan, T. (eds): *Evaluation models: viewpoints on educational and human services evaluation*. Boston: Kluwer Academic Publishers.
- Rajavaara, M. (2007) *Vaikuttavuusyhteiskunta. Sosiaalisten olojen arvostelusta vaikutusten todentamiseen. [The outcome-oriented society. Styles of reasoning about the outcomes in social policy.] PhD Dissertation published by The Social Insurance Institution (KELA) in Finland.*
- Saarinen, E., Hyytinen, K. And Lähteenmäki-Smith, K. (2007) "Kohti kehittävää vaikuttavuuden arviointia T&Ktoiminnassa" ["Towards a more developmental approach to Impact Assessment within R&D", Article proposed for publication at the Finnish Journal for Public Administration.
- Scriven, R. (1991) *Evaluation Thesaurus*. Newbury Park: SAGE.
- Van der Meer, F. and Edelenbos, J. (2006) "Evaluation in Multi-Actor Policy Processes", *Evaluation*, Vol 12(2), 201-218.
- Vedung, E. (2003) *Arviointiaalto ja sen liikkeelle panevat voimat. [The Wave of Evaluation and its driving forces.]* Helsinki: STAKES.
- Walker, R.M. and Boyne, G.A. (2006) "Public Management reform and organizational performance: An empirical assessment of the U.K. Labour government's public service improvement strategy", *Journal of Policy Analysis and Management*, Vol. 25, No. 2, 371-393.

Learning from the construction of programme logic: Cognitive mapping as a participatory approach in the evaluation of regional development programmes

By Ville Valovirta

Abstract

Evaluators often face evaluation missions where the principal endeavour turns out to be figuring out what the programme really is that should be evaluated. The programme logic or programme theory needs to be clarified first, only after which can the programme be sensibly evaluated. In fact, the programme intervention logic often has to be constructed from scratch, since it has not been previously explicitly expressed. The question ex-

plored in this paper is to what extent can the theory construction process, when carried out collaboratively, become a vehicle for learning. What can programme staff learn from their activities if they are involved in deliberating and constructing their programme's logic with an evaluator? This question will be illustrated with observations from a regional development programme in Finland.

Introduction

Coupling learning with evaluation seems natural. What are evaluations good for if people are not able to learn something from them? If evaluations did not contribute to better knowledge and understanding of public policies would they be needed? Policy makers and evaluators have realised the need to increase the capability of institutions and organisations to learn from their experiences. Increasing expectations are set for evaluation to play a significant role in both policy- and organisational learning.

In theoretical terms, however, it has not been an easy endeavour to conceptualise the locus and modes of learning in the evaluation process. Only recently have we begun to gain a better understanding of the stages at which learning takes place and the forms that this takes within the

evaluation process. One of the most fundamental observations here concerns the acknowledgement of the fact that the evaluation process in itself constitutes an important source of learning (Patton 1997, 1998, Weiss 1998, Forss et al. 2002). When actors engage in defining the evaluation mission and participate in data collection for an individual evaluation process they can learn a great deal about their own activities. Carol Weiss has proposed that one of the things that may be learned in undertaking an evaluation is a better understanding of the logic by which the social intervention in a policy or a programme is supposed to achieve its societal goals (1998, 28). More often than not, however, evaluators face situations where the principal task turns out to be one of simply making sense of the na-

ture of the programme, project or intervention to be evaluated. Without a description of the logic or model through which the programme is intended to generate the desired outcomes, the whole evaluation faces serious difficulties. In essence, it is often only once a representation of the programme logic or theory has been first clarified, that evaluation can be meaningfully carried out. In fact, the programme intervention logic has often to be constructed from scratch, since it has not previously been explicitly formulated. The causal assumptions often remain as implicit thinking in programme managers' minds.

The aim of this article is to explore the approaches and practices of programme logic construction and its potential benefits for evaluation stakeholders through one concrete evaluation case. First, we will discuss the major ap-

proaches one can apply when identifying and determining the programme logic. We will argue that a participatory approach is a useful alternative when engaging stakeholders is a viable secondary objective for evaluation. Second, we will explore in more detail the benefits for stakeholders and for the evaluation process of a participatory approach. Learning will be identified as an important outcome of the theory construction process. The specificities of learning through involvement in deliberating on and constructing the programme logic will be studied through a regional policy programme evaluation case. We will conclude by raising some issues in relation to the closely related notions of programme theory, programme logic and programme strategy and their applicability in various settings

Constructing a programme logic: models and practices

The recent discussion on evaluation methods has paid increasing attention to the issues of programme logic and programme theory. Producing a schematic representation of the causal logic of the programme interventions and its intended outcomes is thus now becoming a standard tool in the evaluators' toolbox. The basic idea here is that the evaluator should identify how the programme is thought to attain the intended outcomes. The outcome of this exercise should be a map-like representation of the programme elements. The programme as the target of evaluation should be decomposed into its constituent components: inputs, activities, outputs and outcomes. And most importantly, causal links between the components should be established. This kind of exercise should give the evaluator a better understanding of the programme and thus contribute to a high-quality and analytically sophisticated evaluation study.

Various notions have been used in the evaluation literature to conceptualise the task at hand. Carol Weiss speaks about 'program theory' and 'the program's theory of change' (1997, 1998); Evert Vedung uses the concept 'intervention theory' (1998); and Michael Q. Patton refers to the program's 'theory of action' (1997) as a major step in the evaluation process. Despite somewhat diverging emphasis they all endorse the same fundamental notion that identifying and depicting causal links between programme inputs, activities, outputs and outcomes is an important and useful step in an evaluation process.

Identification of the programme logic will also help to determine whether the programme design is valid and sound. Martin Rein has claimed that "a programme is a theory and an evaluation is its test" (1981, quoted in Weiss 1998). The programme designers are making assumptions about how the planned activities are expected to change

the conditions of the programme beneficiaries. The validity of these 'if-then' assumptions – the programme 'hypotheses' (Kettner et al. 1999) – will then be tested once the programme becomes implemented.

Analytically speaking, then, the inability of the programme to deliver the intended outcomes can be attributed to two kinds of failures. The first is a 'theory failure' (Weiss 1997) where the causal assumptions prove to be false. The programme activities do not generate desired social changes (outcomes) in the target population. The theory of social change is thus flawed. Sometimes an evaluator can already point this out at a prior assessment phase and suggest modifications to the underlying programme theory. In other cases not enough knowledge is available to test the validity of the logic prospectively. It needs to be tested through actual deployment of the programme to see whether it works or not. The second type of failure is an 'implementation failure'. In this case the actual deployment of the programme has failed to deliver the planned outputs in the intended manner. Poor effectiveness is in this case not caused by a flawed theory of social change, but because the planned outputs could not be delivered due to difficulties in implementation. The reasons for implementation failure can be various, such as insufficient human resources, lacking stakeholder support or lack of administrative competencies.

A more sophisticated level of understanding of the programme logic can be gained by examining and modelling drivers of social change at the micro level. One may identify 'mechanisms' (Pawson & Tilley 1998) or 'mediators' (Weiss 1997, Dahler-Larsen 2001) through which the social changes are expected to occur. It should be noticed however that this type of theory-led approach operates at the micro level. One is dealing with behavioural mecha-

nisms which interventions can activate in target populations under certain conditions (context).

How should the programme logic be charted in practice? Patton proposes three ways to identify the programme theory (1997): *deductive*, *inductive*, and the *theory of action* approach. In the deductive approach, the programme theory is derived from academic research. This is the strategy most clearly elaborated by Chen (1990). He argues that a programme theory should be grounded in theories of the social sciences. This type of programme theory would be firmly supported by scientific studies about the theories of social change. The second approach is an inductive one, where the programme logic is formulated through an analysis of empirical data from the field. As a research-intensive method it needs to be carefully embedded in the overall evaluation design.

The third approach is that preferred by Patton's in his user-focused evaluation thinking. It aims at deriving the

programme logic from stakeholders' implicit models of thinking. It builds on the 'theory of action' approach, developed by Argyris and Schön (1978). The model has its theoretical roots in the cognitive psychology theory of social constructs (Kelly 1955). It holds that people base their actions on causal assumptions represented in cognitive schemes and maps. For the purposes of practice-oriented formative evaluation projects which aim at increasing the commitment and empowerment of stakeholders, the theory of action approach is well suited.

Despite lengthy discussions on the epistemological issues related to the programme logic and programme theory, not much material has thus far been available in respect of how a collaborative theory of action should *in practice* be formulated through a joint stakeholder effort. We will now turn to a case which presents a collaborative programme logic formulation process in a regional development programme.

Evaluation of a regional partnership programme in Finland

A collaborative theory of action approach was applied in a regional development programme in Finland. The target of the evaluation was a nationwide regional partnership development programme initiated by the national government in 2001. The initiative, labelled *The Regional Centre Development Programme* (hereafter the RCDP), aims at enhancing urban regional development in non-metropolitan urban regions. Thirty-five partnerships were selected as eligible in the programme on the basis of *a priori* criteria. One of the most important criteria was the extent to which the candidate regions represented functionally integrated local labour market areas. This aspect was emphasised since the programme builds on the idea of activating cooperation between various actors in city regions to uncover promising practices and procedures for networking and partnership. Local labour market areas are believed to constitute a territorial unit which has the potential to drive economic and social development if functionally integrated in a proper manner. The programme *rationale* stresses the importance of local priority-setting in the regions rather than top-down steering via national policy guidelines.

On a general level, the programme is designed along the basic tenets of a new regional development paradigm which has its theoretical roots in institutional economic theory (Amin 1999) and communicative planning theory (Healey 1992). Institutional theory sees a successful development process as "bottom-up, region-specific, longer-term and plural-actor based policy actions" (Amin 1999, 366). The role of knowledge creation, information ex-

change and learning is crucial. Ash Amin has stated the normative dimensions of this approach which include the following imperatives:

1. strengthen networks of association;
2. encourage voice, negotiation and the emergence of procedural and recursive rationalities of behaviour, in order to secure strategic vision, learning and adaptation;
3. mobilize a plurality of autonomous organizations;
4. enterprise support systems, political institutions and social citizenship; and
5. local path-dependencies. (Amin 1999, 368.)

The communicative approach to the regional planning emphasises the political and pluralistic base of the planning process. Accepting different interests and points of view should be the starting point for regional planning. Common deliberation and negotiation in an inter-subjective communication process is the means by which we may proceed in the direction of finding the common ground for regional strategies.

The Regional Centre Development Programme is clearly founded upon this new regional thinking. The programme is strongly structured along the basic principles of the new paradigm:

1. The territories selected in the programme are defined on a functional basis, not by administrative borders, structured by a network of work places, housing, com-

- muting, and services.
2. The programme planning stage has aimed at generating a regionally self-induced planning process by all relevant actors. This aims at mobilising local knowledge and strong engagement in the programme.
 3. An extensive discretion over the objectives has been left for the regions in a bottom-up strategy process.
 4. One of the programme's goals is to build effective networks both intra- and inter-regionally.
 5. An extensive partnership between the regional authorities, the municipalities, research and development organisations and the third sector is required in the implementation of the programme.
 6. Regional development has been defined holistically, balancing various aspects of development like business support, knowledge creation, public services, spatial planning, regional marketing and image construction etc.
 7. Strong emphasis is placed on citizen participation in the development process.

Due to the regional development paradigm emphasising the endogenous nature of local social and economic development through joined-up local partnerships, the programme logic was left relatively loose and open. No uniform national programme logic was employed by the government and the programme management. Instead a loose framework for local collaborations was given to work out and agree on relevant targets and measures. At the initial stage, the local goal setting and programming was of varied quality and consistency. In most of the regional cases, the precise operational definitions of the local partnerships thus remained unspecified and rather vague.

This fuzzy nature of the programme logic made it imperative for the evaluators to support the programme implementation by means of a programme logic clarification exercise at the local level. The first phase of the evaluation therefore involved constructing programme logic models for each of the 35 regional centres in collaboration with programme managers and stakeholders.

Programme logic construction

The first evaluation of the RCDP programme was carried out as a parallel process to programme implementation during the period 2001 to 2003. This formative evaluation was intended to support the programme management both at the national and regional levels through monitoring, facilitation and systematic data analysis. It provided external feedback for the evolving programme during its first years of implementation.

The theory of action approach was our point of departure in mapping the programme logic for two reasons. First, when evaluators do not participate in the programme planning phase, but enter the scene only after the programme has already begun, it seemed promising to map the managers' thinking from a theory of action perspective. The second reason is related to the nature of the RCDP programme itself. The elaboration of the programme theory requires clear objectives to begin with. In the RCDP however there is a strong emphasis on the bottom-up definition of objectives and context-dependence, which makes each regional project unique. The national programme goals are set on a relatively general level. The regions have been given extensive discretion over particular objectives. The national guidelines only emphasise process issues: networking, partnership, regional cooperation on a new level between the municipalities and functional urban regions, the actor's active engagement in the chosen strategy and the coordination of programmes and projects.

At the outset of the evaluation it became clear that the evaluation task requires an innovative approach to pro-

gramme evaluation. The programme follows a rather different logic than that of many previous regional development programmes in Finland, e.g. the European Union structural funds programmes. In these programmes, the wider objectives and priorities are defined at the European level and subsequently rolled down to encompass lower-level priorities on the national programme and priorities levels. This top-down logic has been criticised for being inattentive to local development needs and being at odds with the more contemporary understanding of the dynamics of regional development.

The programme logic and design however becomes increasingly complex and difficult to map. There is no one specific theory behind the RCDP programme, but instead 35 regional strategies for achieving the intended effects. Moreover, implementation cannot be located in one individual organisation, but the effects are expected to emerge from cooperation and interchange between different actors in the region. Thus clear programme logic is difficult to conceptualise. Weiss has also acknowledged that "when a program is very complex and multi-faceted [...] the evaluator may find it impossible to follow every branch or every sub-theory underlying the program" (1998, 65). Much of the work in theory-based evaluation has not been done on large public policy programmes, but in the fields of health promotion and risk prevention (Weiss 1997, 44), where the mechanisms may be more easily identifiable. The RCDP programme's complex characteristics draw the evaluators' attention to the theory of action approach, originally developed by Argyris & Schön (1978), and ap-

plied to programme evaluation by e.g. Patton (1997).

The regional strategies included some more specific ideas about how the positive effects are supposed to be attained. A number of “success factors” that characterise successful regions in the new economic context have been identified. The programme philosophy is based on the idea of promoting certain practices: (1) territorial focus, (2) partnerships, (3) networking, (4) engagement of stakeholders, and (5) civic engagement and social inclusion. By promoting these practices, the regions are expected to build up better competences for coping with the global knowledge-driven economy where regions are obliged to compete. However, these practices are hard to conceptualise exclusively in the linear cause-effect mode required by the programme theory approach. They seem to be of a different systemic nature. A further complication thus con-

cerns the mode of causality in the programme. In recent economic theorising, a positive regional development is seen as a product of systemic interplay and mutually reinforcing occurrence of different factors, of which the knowledge base is the most important. Not always are the causal linkages of a linear nature. Toulemonde has reviewed four types of causal relationships: *linear*, *circular*, *reflexive* and *irreversible* causality (1995). His examples show how all the modes of causality may be found in regional development contexts. He concludes however that circular or irreversible causality is the evaluator’s rule, while linear and reflexive causality only applies to a limited number of ‘happy exceptions’ (ibid, 188). The RCDP programme philosophy clearly builds on a notion of regional development as a systemic entity, where reflexive and irreversible causality is the dominant mode of causes and their effects.

Cognitive mapping as a collaborative process

Due to the complexities of the RCDP programme, a specific method for programme logic elaboration was employed. The cognitive mapping approach (Eden & Ackermann 1998) was applied as the method to identify programme managers’ and stakeholders’ theories of action. Cognitive mapping is an approach which has its background in the study of managerial and organisational cognition (Huff & Jenkins 2002). It aims at clarifying people’s theory of action in a dynamic and context-sensitive way. Cognitive researchers have noticed that people’s thinking is structured into cognitive maps which also guide their action (Eden & Ackermann 1998). The notion of cognitive maps is related to other conceptual models in cognitive psychology, such as mental models (Johnson-Laird 1983), cognitive schemas (Anderson 1978), frames (Rein & Schön 1993) and knowledge structures (Kelly 1955). In contrast to these more all-encompassing notions, the cognitive mapping approach emphasises assumptions about cause-effect relationships between actions, events, contexts and phenomena. They are particularly useful constructs for the purposes of studying intentional action. In contrast to *espoused theories*, which are explicitly stated, people’s thinking and acting is guided by *theories-in-use*, which usually remain implicit and do not usually become communicated as whole (Argyris & Schön 1978). Practitioners are seldom thoroughly aware of their own theories-in-use. That is why mental models need to be made explicit through a mapping exercise.

The cognitive mapping approach has been extensively applied by strategic management theorists and practitioners, especially because it allows us to detect emergent strategic thinking (Eden & Ackermann 1998). Emergent strat-

egies are fragments of strategic thinking in individual minds which need to be taken into common discussion. By building and negotiating consensus from diverse pieces of strategic thinking one may construct more coherent guidelines for effective future-oriented action which take into account the realities set by the context.

With the help of the cognitive mapping technique, one may draw *strategy maps* for a development programme like the RCDP. The strategy mapping allows us to detect the dynamic and interpersonally evolving conceptions of how the programme should be promoting societal change. The strategy here refers to a “systemic network of interconnected statements of strategic intent” (Eden & Ackermann 1998, 5). This definition of strategy points to the importance of negotiating and agreeing upon the strategic action. The conversation on the desirable directions for action is crucial here as is the process of establishing consensus and the refinement of that strategy.

Our approach to picturing a public programme is thus to map the intended programme strategy by cognitive mapping. In the case of Regional Centre Development Programme this seems to have certain advantages. The concept of strategy is extensively used in the field of Finnish regional policy. Actors are therefore familiar with regional strategy thinking. The development programmes are often based on strategic analyses of the strengths, weaknesses, threats and opportunities (*swot-analysis*). Along the theoretical advances in communicative planning thinking, the strategy is now seen mostly as a “soft strategy”, a forum for common deliberation, consensus building, and the mediation of various interests. This communicative process is expected to result in better understanding and engagement in the chosen strategy. Since the regional

development context is not hierarchically structured and there is no overarching regional level of strategy formation in Finland, but instead a multitude of different (and often

competing) strategies, their coordination is also needed. The communicative perspective tries to leave room for this coordination task.

Learning from the construction of programme logic

The cognitive mapping exercise consisted of a preliminary analysis and an evaluation workshop for stakeholders. The preliminary analysis was carried out by reading documents and interviewing the programme managers. This provided an initial understanding of the 35 regional sub-programmes. One day workshops were then arranged for all 35 participating regions. The workshop had two collaborative exercises on the agenda. The first was an interactive construction of a strategy map on the basis of the programming documents, participants' thinking and common deliberation. The strategies written in the programming documents formed the basis for the discussion, but the leading idea was to map the linkages between the different levels of programme outputs and outcomes. We used a computer-assisted working method¹ allowing quick documentation and modification of cognitive maps on the basis of general discussion. A draft map was continuously projected on a large screen, where all the participants could easily follow the modification process in real-time. The collective deliberations produced graphic presentations of the dominant strategy of the city region. Figure 2 presents the output of the Kokkola region mapping exercise as an example.

In the second phase of the workshop, the graphic strategy map was then transferred into the logical framework matrix². The different levels of programme strategy were discussed in this framework in greater detail. The logical framework approach contains a discussion of the specific indicators and sources of verification for each level of intervention, which makes the strategy more operational.

For evaluators, these two exercises provided a structured framework for assessing and understanding the different regions. It also provided a uniform approach enabling the comparison of the different strategies; while at the same time, allowing the regional characteristics to emerge without coercing the programmes into too rigid a framework.

As far as the programme managers and stakeholders were concerned, the experiences and feedback indicated that some level of learning *was* taking place. But what, more specifically, is it that actors can learn in this kind of

undertaking? When asked about the usefulness of the workshops, the participants pointed out a number of things many of which involve learning aspects.

First, the participants credited the workshops for helping them to attain a more holistic and coherent picture of their own thinking. The regional programme strategies are outputs of more or less unstructured planning processes which often do not involve very analytical deliberation on the expected chains of causes and effects. The cognitive mapping produces a graphically illustrated view of the programme strategy which may be more easily stored in the memory for future thinking and action. The coherence and plausibility of the programme strategy may thus be assessed critically and thereafter modified. What is learned in this process is an understanding of the 'big picture' in respect of the programme in its wider context.

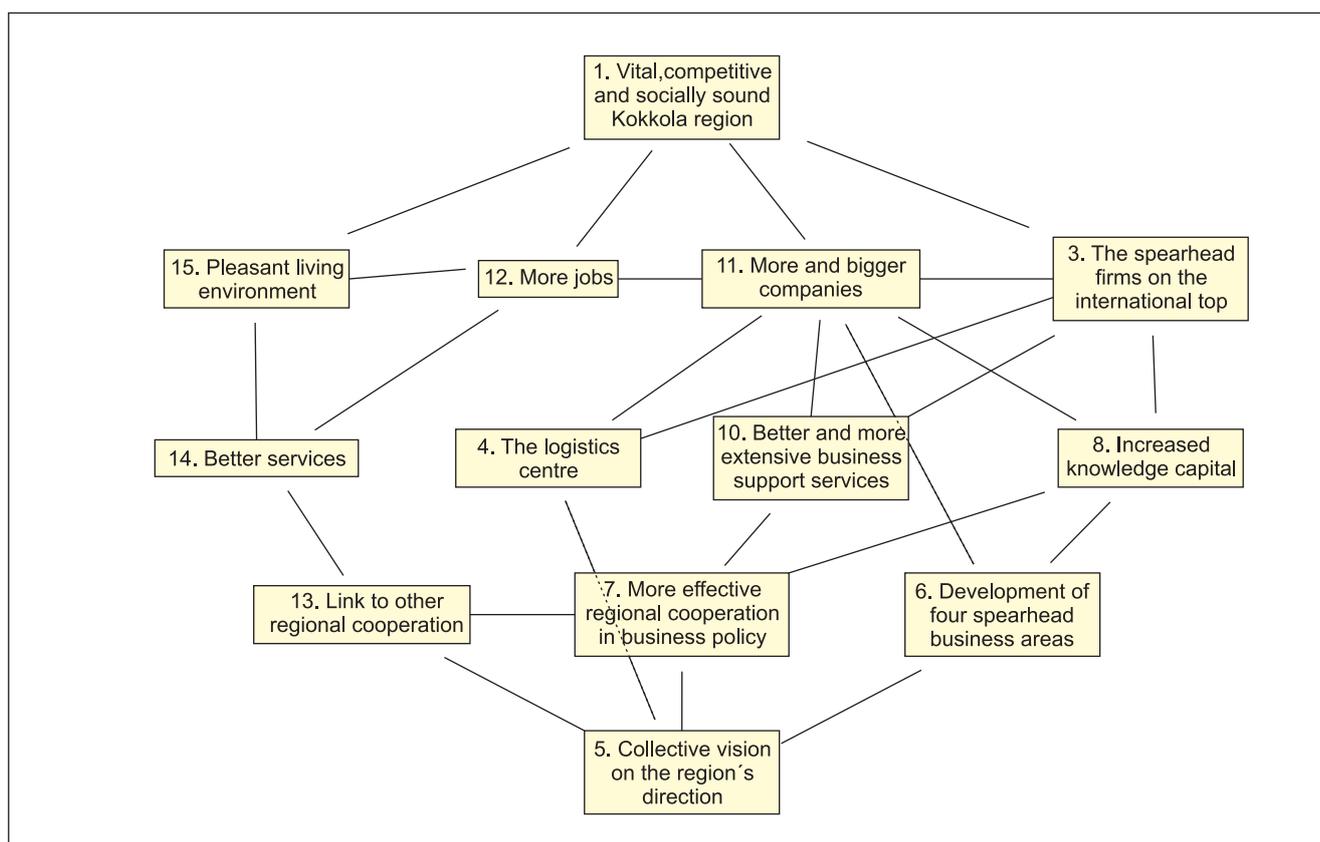
Second, the goals and objectives of the programmes were discussed, modified and fine-tuned during the workshops. The mapping process clarified various levels of goals and revealed some inconsistencies, thus providing the participants with a more structured model of the programme logic. In many cases, the participants soon realised that the goals were set at a far too ambitious level. The process of systematically reflecting the expected outcomes with the inputs and the planned measures revealed the lack of parity between the steps. Here the learning aspect is focused on the programme logic and on the plausibility of this logic. The modification of goals and objectives and increased understanding about them is close to what Peter J. May (1992) has called *social policy learning*. It entails, in contrast to *instrumental policy learning*, "lessons about the social construction of policy problems, the scope of policy, or policy goals" (332).

Third, the workshops seemed also to produce a particular kind of instrumental policy learning. Among the most important things that the mapping process seemed to reveal were the connections between the ultimate programme objectives and the chosen measures. The assumed paths between planned measures and effects on the regional level were considered on a more explicit basis. The identification of cause and effect relationships became easier, when the linkages between different levels of measures and outcomes were visually presented on a screen visible to the whole workshop group. It is however important to note that since the RCDP programme was only in its starting phase at the time that we should not expect the process to

¹ The computer programme used was Decision Explorer by Banxia Software.

² The logical Framework approach is a project planning and monitoring tool which has been widely applied especially in the field of foreign development aid. To learn more about the Logical Framework, consult Sartorius (1991) and European Commission (2001).

Figure 2. Kokkola region Strategy Map



produce knowledge or understanding of the causal links on the basis of empirical knowledge. The interconnected “if...then” propositions are, first and foremost, *assumptions* about how the world *will* behave when we exert our programme influence upon it.

Fourth, following the social constructionist views on the relations between thinking and action (e.g. Weick 1995), the mapping process enables the formation of more coherent mental models in different actors' thinking. The creation of shared mental models is one form of interpersonal learning. To some extent, the workshops provided an occasion for the social construction of the RCDP programme. It should again be noted however that the evaluators facilitating the participatory workshops did not, in the main, take the role of experts in regional development policy. The logic and internal coherence of the programme were the main concerns here, not influencing the substantial contents of the programme.

When comparing the different kinds of learning issues,

we may well assume that at this stage of the programme cycle, the effect of creating shared understanding between implementing parties would be the most important. The new regional thinking strongly emphasises the crucial importance of generating a certain level of consensus about the wanted directions. Without a shared understanding of the logic of the programme, it has more modest chances of causing visible effects. Successful implementation of this type of programmes is dependent on the engagement of actors in co-directed action, sharing and pooling resources and the commitment to multi-directional communication. Therefore, a shared understanding of the nature of the programme, its measures, intended effects and working logic is a necessary prerequisite for effectiveness. This is clearly a collective learning effect, which is not simply an aggregation of individual learning effects. Furthermore, it may not be called organisational learning, since the programme is only coordinated by single organisations but implemented by networks of actors.

Conclusions

This paper has tried to illustrate and discuss the need for the collaborative construction of programme logic and the benefits gained through increased learning. This task is

pertinent particularly in the early phases of a formative evaluation where expectations are set for the evaluation to contribute to the clarification of goals, outcomes and the

programme logic.

In the case study, the evaluation of a regional development programme in Finland, we adopted a theory of action perspective to programme logic construction. This was found to be a useful approach due to the complex nature of the programme at hand, which did not allow us to detect each sub-branch of the 35 regional programmes in order to identify the underlying social mechanisms of change. Instead, a more general level understanding of the strategy was established, which may provide a framework for more detailed, micro-level investigations about the mechanisms of change.

We may conclude that the cognitive mapping approach is an efficient method for identifying the programme thinking in a participatory evaluation setting, particularly in the early stages of a programme. It explicitly lets us map the causal assumptions and the context in which a programme is designed. It is thus a plausible prescription of how to construct programme logic in a participative setting, with a strong input into facilitating learning. Where the programme is too complex to be tracked down into the level of mechanisms of change, the programme theory may be pictured as a strategy map. This may be the case in complex programmes in complex contexts, like for instance the RCDP, where the causality issue may not be sensibly elaborated very far. There are many factors which make the identification of programme logic very difficult, such as the fuzzy borders of the programme, the non-linear causality context and the multitude of relevant implementing actors.

Whenever it is possible however, the cause-effect relationships should be elaborated in greater detail, by linking the programme theory to investigations on the mechanisms which drive the change process. Cognitive mapping may then be seen as a first step, which is succeeded by the process of tracking a detailed programme theory including sophisticated understandings about the mechanisms or mediators of change. The process may be continued in the context of a less participatory process of testing the causal assumptions with a more rigorous theory-driven evaluation approach.

The case study results suggested that some learning benefits may be gained from a collaborative process of drafting programme logic model, such as was done in the regional centres programme case. Four distinct modes of learning were identified: (1) getting a coherent picture of the whole programme, (2) learning about its goals and objectives, (3) understanding the causal assumptions, and (4) increasing shared understandings.

Cognitive mapping, as an instrument in evaluation, is a technique which can help promote the construction of programme strategies which are *internally* more consistent. In other words, irrespective of the truth-value of the content of the strategy and its ability to generate the desired outcomes, the programme logic may become more coherent in terms of drawing plausible connections between actions and their expected effects.

References

- Amin, Ash (1998) An Institutional Perspective on Regional Economic Development. *International Journal of Urban and Regional Research*. 23 (2), 365-378.
- Anderson, Richard (1978) Schema-Directed Processes in Language Comprehension. in Alan Lesgold, James Pellegrino, Sipke Fokkema, Robert Glaser (eds.): *Cognitive Psychology and Instruction*. Plenum, New York.
- Argyris, Chris – Donald Schön (1978) *Organizational Learning. A Theory of Action Perspective*. Addison-Wesley, Reading.
- Chen, Huey-Tsuh (1990) *Theory-Driven Evaluations*. Sage, Newbury Park.
- Dahler-Larsen, Peter (2001) From Programme Theory to Constructivism: On Tragic, Magic and Competing Programmes. *Evaluation* 7 (3), 331-349.
- Eden & Ackermann (1998) *Making Strategy: The Journey of Strategic Management*. Sage.
- European Commission (2001) *Manual: Project Cycle Management*.
- Forss, Kim – Claus C. Rebién – Jerker Carlsson (2002) Process Use of Evaluations. Types of Use that Precede Lessons Learned and Feedback. *Evaluation* 8 (1), 29-45.
- Healey, Patsy (1992) Planning Through Debate: The Communicative Turn in Planning Theory. *Town Planning Review* 63 (2), 143-162.
- Huff, Anne – Mark Jenkins (eds.) (2002) *Mapping Strategic Knowledge*. Sage, Trowbridge.
- Johnson-Laird, Phil (1983) *Mental Models. Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.
- Kelly, George (1955) *The Psychology of Personal Constructs*. Norton, New York.
- Kettner, Peter – Robert Moroney – Lawrence Martin (1999) *Designing and Managing Programs: An Effectiveness-Based Approach*. Sage, Thousand Oaks.
- May, Peter J. (1992) Policy Learning and Failure. *Journal of Public Policy* 12 (4), 331-354.
- Patton, Michael (1997) *Utilization-Focused Evaluation. The New Century Text*. 3rd edition. Sage, Thousand Oaks.
- Patton, Michael (1998): *Discovering Process Use*. *Evaluation* 4 (2), 225-233.
- Pawson, Ray – Nick Tilley (1998) *Realistic Evaluation*. Sage.
- Rein, Martin & Schön, Donald (1993) Reframing Policy Discourse. In Frank Fischer & John Forester (eds.): *The Argumentative Turn in Policy Analysis and Planning*. Duke University Press, London.
- Sartorius, Rolf (1991) The Logical Framework Approach to Project Design and Management. *Evaluation Practice* 12 (2), 139-147.
- Toulemonde, Jacques (1995) Should Evaluation Be Freed from its

Causal Links? An Answer Illustrated by European Economic Development Policies. *Evaluation and Program Planning* 18, 179-190.

Vartiainen, Perttu – Janne Antikainen (1998) Kaupunkiverkkotutkimus. The Ministry of Interior, Helsinki. (*A Study on City Networks*).

Vedung, Evert (1997) Public Policy and Programme Evaluation. Transaction Publishers.

Weick, Carl (1995) Sensemaking in Organizations. Sage, Thousand Oaks.

Weiss, Carol H. (1997) Theory-Based Evaluation: Past, Present, and Future. *New Directions for Evaluation* 76, 41-55.

Weiss, Carol H. (1998) Evaluation. Second edition. Prentice Hall.

The Impact Chain Model as an Analytical Tool for Evaluation within Research and Development

By Soile Kuitunen

Abstract

This article deals with questions relating to our ability to better assess the societal relevance and significance of public RD&I (research, technological development and innovation) interventions. Despite the rigorous efforts made since the 1960s evaluation practices, in this field, remain relatively poorly standardised. As such, many of the procedures developed lack the analytical power and clarity to be of value for strategic, development and learning purposes in the context of public RD&I. This also has implications for the relevance of RD&I as a source of inspiration for other policy areas, where otherwise fruitful cross-fertilisation may exist. Here the linkages between RD&I and regional development are particularly interesting in the context for this publication, as R&D and innovation have become increasingly central resources for regional development, while also being formally incorporated into the Structural Funds programme.

In this article, an attempt is made to develop an analytical and evaluative framework or tool for the assessment of societal impacts, and the relevance and significance of

public RD&I. This is done through the introduction of an *Impact Chain Model*. The model is based on the following characteristics: recognition of the multidimensionality of impacts, the differentiation of five levels at which impacts may come into existence, and the identification of the clusters of factors conditioning the emergence of impacts across all these levels. As the proposed framework is based on impact-chain thinking, as well as the cognitive tool suggested by it, this is of broader relevance and can be applied to the assessment of various types of interventions, with regional interventions included, in which RD&I plays a role.

The article concludes that by using the Impact Chain Model and by integrating the strategic, evaluative and foresight approaches implied by it, evaluation studies can strengthen their explanatory potential and attain a role beyond that of its previous justification-based nature. By developing a credible Impact Chain Model, evaluative knowledge can become a source of opportunity, innovation and competitive advantage for the interventions under scrutiny.

Introduction

RD&I (research, technological development and innovation) is not a new concept in the evaluation context. Evaluation studies focusing on this type of activity date back to the 1960s when the first such studies were produced¹.

¹ It should be noted that by early 2000, the focus of R&D evaluations had become larger as implied by the introduction of the new term R&D&I. This concept has been adopted in central policy documents such as the Community Framework for state aid for research, development and innovation (CEC 2006).

Originally, the major aim of such studies was to determine the quality of research, although at the same time attempts were also made to identify and assess some of the more complex issues, including the impacts and benefits R&D interventions potentially generate for society. The major aim of the dominant framework that subsequently emerged was however to demonstrate the additional value of public R&D on the R&D activities of private compa-

nies. The framework based on the concept of *additionality* understood in this manner was – and remains – one of the key determinants of the fact that the public R&D interventions produce (additional) effects that would not have taken place without the intervention. (Luukkonen 2004)

Despite the embeddedness of the concept of *additionality* in RD&I evaluation, the concept itself displays different facets depending on the context in which it is deployed. In relation to the Structural Funds, the notion of *additionality* at base simply reflects the requirement that contributions from the Structural Funds shall not replace public or equivalent structural expenditure by a Member State, though when considered more broadly, in this context *additionality* also includes potential behavioural and substantive dimensions. These dimensions are however occasionally distinguished from *additionality* by linking this to added financial resources, while the behavioural or substantive aspects are referred to as ‘community value added’. In the other contexts, such as that of the EU Framework Programmes, the concept of *additionality* is defined in somewhat different terms, as will be illustrated in the section 2.

Despite the inevitable advantages the framework based upon *additionality* has in focussing on the relevance and necessity of interventions financed with public resources, it has nevertheless been shown to provide only a limited perspective into R&D evaluation. It does not deal particularly well with questions relating to overall societal merit, worth and value of public R&D² – all issues which are of increasing importance in the prevailing context of the need for an evidence-based understanding of the rationales of public intervention and their contribution to overall welfare. A number of additional questions emerge relating for instance to whether the goals of public R&D intervention cover the most relevant and topical societal problems in the field of R&D and how are these may be prioritised; what kinds of intended and unintended impacts do R&D interventions produce on society as a whole and in by what means are they likely to come into existence; who are the immediate and final beneficiaries; and how are the costs and benefits of the intervention to be distributed in society.

The increasing demands on determining the societal impact and overall relevance of public R&D is, in the main, to be explained by the more general shift in the understanding of public action and management. Since the 1980s public sector reform has been implemented across all of the OECD countries (Pollit & Bouckaert 2004) thus paving the way for the evolution of a new paradigm, “management by results” (NPM) (and consequently, for its critics). Accordingly, the focus of accountability has been extended from demonstrating how much public money was spent, and for what purposes, to demonstrating that the

² The definition of evaluation as a process of determining the merit, worth, or value of something, or the product of that process, was originally put forward by Scriven (1991, 139).

resources were *efficiently* deployed and that they contributed to the emergence of societal benefits.

The evolution of R&D makes it apparent that numerous challenges remain to be tackled by the professional evaluators³. We have already witnessed a genuine “evaluation boom” fuelled by developments in the public sector. It is now worth asking whether the measures put forward thus far remain capable only of capturing the more immediate outputs and outcomes of public R&D interventions while failing to grasp the deeper underlying procedures by which societal impacts emerge. Increasing methodological interest has thus been targeted at the mechanisms conditioning the emergence of impacts, as well as at the issues linked to the societal relevance and significance of public R&D interventions more generally. It does however seem highly questionable whether the currently available evaluation studies methodologies are actually capable of feeding into the strategic development processes accompanying public interventions, and, whether indeed they contribute to mutual learning among stakeholders and evaluators.

This article seeks to shed light on the development needs of R&D evaluation methodologies, as well as on what implications they may have for evaluation exercises more generally, with special reference to the EU Structural Funds. The main argument here is that R&D evaluation can, at its best, contribute not only to the assessment of accountability and justification of public support, but also to R&D policy development, debate, and organisational learning. By undertaking evaluations, public R&D organisations are provided with the opportunity to prioritise and address the broader issues on which their activities can make the largest impact. In the increasingly intense competition among private knowledge producers and between them and public research organisations there is a more pressing need to establish the best possible basis upon which strategic decisions can be made relating to the competitive advantages of organisations in the innovation *milieu*.

The article is structured as follows. It begins with a brief description of the history of R&D evaluation and continues by focusing further on the evolution of the concept of *additionality* within R&D evaluation. In section three the most prominent tendencies transforming public intervention are discussed and further elaborated by focusing on their implications for R&D evaluation in practice. In section four the *Impact Chain Model* is presented. This new analytical tool addresses some of the key aspects of R&D evaluation. Section five summarises the findings and proposes a number of critical considerations worth taking into account in future evaluation exercises.

³ It should however be acknowledged that several studies which have contributed to the overall development of evaluation methodologies and approaches have already been undertaken. Among the most prominent contributors here have been the researchers from VTT Innovation Studies (formerly VTT Technology Studies) who have carried out a large number of theoretical and empirical studies on the societal impacts of the R&D since the late 1990s.

The 'state of the art' in R&D evaluation: what has been assessed and how?

The assessment of R&D is not a novelty in the evaluation field, indeed, such activity dates back to the 1960s, in both the USA and in Europe, with the Nordic countries also being among the first to complete R&D evaluation exercises. The first evaluation studies carried out in the Nordic countries were focused on assessing the quality of research. At the same time, however, attempts were made to determine the socio-economic benefits and impacts of research and technology initiatives, in the Nordic countries and across Europe (Luukkonen 2002, 82).

Two evaluation studies surfaced as the archetypes for the emergence and development of the R&D evaluation scene (ibid, 82). The first study was a large-scale evaluation of a broad research programme called the *Alvey Programme* (Alvey programme for Advanced Information Technology) in the UK in 1980s. The second important study was the assessment of the *Eureka Initiative*. Both of these evaluations were carried out during the period from the mid-1980s to the 1990s, and were the first to try to assess the expected benefits for their participants (Georghiou 1999).

In the 1980s, the impact studies of R&D programmes also started to incorporate the notion of *additionality* in assessing the worth and value of public support to technological advancement in companies. It had clear and evident links to the concepts of impacts and effectiveness due to the fact that all of the various impacts of an initiative in reality constitute the *additionality* it has brought about. (Luukkonen 2000, 713).

The concept of *additionality* quite rapidly developed into a dominant framework offered for the evaluation of RTD programmes within the European Union, providing an expedient political rationale for public programme support, not only supporting company RTD but also RTD at the EU level in general (ibid, 711-712). The background to *additionality* stemmed from the discussion that developed around the notion of market failure. It was shown that left to themselves private companies were reluctant to invest sufficiently in innovative activities (Metcalf & Georghiou 1997). Moreover, such firms were often unable to absorb all of the benefits that arise from innovative activity (Arrow 1962; Nelson 1959). Due to the uncertainty, imperfect monitoring and intellectual property rights related to R&D, competitive markets were unlikely to result in the attainment of a socially optimal level of R&D investment.

In practice, as the notion of *additionality* highlighted, the costs of transferring and exploiting scientific and technological knowledge are so high that they affect the success or failure with which such knowledge can be utilised. This makes much of the knowledge a private good. (Mowery 1983, see also Luukkonen, op.cit., 712) As a result, there is a need for public intervention which supports not only the

creation of scientific and technological knowledge, but also the transfer and dispersion of this knowledge through collaborative networks and projects (Mowery 1994). From here it is not a great step to acknowledge that public funds, where they can be made to work efficiently, can encourage otherwise reluctant companies to collaborate in R&D activities.

Additionality was expected to bridge the difference between the presumed under-investment in RTD by firms and the actual joint investment by firms and public agencies in RTD prompted by public programmes. Considerations of *additionality* however reflected a fear that public support would simply act as a substitute for indigenous R&D investments made by the companies themselves. In its simplest form, *additionality* was assessed by the difference the government-sponsored programmes had made on their recipients, mostly private firms, in terms of R&D activities (Luukkonen 2000, 711). Evaluation studies were, and still are, thus expected to prove that public intervention produces additional effects that would not have taken place without the intervention.

The major questions relating to *additionality* relate to the types of changes that occurred as a consequence of the disbursement of R&D subsidies, the presumption being that public money supporting R&D activities affects the content, scale and results of such activity. In the 1990s, a more refined conception of *additionality* was put forward by Georghiou, who distinguished between three types of *additionality*: *behavioural*, *output* and *overall additionality*. *Behavioural additionality* refers to changes in the scale or scope of R&D projects. *Output additionality* highlights the fact that the outcomes and end results of R&D projects are different due to the subsidy. and *overall additionality* can be seen for example in the EU Framework Programmes, where the main question concerning overall *additionality* is whether R&D projects would have been carried out at all without the support drawn from the Framework Programmes.

Based on the elaboration of these analytical distinctions, a more highly standardised approach has emerged encompassing the following questions (see, e.g. Georghiou et al. 1993; Kuitunen & Oksanen 2002; Luukkonen & Hälikkä 2000; Luukkonen & Niskanen 1998):

- whether the R&D would have been carried out without EU funding
- whether public funds enabled the project to be conducted at a faster pace, sooner, on a larger scale, or differently
- whether the project would have been carried out at all without public funding.

Luukkonen subsequently (2000) developed the framework of *additionality* further by adding a strategic dimension to it. This model resulted in a cross-tabulation in which the categories of the ideal, substitution, trivial and marginal cases were identified. Other elaborations of the notion of *additionality* have been developed within the Structural Funds contexts, similar to the strategic and behavioural aspects referred to above. Bachtler *et al.* (2003) for instance have distinguished five dimensions of *additionality* or “added value”, which as a term is more common in the EU Structural Funds context:

- Political added value: e.g. improving the visibility of the intervention, as the result of which support for it may increase
- Policy added value: strengthening strategic thinking (an integrated, coherent point of view)
- Operational added value: benefits gained from the partnership (vertical coherence, greater inclusion of local actors etc.)
- Learning added value: institutionalisation of the learning reflex, promoting the dynamics of learning

and innovations

- Cohesion added value: e.g. increases in the regional economic development resources, economic impacts in the regions more generally.

At a general level, as indicated by the examples given above, the framework of *additionality* comes very close to providing the basis for a multi-criteria evaluation, as both aim at assessing impacts according to more or less fixed criteria based either on a descriptive or a prescriptive approach (see, for example, Shadish *et al.* 1995)⁴. The added value categories may thus be defined as impact categories. One of the most evident shortcomings in this framework is however that it leaves very important questions concerning the mechanisms and causal paths between intervention and its outcomes (whether these are immediate, intermediate or ultimate outcomes) unexplored. Furthermore, it also pays little or no attention to the levels at which the impacts can come into being. We will return to these questions in section four when we present the *Impact Chain Model* and its attributes.

The transformation of public intervention rationales and their implications for R&D evaluation

Recognising the need for contextualisation, this section provides us with a brief overview of the most prevalent tendencies in the transition of public action and governance. This is particularly important as across Europe, Finland being no exception, the reforms undertaken in the public sector have produced several managerial and administrative challenges, which not only accentuate the need for evaluative information, but also imply the need to incorporate new kinds of criteria into the evaluation exercises themselves. These modifications inevitably also influence the means and implementation of technology and innovation policies. In what follows these trends and their implications are discussed, by reference to the ongoing theoretical debate and to experience gathered through the execution of various R&D evaluations in the Finnish context.

As noted by several authors, public sector reforms have swept over all OECD countries particularly since the 1980s (see, for example, Pollitt and Bouckaert 2000, 2004). In Finland, in the 1980s a dominant doctrine emerged emphasising management-by-results, replacing the old doctrine with its supply-side focus on resources rather than the demand-led focus on the results and outcomes of pub-

⁴ In the prescriptive approach, it is the evaluators who select the criteria, while the criteria themselves are based on goals or needs stated by others, e.g. by legislators or stakeholders in descriptive approach.

lic action. The “*Zeitgeist*” of the old system relied on the fallacy that there were no limits to the amount of public resources available. Retrospectively it can be seen that no explicit requirements to analyse the confines of growth existed. In the late 1980s, however, the once iron-cast legitimacy of the public sector markedly declined. In the 1990s, mostly due to the severe economic downturn, public sector budgets increasingly come under pressure and a new paradigm based upon the notion of management-by-results was adopted. (Virtanen & Wennberg 2005.)

The modern management-by-results rationale has shifted the focus from inputs and resources to outputs and outcomes; it has also implied fundamental changes in the ways they are measured. Budget procedures were transformed and the quality of public services, as perceived by citizen-customers, increasingly became the major principle by which the organisation and delivery of public services was assessed. In Finland the adoption of a new paradigm resulted in significant changes such as a sharp decline in the actual number of public servants through the creation of state enterprises and later in some cases State-Owned Companies; the introduction of framework budgeting and results-oriented budgeting; restructuring central agencies and shifting responsibilities to the regional and especially the municipal level; and finally also the reform

of human resource management including the salary system (Pollitt & Bouckaert 2004, 241-242).

In reality, despite the many positive impulses the management by results paradigm bestowed on the process of change in the public sector, it quite quickly ran out of steam. Among the most intractable problems encountered by public authorities was that of identifying adequate and reliable result-indicators, rendering the targets more or less artificial or random. In addition, weak or even non-existent links existed between individual policy measures and the broader targets concerning the effectiveness of the public interventions. Much criticism was also targeted at the quality of the public service providers and their ability to meet the demands of the citizen-consumers. The more profound structural problems could generally be related to the stagnant governance structures that were not easy to consolidate with the extensive targets set for public intervention. As a result of this clear level of discontent with the system as it was evolving the desire to introduce more transparent *governance* models to replace the traditional and hierarchical *government* model emerged. The need to broaden the basis for the assessment of productivity and profitability, and strengthen the customer orientation of public intervention was also stressed. It has thus become increasingly important to acknowledge and satisfy the continuously expanding demands of the citizens. (Virtanen & Wennberg 2005, 29-30)

As a response to this criticism a new *process-oriented paradigm* has evolved, striving to better serve the needs of the citizens, and taking into account the requirements of the effectiveness of public action. A model based on a process approach and wider efficacy rather than narrowly defined result and productivity targets is thus now establishing itself as *the* paradigm for public governance and management in the 2000s. The effects and wider efficacy that public organisations create through their activities are becoming essential for the materialisation of their accountability and for the critical evaluation of those activities. In this process increasing importance is now being placed on the significance, benefits and effects of public organisation interventions in society.⁵

It is evident that many of the general reforms of the public sector have also affected technology and innovation policy and its administration. The changing nature and context of policy is also reflected in a renewed governance model. Technology policy, both in Finland and elsewhere,

⁵ An increasingly active cluster of evaluative research has emerged in Finland around this problematic; see for instance Kuitunen & Hyytinen 2004; Lähteenmäki-Smith et al. 2006; Lähteenmäki-Smith & Hyytinen 2006; Kutinlahti et al. 2006; Hyytinen & Konttinen 2006.

is thus perceived as being altered from its traditional closed, highly hierarchical decision-making structures into a gradually more open and flexible system. One of the main dimensions of change here concerns the emergence of horizontal networks, which reflect a blurring of boundaries between the public and private spheres and a restructuring of the interactions between them. (See for example, Caracostas & Muldur, 1997 and Edler *et al.* 2003).

Within the emerging system, technology issues are perceived as no longer remaining wholly under the purview of a closed and internally cohesive technology elite. This can be seen as being part of a broader development encapsulated in the notion of moving from government to governance. The new governance system, as opposed to traditional systems of government⁶, implies that decisions concerning technology and innovation are initiated, prepared and decided upon by a larger group of actors representing a broad array of socio-political sectors, incorporating even the political sphere with popularly elected politicians at its core.

In addition to this re-structuring of the traditional forms of power and political competence, the transition towards horizontal public policy action in general and innovation policy in particular, implies a change in the rationales and objectives of these policies. With reference to technology policy in particular, setting goals for economic growth and prosperity is increasingly also coupled with the broader aims of producing well-being as a public good in a manner that meets the criteria of good governance. Though the notion of 'good governance' is in itself a highly contested one, it is most often conceived of as being based on selected aspects of democracy, most often incorporating the dimension of participation, accountability, and efficiency (CEC 2001; Tiihonen 2004).

The ongoing development of the public management reform process was a key factor driving the renewal of evaluation procedures, with R&D evaluation being no exception. From the R&D evaluation viewpoint, new kinds of criteria had to be incorporated into the evaluation, particularly in relation to democracy and the other building blocks of the new mode of public action. To put it in more concrete terms, in addition to the impact criterion itself, equity, equality, transparency, and flexibility were also now referred to in determining the significance and societal relevance of public R&D. Table 1 summarises the criteria. These criteria along with their methodological aspects will be discussed and further elaborated in Section 4.

⁶ For an overview of the contents of the two concepts, government and governance, see, for example, Tiihonen (2004) and Kjær (2004).

Table 1. Criteria used in R&D evaluation (modified from the classification developed by Mickwitz 2006).

Criteria	Evaluation questions	Evaluation questions from the perspective of the R&D programme
General criteria		
Relevance	- Do the goals of the instrument cover all the key problems in the sector?	- Do the goals of the R&D programme cover all the key problems in the R&D sector?
Sustainability	Do the effects have a lasting influence on the state of the problem?	- Do the effects generated by the R&D have a lasting influence on the state of the problem?
Flexibility	- Can the policy instrument cope with changing conditions?	- Can the R&D intervention cope with changing conditions?
Persistence	- Are the impacts persistent in such a way that they have a lasting effect?	- Are the impacts generated by the R&D intervention persistent in such a way that they have a lasting effect?
Predictability	- Are the outputs and outcomes of the policy instrument possible to foresee? - Is it thus possible for those regulated, as well as others, to prepare and take into account, the policy instrument and its implication?	- Are the outputs and outcomes of the R&D intervention possible to foresee? - Is it thus possible for those regulated, as well as others, to prepare and take into account, the R&D intervention and its implications?
Criteria related to democracy		
Legitimacy	- To what extent is the policy instrument accepted by individuals and organisations, such as NGOs, interest organisations and firms?	- To what extent is the R&D intervention accepted by the stakeholders, such as firms, intermediating organisations, knowledge producers, financiers, and interest organisations?
Transparency	- To what extent are the outputs and outcomes of the policy instrument, as well as the processes used in its implementation, observable for outsiders?	- To what extent are the outputs and outcomes of the R&D intervention, as well as the processes used in its implementation, observable for outsiders?
Equity	- How are the outcomes and costs of the policy instrument distributed?	- How are the outcomes and costs of the R&D intervention distributed?

The Impact Chain Model as an analytical tool in R&D evaluation

As noted previously, it is clear that frameworks such as *additionality* are useful in assessing the impacts generated by public R&D. It is equally true however that they provide only a limited perspective into R&D evaluation particularly from the viewpoint of addressing the overall societal significance and relevance of public interventions. The concept of *additionality* thus leaves a number of questions unanswered, including:

- Are the targets of the policy intervention relevant – does the instrument cover the key problems of technology and innovation policy? Are these the issues that matter most to the intervention in question or are these the ones on which the intervention is likely to make the largest impact?
- What are the anticipated and unanticipated, as well as the beneficial and detrimental, effects of the policy being implemented? Are the effects persistent enough to have a lasting impact?
- How are the outcomes and costs of the R&D policy instrument distributed in society (according to such criteria as business sectors, size of the firms, different societal groups)?
- What are the mechanisms which either enhance or impede the emergence of impacts? What kinds of interventions and actions are needed to support the emergence of positive impacts and reduce the negative ones?

The complexities related to public R&D as an activity, moreover, pose particular challenges for its evaluation. R&D is a highly intricate activity; with its complexity further asserted by the fact that according to the prevailing understanding of R&D, it is being performed in networks of actors, each of which display some degree of autonomy, while at the same time all are related to, and interpenetrate, each other. Each actor possesses his or her own perceptions, values and interests which may be contradictory or even opposed to each other. (Stankiewicz 1992) These networks may come into existence at all levels of the innovation systems and *milieus* starting from the organisational level up to that of the level of global innovation systems and *milieus*.

Taking the complexities of R&D activity into account, it becomes apparent that there is a need for a framework which identifies the various levels and mechanisms related to the emergence of the impacts of public interventions. What is more, to become a useful tool to be deployed in the strategic development of the intervention, the strategic, evaluative and foresight approaches and methodologies which are consistent with them should be integrated.

In the following sections, we set up an evaluation framework based on impact chain thinking which contributes to the assessment of the societal relevance and significance of the R&D. The major characteristics of impact chain thinking can be summarised as follows (Table 2).

It should be acknowledged that impact chain thinking merely presents a *model* or an analytical tool rather than a *theory* aiming to examine such questions as how public interventions are constituted or their universal logic⁷. As Table 2 reveals, the Impact Chain Model nevertheless looks at the causalities inherent in the intervention by tracing the links between independent and dependent factors.

One of the key aspects in impact chain thinking is the notion of the multidimensionality of impacts. Given their inherent complexity public R&D policies seldom turn out exactly as intended. The effects and impacts of policies, even when extensively planned, are often unanticipated thus highlighting the importance of incorporating the foresight approach into the evaluation design. Furthermore, the effects public R&D generates may also occur outside the technology and innovation policy domain, even while having significant effects on it. The tendency towards the broadening of the scope and means of public policy, resulting in the replacement of formerly isolated policy sectors with rather more horizontal policy frameworks, accentuates the need for evaluating the inter-linkages between R&D interventions and other policy instruments⁸.

Furthermore the effects can vary in terms of their temporary frequency, meaning that some of the impacts may be temporary whereas others have a more lasting character. What is more, the effects may be primary or less important (secondary) as well as beneficial, detrimental or neutral in nature. This multidimensionality of impacts is illustrated in the figure 3. below.

The *Impact Chain Model* has many advantages with respect to strategic development and learning processes. By identifying a wide variety of impacts, it helps in prioritizing those issues that are most likely to have the largest impact on society though it is evident that evaluation studies are seldom, if ever, able to answer fully questions dealing with extremely complex issues such as RD&I. Prioritization is nevertheless essential due to the fact that an organisation (or intervention) cannot address all the societal

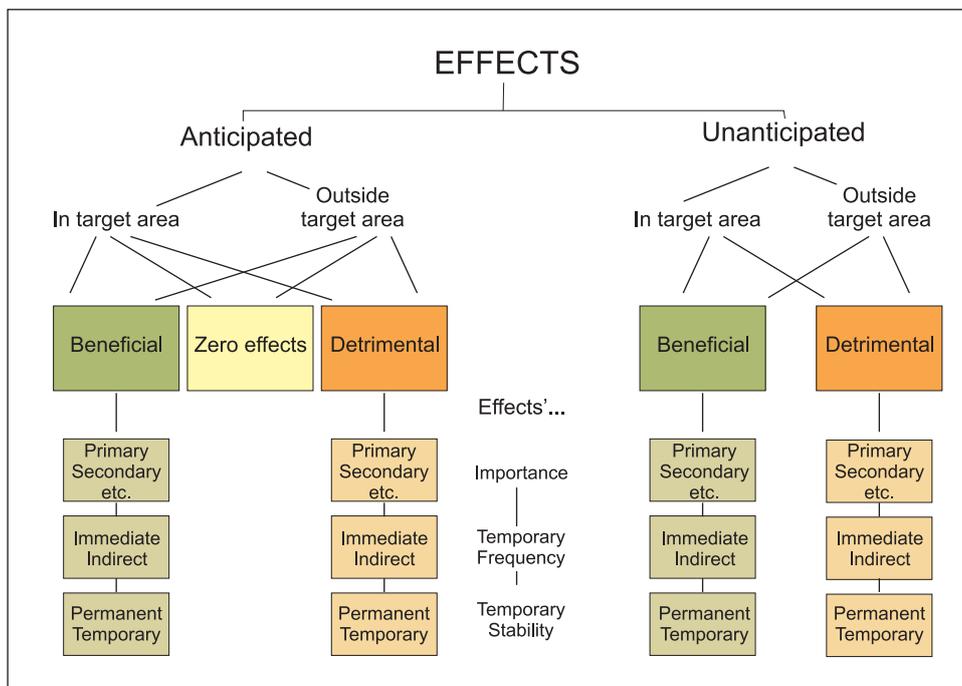
⁷ There are of course a multitude of different understandings as to what constitutes a theory in the social sciences. Noro has provided an interesting division between three types of sociological theories: general theories, research theories and 'zeitdiagnose' (for a more detailed discussion on these, see Noro 2000, 324).

⁸ The tendencies towards a more horizontal innovation policy have been investigated in the Finnish context by Kuitunen and Lähteenmäki-Smith (2006a, 2006b).

Table 2. Characteristics and attributes of the chain of impacts thinking.

Theoretical underpinnings	<ul style="list-style-type: none"> - Based on the realist school of evaluation, inspired by the work of such authors as Pawson and Tilley - Not a theory, but merely an outline of an evaluative framework
Aims and targets	<ul style="list-style-type: none"> - Acknowledges the multidimensionality of impacts - Identifies the mechanisms and factors impeding and enhancing the emergence of impacts - Differentiates between the various levels at which impacts come into existence - Traces causal links and chains between independent and dependent factors - Helps in prioritising the societal issues R&D intervention addresses and/or the ones on which the intervention can make the largest impact
Approaches and methodologies	<ul style="list-style-type: none"> - Due to the long timeframes within which the impacts of societal interventions usually emerge, impact chain thinking incorporates foresight and strategic and evaluative elements, and suggests a methodology which is in accordance with these approaches - Besides monitoring the social impacts, impact chain thinking focuses on identifying the evolving social effects which are likely to come into existence in the future and/or which are assumed to be the most important
Overall characteristics	<ul style="list-style-type: none"> - Universally applicable and may be deployed irrespective of the policy sector or activity the evaluation is focused on

Figure 3. The multidimensionality of impacts⁹



⁹ Modified on the basis of classifications put forward by Mickwitz 2006 and Virtanen 2007

needs or problems even within the boundaries of its core activities, such as R&D. On the contrary, there is significant demand to select issues that intersect with the core activities and are in accordance with the strategic goals of the intervention.

The Impact Chain Model suggests the importance of categorizing social issues into three major classes, namely, generic social issues, value chain social impacts, and the social dimensions of the competitive context (see, Porter & Kramer, 2006). Despite the fact that these categories were originally developed within the context of private companies, they are advantageous also in identifying and prioritizing issues and impacts within the public sector. The usefulness of the differentiation between these three types of social issues is further stressed by the intensified competition between knowledge producers within the public and private spheres as well as between them.

Generic social issues are those not significantly affected by public R&D interventions and those that generally do not materially affect competitiveness in the long term. These issues can therefore also be considered as marginal. The value-chain social impacts, in contrast, are those that are significantly affected by public R&D. These can then be perceived as important issues that need to be addressed in the everyday activities of public R&D interventions. The final category, the social dimensions of the competitive context, refers to issues in the external environment that significantly affect the underlying drivers of public R&D, thus providing rationales for refinements in strategy or even for a redirection of core activities. It is evident that both the *value-chain social impacts* and the *social dimensions of the competitive context* are salient for public R&D interventions, conditioning their legitimacy and thus constituting the means to hold these interventions accountable.

To concretise the chain of impacts we have developed a tool labelled *Impact Map* in which the categories of impacts as well as the mechanisms conditioning them are both identified (Figure 4). The picture illustrates the find-

ings of an evaluation study recently completed in the public health sector showing the five levels at which the impacts may emerge alongside the four impact categories applicable to other policy sectors. The impact map also addresses the fact that besides the four categories of factors affecting the emergence of impacts, there are other mechanisms that should also be taken into account, namely *duplication* and *embed*. Without the wider utilisation (embed) and diffusion (duplication) of the developed solutions, technologies and such, the benefits of the R&D intervention are inclined to be reduced to local and highly incremental amendments, which are hardly significant in contributing to the increase in overall welfare in society.

In the impact map, four categories of factors conditioning the emergence of impacts are identified: these are the factors related to technology; power and decision-making structures; cognitive and socio-cultural factors, and factors related to legislation. The second feature of the map is that the impacts are classified into four types of changes: changes in the intellectual capacity, processes, customers, and structures of organisation all being involved in the R&D.

The impact map implicitly recognises the fact that in order to serve the wider development needs of the public R&D intervention under scrutiny, methodological and theoretical triangulation is required. In our view, this implies the integration of the evaluative, foresight and strategic approaches and methodologies that are consistent with that approach. We have illustrated the main features of the approaches as well as the core questions related to impact evaluation as follows:

It should be stressed that all these approaches acknowledge the changes and overall trends in the innovation *milieu* and policy. The methodologies proposed are consistent with the evaluation questions while the whole model, encompassing the three approaches, suggests a methodological triangulation. In practice this implies the employment of quantitative and qualitative methods and empirical data sources.

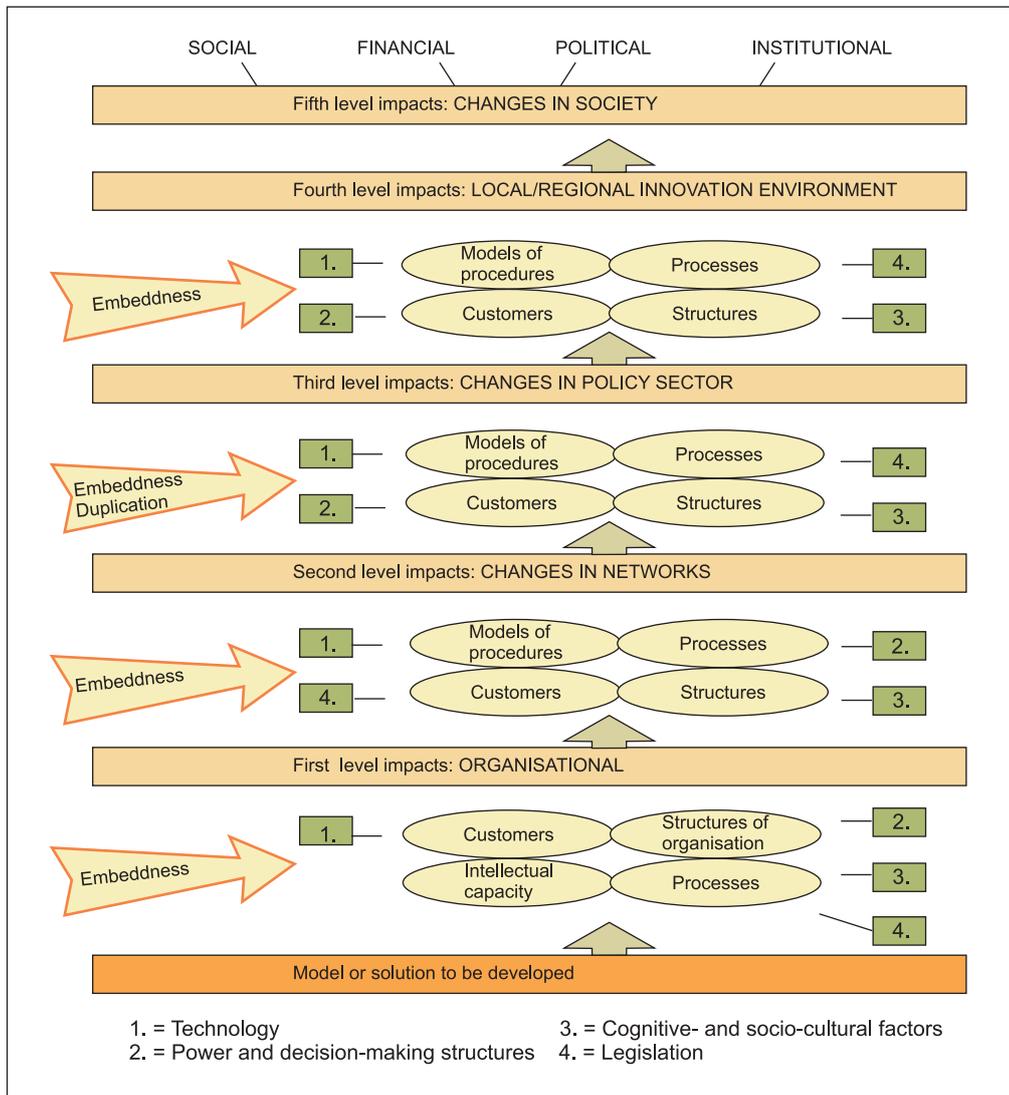
Concluding discussion

Progress on the development and application of methodologies for assessing public R&D has been continuous, both in Europe and the USA, particularly since the 1980s. A dominant framework has evolved based upon the notion of *additionality*, though the notion itself has been profoundly modified especially in recent years. While the first attempts to measure the *additionality* or 'additional value' of public R&D primarily concentrated on demonstrating the changes resulting from the implementation of the intervention, more recent approaches have been extended in

scope to cover such issues as societal, political, policy, and learning 'added value'.

Although an important aspect of assessing public RD&I, the framework of *additionality* has proven inadequate in addressing the societal relevance and significance of RD&I. In an endeavour to shed light on the societal changes and benefits, evaluation exercises should thus pay greater attention not only to anticipated effects but also to unanticipated effects, as well as those more potential impacts, likely to evolve in the future. Furthermore, the iden-

Figure 4. The impact map as a tool to concretise the chain of impacts ¹⁰.



¹⁰ The impact map illustrated in figure 2 was presented in the evaluation study of the FinnWell Programme (Tekes 2007).

tification of the mechanisms conditioning the emergence of impacts is equally important to incorporate into the evaluation exercise.

The profound transition of the public mode of action is embodied particularly on the adoption of the new paradigm based upon the notion of effectiveness and the societal relevance of public interventions which has reinforced the need for broadening the scope and criteria of RD&I evaluation, thus emphasising the employment of criteria related to the functioning of democracy. In more concrete terms, this implies the incorporation of such criteria as acceptability, transparency, equity and equality in addition to the more traditional evaluation criteria dealing with efficiency and effectiveness in the evaluation framework. These criteria are of importance in the evaluation of all kinds of public interventions supporting RD&I irrespective of the level at which they are implemented, which may be local, regional, national or global.

One of the key deficiencies of current RD&I evaluation procedures appears to be their detachment from the core business operations and strategies of public RD&I interventions, thus resulting in narrowed opportunities for strategic development and learning. In this paper, we have presented a tool that aims to lessen the distance between evaluation practice and the strategies and development of the public RD&I interventions. This model, namely, the “Impact Chain Scheme” is universally applicable, and can therefore be employed in *any* policy sector irrespective of policy issues, or of the level at which an intervention is launched.

The model utilises impact maps as a means to demonstrate the multi-dimensionality of the impacts as well as the interrelatedness of the factors conditioning the emergence of impacts. It builds upon the notion that it is neither rational nor in the interest of the entire society to strive to achieve all of the possible societal benefits through

Table 3. Main features of the strategic, evaluative and foresight approaches in R&D evaluation.

Strategic approach	
Fundamental questions	Why is R&D conducted? Why the intervention exists: what is the key societal need or problem the intervention tries to solve?
Questions related to impact assessment	What kinds of impacts are anticipated to emerge which are in accordance with the strategic goals? What must be done in order to produce (positive) impacts? By which mechanisms is intervention assumed to produce impacts?
Data	Who are the intended beneficiaries? Strategy maps, overall elaboration on visions and missions
Evaluative approach	
Fundamental questions	What has been achieved and how have we ended up at the prevailing situation? What happens if R&D activities remain the same?
Questions related to impact assessment	What are the impacts thus far (beneficial, detrimental, inside and outside the target area etc.)? By which mechanisms have the impacts come into existence? Who are the final beneficiaries?
Data	Indicators, statistics, SWOT analysis etc.
Foresight approach	
Fundamental questions	How is the development likely to proceed? How is R&D activity likely to proceed? What are the weak signals or as yet unforeseen changes and challenges?
Questions related to impact assessment	What are the most probable impacts in the future? What are the social issues the organisation is likely to have the most impact on in the future? What are the factors most likely to hinder the emergence of positive impacts? What kinds of factors are most likely to give impetus to their emergence?
Data	Delfoi, panels, scenarios, weak signals, mega trends, simulations

implementation of public RD&I interventions. Quite the contrary, to maximise its societal benefits, public RD&I interventions should make a distinction between generic social impacts, value-chain social impacts, and the social dimensions of the competitive context thus feeding the prioritization process of social issues the intervention should focus on.

What is more, the *Impact Chain Model* recognises the fact that in order to become a distinct tool for strategic development and learning, the three approaches to knowledge creation need to be integrated – meaning here the integration of the evaluative, foresight and strategic approaches. All in all, it is claimed that by using the Impact Chain Model and by integrating the strategic, evaluative and foresight approaches implied by it, evaluation studies

can be much more than simply a means for justification. They can instead become a source of innovation and competitive advantage for the interventions under scrutiny while also playing a similar role for the actors and organisations involved in such interventions.

References

- Arrow, K.J. (1962) Economic welfare and the allocation of resources for inventions, in R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factor*. Princeton, NJ, Princeton University Press
- Bachtler, J. & Taylor, S. (2003) The Added Value of the Structural Funds: A Regional Perspective. *IQ Net Report on the Reform of the*

Caracostas, P. and Muldur, U. (1997) *Society, the endless frontier. A European vision of research and innovation policies for the 21st century*. Brussels: European Commission

CEC (2001) *European Governance: A White Paper*. Brussels, 25.7.2001, COM (2001) 428 final

CEC (2006) Community Framework for State Aid for Research and Development and Innovation, 31.12.2006, Official Journal of the European Union, Available at: <http://eur-lex.europa.eu/>

Edler, J., Kuhlman, S. and Behrens, M. (eds) (2003) *Changing Governance of Research and Technology Policy – The European Research Area*. London: Edward Elgar

Georghiou, L., H. Cameron, J. A. Stein, M. Neveda, M. Janes, J. Yates, M. Piper, M. Boden, and J. Senker (1993) *The Impact of European Community Policies for Research and Technological Development upon Science and Technology in the United Kingdom*. London, HMSO

Georghiou, L. (1999) Socio-Economic Effects of Collaborative R&D - European experiences *Journal of Technology Transfer* 24 (1), 69- 79

Hyytinen, K. And Konttinen, J. (2006) *Vaikuttavaa tutkimusta. Arviointikäytäntöjä julkisissa tutkimusorganisaatioissa: Puolustusvoimien Teknillinen Tutkimuslaitos [Research with an impact: Evaluation practises in public research organisations: Case PVTT]*. Espoo: VTT

Kjær, A. M. (2004) *Governance*. Cambridge: Polity Press

Kuitunen, S. & J. Oksanen (2002) *Mikä rooli rakennerahastoilla on alueiden innovaatiotoiminnassa ja innovaatiopolitiikassa?* Helsinki, Edita. [What is the role of Structural Funds in innovation activity and innovation policy? Research report.]

Kuitunen, S. & Hyytinen, K. (2004) *Julkisten tutkimuslaitosten vaikutusten arviointi*. Käytäntöjä, kokemuksia ja haasteita. VTT Teknologian tutkimus. Espoo. Impact assessment of public research organisations.]

Kutinlahti, P., Lähtenmäki-Smith, K. ja Konttinen, J. (2006) *Vaikuttavaa tutkimusta. Arviointikäytäntöjä julkisissa tutkimusorganisaatioissa: Helia ja SAMK*. Espoo: VTT. [Research with an impact: Evaluation practises in public research organisations. The universities of applied sciences case study]

Luukkonen, T. and P. Niskanen (1998) *Learning Through Collaboration - Finnish Participation in EU Framework Programmes*. Helsinki: VTT Group for Technology Studies.

Luukkonen, T. (2000) Additionality of EU Framework Programmes, *Research Policy*, Vol. 29, 2000, 711-724.

Luukkonen, T. (2002) Research Evaluation in Europe: State of Art. *Research Evaluation* 2 (11), 81 - 84.

Luukkonen, T. (ed.) (2004) *Biotekniikka. Tietoon perustuvaa liiketoimintaa*. Helsinki: Etna, Yliopistopaino. [Biotechnology. Knowledge-based business]

Luukkonen, T. & Hälikkä, S. (2000) *Knowledge Creation and Knowledge Diffusion Networks*. Tekes, Helsinki.

Lähtenmäki-Smith, K.; Hyytinen, K.; Kutinlahti, P.; Konttinen, J. (2006) *Research with an impact: Evaluation practices in public research organisations*. Espoo: VTT.

Lähtenmäki-Smith, K. ja Hyytinen, K. (2006) *Vaikuttavaa tutkimusta. Arviointikäytäntöjä julkisissa tutkimusorganisaatioissa: Maa- ja elintarviketalouden tutkimuskeskus MTT*. Espoo: VTT. [Research with an impact: Evaluation practises in public research organisations. Case Agrifood Research]

Lähtenmäki-Smith, K. and Kuitunen, S. (2007) "Efficiency versus Legitimacy: The Governance of Technology" in Blühdorn, I. (ed) *In Search of Legitimacy: Complexity, Efficiency and the Foundations of Democratic Politics*. Routledge, Forth-coming in 2007

Metcalf J. S. and Georghiou, L. (1997) *Equilibrium and Evolutionary Foundations of Technology Policy*. CRIC Discussion Paper No. 3, The University of Manchester, Manchester: mimeo

Mickwitz, P. (2006) *Environmental policy evaluation: concepts and practice*. Helsinki: Finnish Society of Sciences and Letters

Mowery, D.C. (1983) Economic theory and government technology policy. *Policy Sciences*, 16, 27-43

Mowery, D.C. (1994) *Science and Technology Policy in Interdependent Economies*. Boston, Dordrecht, London, Kluwer Academic Publishers

Nelson, R.R. (1959) The simple economics of basic scientific research. *Journal of Political Economy*, 67

Noro, A. (2000) "Aikalaisdiagnoosi: sosiologisen teorian kolmas lajityyppi" in Rahkonen, K. (ed.) *Sosiologia Nykykeskusteluja*. Gaudamus, Helsinki. [Diagnosis of Contemporary Social Conditions: A Third Type of sociological theory" in *Contemporary Sociological Debates*.]

Pollitt, C. and Bouckaert, B. (2000) *Public Management Reform: a comparative analysis*, Oxford: Oxford University Press

Pollit, C. and Bouckaert, G. (2004) "Public management reform. A comparative analysis", Oxford University Press, Oxford

Porter, M.E. & Kramer, M.R. (2006) Strategy and Society. *Harvard Business Review*, December 2006

Shadish, W.R., Newman, D.L., Scheirer, M.A., & Wye, C. (Eds.) (1995) *Guiding Principles for Evaluators*. San Francisco: Jossey-Bass

Stankiewicz, R. (1992) Technology as an autonomous socio-cognitive system, in Grupp (eds.): *Dynamics of Science-Based Innovation*. Berlin: Springer-Verlag

Tiihonen, Seppo (2004) *From Governing to Governance. A Process of Change*. Tampere: Tampere University Press

Virtanen, P. and Wennberg, M. (2005) *Prosessijohtaminen julkishallinnossa*. [Process Leadership in Public Management] Helsinki: Edita

Virtanen, Petri (2007) *Arviointi: arviointitiedon luonne, tuottaminen ja hyödyntäminen*. [Evaluation: the nature, production and utilisation of evaluative knowledge], Helsinki: Edita.

Top-down or bottom up? Methods for evaluating employment effects in Nordic Structural Fund programmes

By Jon Moxnes Steineke¹

Introduction

Both theory and empirical evidence point to the fact that increased regional inequalities and the increased geographical concentration of economic activities are side effects of the industrial restructuring process produced by ongoing economic integration and technological progress (Boldrin and Canova 2001, Henderson and Morgan 2002, Krieger-Boden 2002, Funck and Pizzati eds. 2003, Dupont and Martin 2003, Midelfart 2004). This makes the objective of employment effects all the more pressing a policy concern and necessitates the assessment of policy interventions which seek to create positive employment effects. Across a range of new economic geography and economic growth models, the effects of regional subsidies to lagging regions or regions suffering from industrial decline are being explored. A variety of regional policy initiatives are evaluated according to their impact on regional inequalities, the welfare of different groups, and the social welfare of the economy as a whole.

By modelling various industrial relocation policies a number of potential disadvantages have been identified. While seeking to promote social and economic cohesion, the preferred regional policies may in fact impede national economies from engaging in regional specialization driven by technological advances (Sapir et al. 2003). In this setting, EU regional policies as implemented through the Structural Funds (hereafter SF) do not appear to have enhanced the economic capacity of the regions to which structural funding is allocated.

¹ I am grateful to Åsa Pettersson for data management and research assistance in the evaluations projects upon which this article is based.

Kokko and Gustavsson (2004) provide some insight into this problem, comparing manufacturing employment in SF supported and non-SF supported remote Swedish counties in the period 1990-2000. They observe that counties qualifying for Objective 1 and Objective 2 programme support did not perform any better in creating manufacturing jobs than counties where such regional support was not available.

In addition to economic research into the effects, there is a regular monitoring and assessment routine built around the evaluation of the Structural Funds, which should also provide us with some insights and rigorous evaluative assessments of such claims at the meso and micro levels. The critical question posed then may be whether these evaluations withstand closer inspection.

In this brief chapter on the evaluation of job creation in the SF context, the need to improve the methods used to evaluate regional policy employment effects will be assessed. Labour market outcomes (new jobs, on the job training and competence development) are important not only in themselves, but also as such outcomes are used as favoured proxies for a range of other regional development policy impacts. In the first part of the text a short introduction on how to measure employment effects will be provided. In part two, the top-down and bottom-up approaches will be contrasted. Previous experiences from *ex post* evaluations of the 1995-1999 Nordic SF programmes will be drawn upon as examples here. Finally, the chapter will close with a brief discussion of the way forward with a view to making the evaluation of employment effects more methodologically robust.

Measuring employment effects

Both the Commission and different national programme management authorities have in the past produced *ex post* evaluations of the SF programmes.² While the Commission calls for synthesis reports at the national level, incorporating all national programmes within an Objective, the national programme management authorities perform evaluations of all programmes separately. The application of different data sources may differ significantly in the different evaluations, even when assessing programmes within the same Objective.

The core indicators used for measuring employment effects are the number of jobs created and the number of jobs maintained. These two indicators were almost universally applied in the Nordic SF 1995-1999 programmes. Although the SF projects themselves tended to focus on new jobs created, many also reported on jobs maintained as an additional effect of the individual intervention. In addition, ESF-funded programmes may also include the number of employees or jobless who receive training as additional employment effects. Methods for assessing such indicators are not included in the review and discussion provided here.

Jobs created and jobs maintained provide merely preliminary, gross measures of employment effects of a particular intervention. To provide an exact estimate of total impacts these effects need at least in theory to be adjusted for displacement or substitution effects (CSES 2006).

Displacement effects occur when the positive effects of an intervention have negative side effects (for instance when enterprises such as call centres re-locate to territorially assisted areas as a result of SF co-funding opportunities, leading to job losses at the previous location of the call centre). Substitution effects on the other hand occur when individuals who have received training, for instance through an ESF-funded intervention, obtain jobs that would otherwise have been offered to an individual (be they an existing employee or a jobless person) who has not received such assistance.

Evaluating the employment effects of Structural Fund programmes: bottom-up vs. top-down approaches

Determining the manner and extent to which expected objectives have been achieved as a result of a programme is a basic issue in programme evaluation. With respect to assessing the employment impacts of programmes there are, as Bartik (2003:34) observes, “far too many cases where state and local development organizations claim credit for

any state or local job growth that the organizations happen to have subsidized”. There is clearly then every reason to ensure that attempts are made to evaluate these effects as rigorously as possible and to avoid counting effects twice (or in fact even more than twice).

When designing the programme, the ideal evaluation scheme involves comparing two groups (be it a group of municipalities, businesses or individuals, or GTSAs - geographically targeted support areas), one of which has been exposed to the programme and one that has not been so - and then attributing differences between the groups to the programme. Causal inferences may then be made by comparing identical groups before and after exposure to the programme. What usually distinguishes evaluations of employment effects is the degree to which such comparisons are made between groups that are identical in every respect save the exposure to the programme or not, or to what extent they in fact are statistically comparable.

Most geographically targeted support areas in the context of Structural Funds programmes are constructed groups in the sense that they share some common socio-economic characteristics (meeting similar criteria or thresholds). On the other hand they are composed of parts of different functional labour market areas, which span different administrative units such as counties.³ Identifying identical groups or ‘twins’ to a SF target area then becomes an exercise in constructing yet another synthetic geographical unit.

The SF policy impact on designated Objective 1, 2 or 3 regions cannot be evaluated experimentally in a field setting by comparing economic outcomes in the SF (‘treatment’) areas unless some eligible local areas have been randomly chosen not to receive such assistance. Since such experimentation is politically impractical in the context of the Structural Funds, various proponents have argued that regional development programmes can and should be evaluated by quantitative (statistical) analyses of economic outcomes in firms or areas benefiting from ERDF and ESF funding, or more intensively using the economic outcomes of regions/firms benefiting from SF co-funding (‘treatment’ groups) and similar outcomes in comparison firms or areas (‘comparison’ groups).

When the treatment group is compared with a single comparison group, there is a pair-wise matching process between the programme area in the SF context and the control group of similar characteristics. Bondonio (2002a) and O’Keefe (2004) have used this method with some success in evaluating the performance of local economic development programmes in the US. O’Keefe suggests that

² In the 2007-2013 period, the responsibility for Ex-ante evaluation lies with the Member States, on-going evaluation is the responsibility of both the Member States and the Commission while the Ex-post evaluation is the responsibility of the Commission.

³ See Karlsson and Olsson 2002 for a discussion of the composition of the Swedish Objective 2 Fyrstad target area (1995-1999) for an example.

although local economic development programmes may display employment growth after the initial designation (corresponding to the EU programming periods), the employment effects does not persist over a longer time period.

What evaluation strategy to use (one group vs. comparison group matching) is primarily based on data availability considerations (Bondonio 2002b). The Swedish authority responsible for commissioning SF evaluations during the previous programming period NUTEK (2003a) may serve here as an illustration of a rather uninformed attempt at pair-wise matching of the Swedish GTSAs.

In the ex post evaluation of the Swedish objective 2 sub-programmes Blekinge, Fyrstad and Bergslagen artificial twin territories to all the three sub-programme areas were identified based on a basic set of selection criteria, making the twin territories comparable with the objective 2 sub-programme areas (see table 4):

In addition to the criteria displayed above, the territorial pairs were matched with respect to their composition

of municipality types, their 1999 commercial climate, the change in industry specialisation 1995-1999 and the 1995-1999 relative change in the daytime population.

At the municipal level, it proved impossible to display any statistically significant differences in terms of relative economic performance(s) in the 1995-1999 period. A weakness of the comparative approach used was that each area included only from 4 to 24 municipalities – much too few to perform a proper quantitative assessment. Another complication might have been that the analytical period used in observing outcomes – impacts were anticipated to be immediately available by the end of the programming period – may have been too short. Durable labour market effects may typically be displayed with a time lag of several years after the end of an intervention of this kind.

The rather crude approach displayed in NUTEK (2003a), creating synthetic twin areas for the pair-wise matching, has not been repeated elsewhere in the Nordic context.

Top-down approaches

A variety of more methodologically robust, statistical approaches can be used to compare firms and areas that use or are assumed to benefit from a SF programme with comparison groups of firms or areas that do not use the programme. The identification and selection of comparison groups provide for various evaluation strategies. In some instances, firms or organisations that have applied for but been denied SF co-financing have been included in the 'treatment' group. In other instances, these SF 'rejects' are pooled and used as 'control groups' in themselves.

At the local area/regional level, the identification of the 'treatment area' or region is quite straightforward in Structural Funds programmes. The objective 2 and 3 areas are explicitly identified and mapped municipality by municipality in the programme documents. However, the identification of a 'control' area or region critically depends on the parameters used in identifying the geographical twins of the various programme areas. Twins may be identified as geographically adjacent to the programme area (see NUTEK 2003a for an illustration), or they may be identi-

Table 4: Objective 2 programme area and twin area characteristics: some comparisons

Area	No. of municipalities	Population (Dec. 31 1999)	Population change 1995-1999 (in %)	Relative change in number of new firms established per capita 1995-1999
Bergslagen Obj, 2 sub-programme	24	384023	-5.71	-1.42
Bergslagen 'twin'	16	383507	-1.23	-1.28
Blekinge Obj. 2 sub-programme	5	150625	-1.56	-1.1
Blekinge 'twin'	4	154326	-1.33	-1.04
Fyrstad Obj. 2 sub-programme	4	153293	0	-1.75
Fyrstad 'twin'	8	130101	-2.99	-1.6

fied and delimited according to socio-economic correspondence indicators that do not include geographical proximity to the SF programme area in question (see NUTEK 2003b for such an approach).

Quantitative approaches are based on statistical theories and models, while qualitative approaches to evaluation consist of a range of sociological, psychological and anthropological sub-disciplines that do not necessarily base their empirical investigations on theoretical grounds such as the testing of hypotheses (Kvitastein 2002). Qualitative approaches may be preferred in formative evaluations, that is evaluations that mainly aim at supporting or correcting programme incentives as they are being performed (*ex ante* or mid-term). Quantitative approaches yield the most benefits *ex post*, in the quest to assess the more durable impacts of SF programmes over the long term. In terms of employment impacts of territorial development programmes, early case studies of urban enterprise zones in the US displayed no relationship between incentives and the extent of new job creation (Rubin and Wilder 1989).

A stream of impact evaluations have emerged the last five years to complement previous case studies or synthesis reports examining job creation in relation to geographically targeted policy incentives, both in the US and in Europe. In particular, the literature assessing the impacts of US state enterprise zones is growing (e.g. Bondonio 2000a,b; 2002a,b, Bartik 2003, O'Keefe 2004). In these studies, analysis is focussed on measuring the net impact on employment and business outcomes of the programme incentives net of influences exogenous to the programme. Bondonio (2002b) provides the first econometric impact evaluation of an Objective 2 programme in a EU setting.

Both the US and Bondonio studies aim at robustly estimating the *ex-post* net employment impact of business incentives offered in the support areas. The evaluations are

developed using a 'comparison group design', where pre-post employment changes recorded in the target area (be it an urban enterprise zone in the United States or an Objective 2 area in the EU) are compared to those of adjacent non-target areas. A number of econometric specifications are then used to estimate the employment impacts (see Bondonio 2000a; 2002a and Bartik 2003 for more detailed presentations).

Bondonio (2002b) concludes that no significant employment change was induced by the SF programme in the Piedmont Objective 2 area in the 1994-96 programming cycle. The finding occasioned two different interpretations – either the Objective 2 programme business incentive did contribute to modifying the investment or hiring policies of the targeted firms, and/or the structural funds incentives induced only miniscule employment changes in the targeted Piedmont areas compared to the total size of the regional economy.

Note that such econometric assessments can only be performed 4-6 years after the end of the programming period due to time required to produce national and regional socio-economic data. This could be an issue for a future thematic evaluation of the employment effects of regional SF programmes of the 1994-99 period (Objective 2, 5b and 6), as well as the Structural Funds programmes of the recently completed 2000-2006 programming period.

In several SF evaluations the evaluators have made a point of calculating the cost of creating new jobs in SF co-financed projects comparing this with the cost of creating new jobs in similar initiatives not funded by the ERDF or ESF (Oxford Research AS 2001). In most of these comparisons, the SF input comes out favourably. Given the unavailability of robust estimates of the job creation effect of SF programmes, such exercises should be avoided as we have already displayed that they tend to be based on mere 'guesstimates'.

Bottom-up approaches

Quantitative, top-down methods may be applied to control the robustness of the results provided by the more qualitative approaches employed by public administration, social anthropology and economic sociology. Is there current evidence to suggest that the jobs reported in the programme registers (as displayed in the end-of-project reports to the national programme authorities) have been maintained beyond the programming period?

The quality of follow-up data registers in the 1995-1999 programming period were found to remain lacking both in Finland (Vehkasalo 2002, Katajamäki 2002) and Sweden (NUTEK 2002a,b; 2003a,b). As a result, only a small minority of jobs entered into the various registers were

found to remain permanent. Auditing a sample of Finnish Objective 5b projects in 2000, Vehkasalo (2002) found that the employment data was highly inaccurate. Both Vehkasalo and Katajamäki (2002) showed that a more detailed control of project outcomes produced an enduring job generation effect at only 15% of the employment outcome registered by the national programme authorities. Similar discrepancies were found in Swedish Objective 6 programmes (Katajamäki *op.cit.*) and Objective 2 programmes (NUTEK 2003).

According to several recent studies (CSES 2003, EC 2004), support for expenditure on R&D, innovation and technology transfer appears to have been particularly ef-

fective in creating new jobs as well as in upholding existing ones.

In the opinion of the Finnish State Audit Office a simple telephone interview with former project managers in order to determine the job generation effect of projects is insufficient as a qualitative, bottom-up measure. *In situ* controls – workplace visits by evaluators – may be required to assess the actual quantitative impact in terms of new and permanent jobs created. Such direct assessments of the job creation effects are however impractical in SF programmes dominated by large framework projects.

Assessing employment impact is however the most sig-

nificant factor in explaining employment effects. Employment effects tend to erode over time – as displayed in table 5. Here we compare employment effects assessed at the beginning of the programming/project periods with reported outcomes some time after the completion of SF programmes. In all but the first instance the information has been collected from the same sources – the project leaders.

As displayed in the compilation, employment effects reported at the project level vary depending on when such information is obtained. The most robust estimates are probably obtained by combining project level information obtained *ex post* with statistical approaches.

A review of the evaluation of employment effects in Nordic SF programmes: how are they assessed?

Estimating employment effects is one thing – it is however quite another to identify which bottom-up approach to use when performing SF evaluations. In addition do associations exist between the evaluation methods selected and the reported outcomes in terms of how SF programmes fare as compared to their initial ambitions in terms of producing sustainable jobs?

In table 6; we display the data sources applied in the *ex post* evaluation of sixteen of the twenty regional Objective 2, 5b and 6 programmes that were in operation in the second wave of EU SF programmes (Hanell et al. 2002). In column seven of table 1 we display the regional programme evaluators' assessment of the employment impact of the various programmes. For comparative purposes we present the assessments of the employment impacts made by the EC's national programme evaluators *ex post*.

Although employment figures are subject to 'some uncertainty' (Oxford Research 2001:10), the Danish Objective 2 programmes and the Åland Objective 5b programme are the only ones where the evaluators have concluded that the employment effect more than fulfilled the goals for the programmes. There are no deficiencies in the quantitative design selected in making these assessments - the Danish evaluators calculated the employment effect on the basis of questionnaire responses from more than 250 projects out of a total project portfolio of 800, while the Åland 5b evaluators interviewed a net sample of some seventy project managers, thus probing close to fifty per cent of the total number of projects in the programme.

EC programme evaluators typically assess the employment impact to be more positive than those obtained in national sub-programme evaluations commissioned by

Table 5: Expected employment effects by stage in the programming period. Job creation in Objective 2 sub-programmes

Period		Objective 2 (1995-1999) sub-programme		
		Bergslagen	Blekinge	Fyrstad
1995	Job creation objective as in SPDs	9 000	2250	5000
1996-1999	Results expected by new projects (as reported in project applications)	11 000	2000	2500
1996-2002	Results reported by finalised projects	6 000	2400	3000
2002	Results reported by national programme management authority <i>ex post</i> (STINS)	4 500	1500	4500
2003	Estimated sub-programme outcomes following <i>ex post</i> interviews with a sample of project leaders	800	500	400

Table 6. Bottom-up data sources used to assess employment effects (jobs created and maintained) in a sample of national evaluations of Nordic SF programmes (Ex Post 1994 (95) – 1999)

SF Programme	National evaluation Data Sources				Programme objective (SPD) obtained in terms of employment impact?		
	File reviews at nat. programme management	Project case studies	Expert opinions at		Surveys (postal or phone)	EC evaluations	National evaluations
			project level	programme level			
Denmark Obj. 2 NJ	√	√		√ ¹	√	No	Yes (>100%)
Denmark Obj. 2 L						No	Yes (> 200%)
Denmark Obj. 5b	√		√	√ ¹	√		...
Finland Obj. 2						Yes (113%)	No
Finland Obj. 5b mainland	√				√ ³		Ambiguous*
Åland Obj. 5b	√			√	√		Yes (120%)
Finland Obj. 6						No (15%)	
Sweden Obj. 2 BFB	√	√ ²		√	√	Yes (169%)	No (8-22%)
Sweden Obj. 2 NN, ÅK	√		√	√	√	Yes (169%)	Ambiguous*
Sweden Obj. 5b VGD	√		√		√		Ambiguous*
Sweden Obj. 5b West	√		√	√			Yes
Sweden Obj. 5b Southeast	√		√	√			No (< 97%)
Sweden Obj. 5b Archipelago	√		√	√			Not determined due to lack of data
Sweden Obj. 5b Gotland	√		√	√			No (88%)
Sweden Obj. 6	√		√	√	√	Ambiguous*	Ambiguous*

¹ Focus group meetings

² Of project portfolio in a sample of municipalities

³ Limited inquiry

... Employment objective not quantified

* Depending on source

the national programme management. Thus, while the national evaluations of the Swedish Objective 2 programmes found that the programme objectives were not obtained in terms of employment impact (NUTEK 2003a,b), the national programme evaluation commission called for by the Commission came to the opposite conclusion (CSES 2003). The discrepancy may be explained by the richer data sources used in the national evaluations. While the EC evaluations were uniquely based on reviews of second-

ary sources (e.g. monitoring data; typically data file reviews and administrative records), the national programme evaluations benefit from input from both primary and secondary sources (see table 6). And the selection of data sources has repeatedly proved to be a critical issue in terms of estimating the job creation of SF programmes:

Evaluating the employment impact of the Swedish Objective 6 programmes for the period 1995-99, Wiberg *et al.* (2002) randomly chose 20 projects for interview, and

found that the job creation effect of these projects were below 10% of what was reported in the database maintained by the national programme management authorities (STINS). This finding was confirmed by interviews with key interviewees both at the national and the regional level.

A similar discrepancy was reported by Vainio and Laurila (2002) who performed the same exercise on the Finnish Objective 6 programmes for the 1995-99 period. On the basis of expert interviews, the evaluators concluded that the national programme management database – FIMOS – exaggerated the actual employment effect by a factor of seven or eight. Thus Vainio and Laurila estimated the actual employment impact to be 15 per cent of the project outcome figures registered in the FIMOS database.

Most of the SF evaluations surveyed for table 6 provide

only weak quantitative evidence in terms of providing assessments of jobs created or maintained, and none in terms of jobs relocated.

The main lesson drawn from the compilation of national evaluations of Nordic Structural Funds programmes (1995-1999) is that expert opinions or case studies at the project level in most cases need to be complemented with survey material. Register data made available by the national programme management authorities in many instances remains of such dubious quality that it needs to be complemented with other primary data sources. EC evaluations should take advantage of national evaluation exercises and be timed so as to be able to incorporate findings from the national evaluations. After the completion of the 1995-99 programming period, the phasing between EC and national evaluation was the opposite, with EC *ex post* evaluations preceding the national ones.

The way forward

Assessing the job creation effect of SF programmes is by no means a straightforward exercise. For employment outcomes it is important to distinguish between the pre and post intervention situations and thus between the changes accounted for by three different types: new jobs created, jobs maintained and jobs retained. In many SF programmes such distinctions are still unavailable. Programme objectives may have been set for the former type but not for the latter two. This is a serious hindrance to performing any quantitative assessment of the employment impact of SF programmes.

The counter-factual situation is often another major challenge in evaluation terms. It is often not possible to imagine what outcomes would have come about without the interventions that actually takes place.

Although national territorial support zones and EU support zones frequently overlap, they are not totally congruent. It is often difficult to disentangle outcomes that are due to nationally developed interventions from those that are developed within the context of Structural Funds. Only continuous refinement of top-down approaches, which controls statistically for the effects of competing national initiatives, points to a way out of these shortcomings. This calls for a greater involvement of the research community in programme evaluations.

The most pressing issue in terms of enabling good evaluation practices relate to programme management issues. It is essential in the management of SF programmes to

quantify employment objectives at the programme level in terms of jobs created, jobs maintained and – hopefully – also in terms of jobs relocated from the very outset. This bottom-up requirement would ask programme managers to call to the attention of prospective project partners the significance of setting quantifiable targets for all these diverse types of labour markets outcomes from the very outset – in itself a rather ambitious requirement.

Not all project managers adhere to or stay with initial targets. This might even be a blessing in disguise. An additional requirement could be to call upon programme managers to ensure that it is possible to identify projects where employment effects are registered even when such results are not anticipated at all from the outset. In the context of the Structural Funds project cycle, identifying and following up on the sub-set of projects within a SF programme that produce *unanticipated* employment outcomes (of all kinds) would be of great value in preparing programmes for the future.

Another issue concerns the need to acknowledge the fact that employment effects may be incidental, and that robust estimates may be available only some time after the closure of Structural Funds programmes. *Ex post* evaluators could however take better advantage of recent developments in quantitative analyses (top down approaches) so as to control the reliability and validity of information made available to them at the project level.

References

SF evaluations (programme evaluations, national evaluations and syntheses) reviewed

CSES (2003) *Ex Post Evaluation of 1994-99 Objective 2 Programmes – Synthesis Report*. Report to the European Commission's DG Regio prepared by Centre for Strategy and Evaluation Services, Sevenoaks (June).

Glesbygdverket (2001a) *Slutlig utvärdering Mål 5b Gotland. Ex Post evaluation of the Objective 5b Gotland programme from Ledningskonsulterna i Stockholm AB* (Oktober).

- (2001b) *Slutlig utvärdering Mål 5b Västerbotten/Gävle/Dala. Ex Post evaluation of the Objective 5b North programme from ITPS, Östersund* (Oktober)
- (2001c) *Slutlig utvärdering Mål 5b Sydöstra Sverige. Ex Post evaluation of the Objective 5b Southeastern Sweden programme from Ledningskonsulterna i Stockholm AB* (Oktober).

IM Gruppen (2001a) *Slutlig utvärdering Mål 5b Skärgården. Ex Post evaluation of the Objective 5b Archipelago programme / Report 2001:3* IM Gruppen, Uppsala.

- (2001b) *Slutlig utvärdering Mål 5b Västra Sverige. Ex Post evaluation of the Objective 5b Western Sweden programme / Report 2001:2* IM Gruppen, Uppsala.

Kalliomäki, A. and V. Vehkasalo (2000) *Projekt för utvecklande av landsbygden: sysselsättnings-effekterna av program 5b*. Förvaltningsrevisionens berättelser no. 11/2000. Statens Revisionsverk / State Audit Office (SRV), Helsinki.

Katajamäki, H. (2002) *Ex-post evaluation of Objective 6 programmes in Finland and Sweden for the period 1995-99*. Synthesis report to the European Commission/DG REGIO. University of Vaasa (December).

Moos, J.U. (2001) *Slutevaluering af program for de danske Mål 5b-områder 1994-99*. Report to the Danish Ministry of Agriculture by Teknologisk Institut AS (August).

NUTEK (2002a) *Slutlig rapport för Mål 5b Skärgården 1995-1999*. Stockholm: Swedish Business Development Agency (NUTEK).

- (2002b) *Slutlig rapport för Mål 5b Västra Sverige 1995-1999*. Stockholm: Swedish Business Development Agency (NUTEK).
- (2003a) *Regional kraftsamling – Slututvärdering av Mål 2 Bergslagen, Fyrstad och Blekinge 1995-1999*, NUTEK Info 008-2003. Stockholm: Swedish Business Development Agency (NUTEK).
- (2003b) *Från regional fördelningspolitik till regional tillväxtpolitik – Slututvärdering av Mål 6 och Mål 2 Norra Norrlandskusten samt Mål 2 Ångermanlandskusten 1995-1999*, NUTEK Info 006-2003. Stockholm: Swedish Business Development Agency (NUTEK).

Oxford Research AS (2001) *Evaluation of Objective 2 programmes in North Jutland 1994-99*. Report to the County of North Jutland (September).

Statistics and Research Åland (2002) *Efterhandsutvärdering Ålands mål 5b-program 1995-1999*. Report no. 2002:6. ÅSUB, Mariehamn (July 2001).

Statistics and Research Åland / Nordregio (2003a) *Halvtidsutvärdering Ålands mål 2 program* [Mid Term Review Åland Objective 2 programme], ÅSUB (Statistics and Research Åland) Report no. 2003:7. ÅSUB, Mariehamn.

- (2003b) *Halvtidsutvärdering Ålands mål 3 program* [Mid Term Review Åland Objective 3 programme], ÅSUB (Statistics and Research Åland) Report no. 2003:8. ÅSUB, Mariehamn.

Vainio, A. and P. Laurila (2002) *Ex-post evaluation of Objective 6 programmes for the period 1995-99. Country report for Finland*. Report to the European Commission by the Levón Institute, University of Vaasa (December).

Wiberg, U., B. Jansson and L. Lundmark (2002) *Ex-post evaluation of Objective 6 programmes for the period 1995-99. Country report for Sweden*. Report to the European Commission by the Department of Social and Economic Geography, Umeå University (December).

Other references

Bartik, T.J. (2003) "Evaluating the impacts of local development policies on local economic outcomes: what has been done and what is doable?" Upjohn Institute Staff Working paper no. 03-89, W.E. Upjohn Institute for Employment Research, Kalamazoo.

Boldrin, M. and F. Canova (2003) "Regional policies and EU enlargement", CEPR Discussion Paper no. 3744, Centre for Economic Policy Research, London. <URL> <http://www.cepr.org/pubs/dps/DP3744.asp>

Bondonio, D. (2000a) "Evaluating decentralized policies: how to compare the performance of state and local economic development programmes across different regions", Paper presented at the 4th EES Conference Lausanne October 12-14th.

- (2000b) "Statistical methods to evaluate geographically-targeted economic development programs", *Statistica Applicata* vol. 12 no. 2, pp. 177-204.
- (2002a) "Evaluating decentralized policies. A method to compare the economic performance of economic development programmes across different regions or states", *Evaluation* vol. 8 no.1, pp. 101-124.
- (2002b) "Evaluating the employment impact of business incentive programs in EU disadvantaged areas. A case study from Northern Italy", Working Paper – Economics series N. 28, POLIS – University of Eastern Piedmont. <URL> <http://polis.unipmn.it/pubbl/index.php>

Bondonio, D. and R. T. Greenbaum (2006) Do business investment incentives promote employment in declining areas? Evidence from EU Objective-2 regions *European Urban and Regional Studies* vol. 13 no. 3; 225-244.

CSES (2006) Study on Measuring Employment Effects – Final Report and Guidelines. Report to the EC (June). Centre for Strategy and Evaluation Services, Sevenoaks UK.

Dupont, V. and P. Martin (2003) "Subsidies to poor regions and inequalities: some unpleasant arithmetic", CEPR Discussion Paper no. 4107, Centre for Economic Policy research, London.

European Commission (2004) *A new partnership for cohesion*. Third report on economic and social cohesion. EC, Brussels.

Funck, B. and L. Pizzati (eds.) (2003) *European integration, regional policy, and growth*. Washington DC: The World Bank.

- Henderson, D. and K. Morgan (2002) "Regions as laboratories: the rise of regional experimentalism in Europe", in Gertler and Wolfe (eds.) *Innovation and Social Learning: Institutional Adaptation in an Era of Technological Change* Basingstoke: Palgrave, pp. 204-226.
- Karlsson, C. and M. Olsson (2001) Methods for identifying functional regions: theory and applications, in Johansson, Karlsson and Stough (eds.) *Regional Policies and Comparative Advantage* Cheltenham: Edward Elgar, pp. 367-389.
- Kokko, A. and P. Gustavsson (2004) *Regional integration, foreign direct investment, and regional development*, pp. 111-135 in EIB Papers vol. 9.1. Luxembourg: European Investment Bank.
- Krieger-Boden, C. (2002) "European integration and the case for compensatory regional policy" Kiel Working Paper no. 1135 Kiel Institute for World Economics, Kiel (November).
- Kvitastein, O.A. (2002) Offentlige evalueringer som styringsinstrumenter: kravspesifikasjoner og kontrollproblemer, SNF Report no. 30/2002, Institute for Research in Economics and Business Administration (SNF), Bergen <URL> <http://www.snf.no>
- Malan, J. (2000) Counting the jobs – fact or fiction? Paper presented at the RSA EU Regional Policy Conference Aix-en-Provence (14-15. September).
- Midelfart, K.H. (2004) Regional policy design: an analysis of relocation, efficiency and equity. Mimeo, Norwegian School of Economics and Business Administration, Bergen (January).
- OECD (2004) Evaluating Local Economic and Employment Development: How to Assess What Works Among Programmes and Policies. LEED (Local Economic and Employment Development) OECD, Paris.
- O'Keefe, S. (2004) "Job creation in California's enterprise zones: a comparison using a propensity score matching model", *Journal of Urban Economics* vol. 55 no.1, pp. 131-150.
- PLS Rambøll Management (2002) Employment effects of the structural funds: an assessment based on theoretical considerations and on empirical evidence. Report to the EC (August).
- Rubin, B. and M. Wilder (1989) "Urban enterprise zones: employment impacts and fiscal incentives", *Journal of the American Planning Association* vol. 55 no. 4, pp. 418-431.
- Sapir, A. et al. (2003) *An agenda for a growing Europe – Making the EU economic system deliver*. Report of an independent high-level study group to the President of the European Commission (July).
- State Audit Institution (2005) Regionala stöd – styrs de mot ökad tillväxt? Riksrevisionen report RIR 2005:6. SAI, Stockholm (March).
- Treasury Board of Canada (1998) *Program evaluation methods: measurement and attribution of program results* (3rd ed.) <URL> <http://www.tbs-sct.gc.ca/eval/pubs/method/pem.pdf>
- Vehkasalo, V. (2002) "Effectiveness of rural employment policies: evidence from Finland", Paper presented at the 5th EES Conference Seville October 10-12th.

Sitting on both sides of the fence: Meshing theory and practice to construct useful regional pro- gramme evaluations

By *Jörgen Gren*¹

Abstract

The main questions addressed in this chapter involve those of combining theoretical insight into evaluation methodology with the practical concerns of the commissioners and the contractors of evaluation projects. Is there an ideal balance between theory and practice, and how can these two spheres interact and communicate in a constructive way? Finally, can an evaluation inclusive approach (to policy) be said to have developed across Europe? The present chapter seeks to explore these issues further in a wider European context using examples from the Structural Funds programmes mainly, though not exclusively, in France.

The issues analysed through a set of sub-chapters include:

- The development of evaluation awareness and the uses

of evaluations.

- The construction of impartial evidence as the basic building block of an evaluation.
- The nature of indicators and evaluation reports: quantitative or qualitative (processes or figures).
- The role of the evaluator and the needs of decision-makers, and,
- How to design evaluation to make it useful.

It is important to note that the article will address evaluations of larger socio-economic programmes or policies and *not* particular projects. This differentiation in the degree of activity, as well as in the unit of analysis is very important when terms like evidence, evaluation or action are used.

Increased evaluation awareness in the EU and the operational nature of evaluations

With increasing questioning and scrutiny of public intervention in both the economic and social spheres, governments are turning to research and evaluation evidence for legitimacy which is no longer guaranteed by democratic political processes alone. The decision-makers basically

¹ NB. All views and opinions expressed in this article are personal and do not necessarily reflect official European Commission policies or opinions.

need to know what works, what does not and why.²

² There are many publications on this theme of which perhaps the most noteworthy are Davies, Nutley and Smith 2000, White 1999, Van der Knaap 1995, Halachmi and Bouckaert 1996, Gren 2001, McVittie/Kim Swales 2003. See also the EU funded website for evaluation methodology development www.evalsed.info, introduced elsewhere in this report. In addition, the OECD's major overview of evaluation practices and measuring performance in European states from 1997 is also useful.

In order to know what is happening, public decision-makers at all levels increasingly turn to evaluation using it as a management tool with operational effects i.e. an evaluation showing little or no results in a programme often means a cut in resources. This trend does seem clear enough and becomes even clearer if we consider the European level where, for instance, major changes in the regional programmes for the 2000-06 period – the so-called ‘mid-term’ review – were accompanied but also driven by evaluation (CEC 1999, CEC 2000). This trend is also compounded by the necessity, in the new financing period 2007-2013, to submit the Structural funds operational programmes with an *ex-ante* evaluation and, where relevant, a strategic environmental assessment (CEC 2006).

The mid-term evaluations, as they were done in 2003-04, are also interesting from the viewpoint of the “performance reserve”. This instrument sliced off 4% of the total endowment of a programme at the start of the 2000-2006 financing period only to redistribute this money half-way through the programme, in 2004. This type of reserve is also possible to implement, on a voluntary basis, indeed, for the period 2007-13 some countries have chosen to apply this instrument as a means of rewarding good performance and quality in programme work. The original objective was to stimulate programmes into performing better than in the 1994-99 period, hence the reserve only being made available after proven results or achievements

Table 7. Performance reserve in France (Objective 2 regions)

Région	Total (in €)	% of initial funding
Alsace	1 868 116	2.00%
Aquitaine	9 495 585	2.10%
Auvergne	5 991 387	2.00%
Basse-Normandie	9 633 418	3.63%
Bourgogne	14 857 738	6.36%
Bretagne	15 456 568	3.78%
Centre	11 738 568	5.89%
Champagne-Ardenne	12 619 483	6.08%
Franche-Comté	10 037 704	5.46%
Haute-Normandie	10 999 220	3.58%
Île-de-France	4 375 624	3.07%
Languedoc-Roussillon	17 099 710	6.33%
Limousin	2 751 252	2.00%
Lorraine	13 364 070	3.51%
Midi-Pyrénées	23 327 061	5.76%
Nord-Pas-de-Calais (Obj 2)	27 953 995	4.60%
Paca	10 478 003	3.41%
Pays de la Loire	22 533 054	5.62%
Picardie	17 007 265	6.55%
Poitou-Charentes	16 207 270	6.10%
Rhône-Alpes	14 431 576	3.53%
Total	272 226 667	

in the programme. These expected results were analysed during the *ex-ante* evaluation stage and were subsequently set out in the official programming documents. There was moreover a further form of competition built into the performance reserve since e.g. France chose a system which would penalise under-performing regions by distributing, wholly or partially, their part of the reserve to their neighbours.

In concrete terms, the distribution of the reserve in France took place in March 2004 and can be illustrated as follows for objective 2 regions:

As can be seen from the table 7. above, 9 out of 21 Objective 2 programmes were viewed as performing above average and thus received more than their “normal”³ 4% share. In other words, a genuine competition between regions was established in France. The basis for measuring success was evaluation and the results had a direct impact on the distribution of funding i.e. an outcome measured in financial terms.

This is only one of many indications of evaluative information with tangible financial implications for the regions. Not surprisingly there has been a clear increase in evaluation awareness on a European scale, as well as an indisputable political will to follow up evaluations through (negative or positive) actions both within the EU Structural Funds and in other areas. Indeed, the maintenance of the evaluation requirements in the new financing period (CEC 2006, Art 47), even if they of course reflect the general trend towards further decentralisation, could be seen as evidence of the mainstreaming of the evaluation instrument.

Looking at the phenomenon from a theoretical point of view, it could be said that clear system effects (Edwards, Spence 1994) have been produced by the introduction of this principle into the Structural Funds Regulation, as well as by the inclusion of the evaluation tool as a vital component in virtually all of the areas in which the European level has policy responsibilities. Indeed, European Commission publications such as the MEANS collection (CEC 1999b) or its web-based follow-up (www.evaled.info) indicate that evaluation practices already developed in the mid-1990s across all Member States, even in those where evaluation for various political, cultural or administrative reasons has been previously absent or viewed with a certain suspicion. The enlargement process could

³ ‘Normal’, in the sense of average, i.e. if every region in France was considered to be performing well, each would receive a 4% proportional share of the total.

also be said, as has been shown by the new Regulations, to have strengthened the process of making evaluation more widespread across Europe, as the new Member States have embraced it as a functional management tool and have also introduced this into their national systems (Estonian Ministry of Finance, 2006). Indeed, as the Council Regulation states (CEC, 2006, art 47.1): "Evaluations shall aim to improve the quality, effectiveness and consistency of the assistance from the Funds and the strategy and implementation of operational programmes with respect to the spe-

cific structural problems affecting the Member States and regions concerned, while taking account of the objective of sustainable development and of the relevant Community legislation concerning environmental impact and strategic environmental assessment."

It would then seem possible to speak of a generalised evaluation-inclusive and even mainstreamed situation for policy development and programme management in Europe, similar to the system effects analysed by Spence and Edwards as early as 1994.

The basic building blocks in an evaluation

Public policies are developed in a context that is marked by increasing social complexity. This means enormous challenges for elected politicians and various pressure groups to produce "hard evidence" of "what works" in this very same complex environment (Amin and Hausner 1997, Clarke and Newman 1997).

Evaluations based on evidence, constructed on data stemming from indicators or measurements of quantifiable tangibles, as well as qualitative information, is a theory and a concept inherited from the medical science world and basically means, in that context, to use research evidence along with clinical knowledge and reasoning to implement interventions that are effective.

Evaluation in a regional policy context generally implies the use of a mix of various social sciences (economics, sociology, geography, political science or even psychology, as well as statistics) to create a product for political use and/or as a tool for stakeholder negotiations (Monnier 1992, Conan 1998, Beutel 2002).

Hard evidence, solid objectivity or irrefutable proof, could an evaluation – the mix of social sciences and statistics - provide this? Experience suggests that it is difficult to reach such a status in evaluation (see Pawson 2002, Sand-

erson 2000 but also Walker 2001 for an analysis of this issue in the UK). Evaluation, by its very nature, concerns interpretations, initial hypotheses, empirical studies of generalised phenomena, assumptions, variables, estimates (sometimes 'guesstimates') but also a clear methodology, viable information, solid analysis and credible findings (DATAR 2003).

A practical approach to evaluation would then imply that evaluation can produce robust information and impartial conclusions but that this will constitute only *one* basis among many for judging programmes, policies, interventions or for preparing future policy development (or in fact both).⁴

Indeed, in the three most recent Structural Funds regulations (1994-99, 2000-06 and 2007-13), it is mentioned in various forms that competent authorities are obliged to assemble data, evaluate this data and make the results available to the public. The impartiality and independence of evaluations are concepts that are mentioned directly, e.g. in Regulation 1083/2006 (art 47.3). It is also clear, however, that management authorities are not bound to particular evaluation findings or recommendations.

The theoretical conflict between quantitative and qualitative approaches

There is, for reasons unknown, an ongoing discussion on whether evaluations should favour a quantitative or a qualitative approach (for information on the subject of both approaches and the discussion surrounding them see e.g. Shaw 1999, Conan 1998, Monnier 1992 or McVittie/

Kim Swales 2003). However, in real evaluation situations, there is no conflict and both are used to produce a report. Indeed, an evaluation is never complete without both approaches (Pawson and Tilley 1997).

To illustrate this complementarity, a few actual evaluation situations can be used: an evaluation that only treats financial input and spending percentages as proof of activity is not particularly useful for future policy decisions be-

⁴ For a discussion of the processes of deliberation and different evaluation formats see Dryzek 1990, Henkel 1991, Van der Knaap 1995, Gren 2001 and DATAR 2003.

cause it does not tell you what worked and how as it only tells how much money was spent and under which headings.

On the other extreme, evaluations without any quantification to support their arguments e.g. concerning perceptions of programme activity are not useful either as they provide little or no justification for the conclusions. The question of the magnitude of the stakeholders' perceptions – how many thought what and is that figure high/low compared to the total population or group - should be addressed and thus included in a coherent quantitative framework.

These two extremes do exist but most evaluations take the middle road and use both approaches where possible. Nevertheless, there are some perceptible trends in the current state of affairs concerning evaluation. It seems as if the regions (paymasters and authors of the terms of reference) and the contractors (evaluators) favour process evaluation which in essence is a qualitative type of analysis (Conan 1999). Indeed, greater emphasis seems to be placed on *how* it works rather than on *whether* it works.

The operational experiences from the mid-term review process – which is the latest ongoing activity-based large scale evaluation in the EU - for the Objective 2 Structural Funds programmes in France (2003-2004) and elsewhere would support part of this argument, although not fully. Indeed, the management and monitoring systems – the processes of delivery - have been thoroughly investigated in all Objective 1 and 2 mid-term evaluations for the period 2000-2006, seemingly at the expense of the results of the programme.

The apparent reasons for this lie in the commissioning of the evaluation in the timeframe and with the constraints of availability of monitoring data over and above financial input information. In other words, the evaluators' mandate (through the terms of reference) clearly stated that the management and quality of monitoring systems are key issues in the analysis. Further, the programmes had been active for only 2 years which does not give enough timeframe for the evaluators to go beyond anything but the initial results and financial analysis on the rate of consumption of funds i.e. how much was spent and in relation to which measures? In addition to this, for various reasons such as uncertainty over the robustness of information and large interpretation margins e.g. is the number of jobs created/maintained indicated by a project manager correct or not, time-frame issues as stated above or simply the lack of human resources, the result indicators were only partially entered into the computerised monitoring systems.

However, the focus on process evaluation was far from total and certain evaluators, such as in *Martinique* and *Corsica* (Objective 1) or *Nord-pas-de-Calais*, *Franche-Comté* and *Aquitaine* (Objective 2) were able to use existing data

or create information that enabled quantitative “causality” type of analyses. In addition, most evaluations also quantified the processes e.g. measuring lead times, payments in time or delays, level of indicator input etc.

In general, including initial information on the current *ex-ante* evaluation process preparing for the 2007-13 period, it could be said that there have been great qualitative leaps forward in terms of evaluation practices in most European countries, including the new Member States. Indeed, the latter, as exemplified by the annexed list of evaluations undertaken by Estonia in 2006, have often undertaken complex and thorough ongoing evaluations of their programmes. It should however be noted that there was no regulatory obligation for the new Member States to undertake any evaluation at all in the 2004-06 period⁵. Further, going back to the latest European-wide round of common evaluations - the mid-term evaluations 2003-04 - in six different countries⁶, it is possible to draw the following general conclusions:

- Basic information was made available to the evaluators to a greater extent than in the 1994-99 period.
- A clear evolution towards more thoroughly constructed indicators (input, output, results). This also implies that improved monitoring data and information was made available in a ready-to-use form i.e. less need to reconstruct or gather raw data for the evaluators.
- A better understanding of the goals and objectives of the evaluation tool.
- Improved quality of the analysis as well as improved knowledge or professionalism in terms of the human resources attached to the evaluation tasks (both in the consultancies and in the administrations).
- Timetables have been respected in all cases (financial penalties were attached to not finishing on time i.e. before 31 December 2003).
- Evaluations were mostly undertaken in partnership between the evaluator and the regional programme actors.

In these new (and improved) circumstances, evaluation is as much about *identifying good practice* and undertaking participatory evaluations with the regional authorities, as about *objectively identifying causalities* or the counterfactual. The perceived role of the evaluator thus changes slightly from that of objective observer to simply another actor in the programming process.

⁵ Note that the financing period for new Member States is shorter than the normal financing period due to their membership as of 1/5/2004.

⁶ Based on a personal overview of the evaluation exercise in six Member states (S, SF, UK, F, D, and E).

The role of the evaluator and decision-making needs

Authors like Fetterman (1996) discuss participatory approaches in what is called an 'empowerment evaluation'. This type of evaluation implies that the evaluators collaborate, in partnership, with the participants to facilitate a sort of auto- or self-evaluation. Indeed, when questioning the regional partnerships on the positive side of the evaluation exercise, it is invariably mentioned that the most appreciated effects were the possibility of discussing – iteratively or 'ping-pong' style - the programme with an independent assessor and thus benefiting from the ability to step outside the day-to-day management situation. It should be noted that partnership implication in the evaluation is mentioned in the EU-regulations (monitoring, responsibility, provision of data etc.,) while the use of independent evaluators is also explicitly mentioned as an obligation in the recent Structural Funds Regulations (1260/1999 art 42 or 1083/2006, art. 47). In any case, this type of participatory method has become the norm (DATAR 2003). Judging from the experiences of the current *ex-ante* evaluation round, this conclusion still holds true.

This however begs the question of how independent the evaluator really can be, given that his paymasters have certain, political and economic interests to cater for within the context of a programme. Independence is also limited by the mandate given in the call for tender i.e. what is to be evaluated is mostly decided by the buyer, not the evaluator. In addition, in EU regional policy, the client (or rather its handling of the programme) is at the same time the subject of the evaluation.

Independence is then, to say the least, difficult to define. There are of course minimum criteria i.e., that the evaluator needs to come from outside the authority managing the programme, this is also the meaning of the words "independent assessor" as mentioned in the Council Regulations. Indeed, the latest Regulation also specifies this role further (CEC 2006, art 47): "Evaluations shall be carried out by experts or bodies, internal or external, functionally independent of the authorities referred to in Article 59(b) and (c). The results shall be published according to the applicable rules on access to documents". Further on this subject, in the minutes of a conference bringing together all Objective 2 regions in France in April 2003 (DATAR 2003), this independence of the evaluator is understood as meaning "impartial conclusions" specifying that (inevitable) controversial issues are to be presented in a balanced way ("*de façon équilibrée et équitable*", p. 74). This then is what evaluator independence, mixing theory and practice, boils down to.

However, a theoretical discussion on the meaning of independent evaluators has a tendency to obscure more pertinent questions when it comes to programme evaluation. It could be argued then that there are two key ques-

tions for evaluations in general and for the role of evaluators in particular, and they are:

1. To what extent is evaluation being genuinely utilised to inform the development and implementation of policies?
2. To what extent are evaluation findings being used to support politically driven priorities?

The elected representatives at the heart of regional policy prioritization processes and, in many countries e.g. the Nordic ones or France, are also at the heart of project selection and programme management. They evolve in a context in which a constant stakeholder debate is taking place.

To use the social-constructionist paradigm: policy development is a process of deliberation (Dryzek 1990). This process compares and weighs beliefs, principles and actions under multiple perspectives. This means, in short, that policy learning and decision-making involve a socially conditioned argumentative process (Parsons 1995). In evaluation practice, this means that evaluators are participants in the policy discourse but they are only one of many that have legitimate call on the policy makers' attention.

In concrete terms, and to respond to the two questions above, in this context evaluation is genuinely used to implement further policies but *only partly or selectively*.

Is this a problem? Not really, since if an evaluation is produced in a participatory and transparent manner it will in any case become an important part in the process of deliberation. This was also the conclusion of a meta-evaluation of the Structural Funds evaluations in Finland (Uusikylä and Virtanen 2000). In short, the information the evaluation contains will be used differently by different stakeholders opening up a reasoned dialogue concerning the achievements and objectives of the programme. An evaluation will also, if produced in a transparent manner, generate a basis for argumentation between stakeholders which in turn decides the degree of pressure for change as well as the framework inside which conflict is dealt with or consensus is built (see e.g. Valovirta 2002).

This may, however, signal a change for evaluators. Authors such as Majone (1989:xii) note that evaluators need to: "learn rhetorical and dialectical skills [...] to define a problem according to various points of view [...] and to adapt the argument to the audience [...]". See also Maynard (2000) and his article: 'Whether a Sociologist, Economist, Psychologist or Simply a Skilled Evaluator' on this subject. Other authors, such as Eric Monnier (1992), call this evolution the development of "pluralist evaluations" in order to face up to and assess complex socio-economic situations.

Older quasi-objectivist modes of evaluation would assume that it is possible to derive the truth on the effectiveness of policy measures (Bernstein 1983, Van der Knaap 1995). In practice, it is, more than presumptuous for an outside evaluator, often not familiar with the region or the programme to any great extent, to believe that this is possible.

Indeed, all this points to an evolution in the role of the evaluator from the traditional “neutral observer” or even “objective technician” to that which Jenkins-Smith and Sabatier call an “advocate” of a certain viewpoint (1993:4).

It should however be noted that most programmes are the fruit of a fierce negotiation between stakeholders. Indeed, most authors accept that evaluation is a social process that needs a participatory approach (Gregory 2000). In

this “partnership” type of evaluation, the evaluator does indeed advocate segmented or certain viewpoints that are (or not) taken aboard as the planning proceeds.

The point of advocacy and partnership is valid in all the evaluation stages (*ex-ante*, ongoing, mid-term or *ex-post*) where the evaluator needs to present the findings to the client, making proposals for improvements which then have to be implemented either directly or in the near future (i.e. *ex-post* feeding into new activities). It should however again be noted that in the European regional policy context, these proposals are not binding but are, in line with the argument developed above, one part of the input decision-makers receive but also need to develop, to continue or improve programmes.

Conclusion

To conclude, evaluation has become the very basis for an argumentative process which, at its best, makes regional policy and regional programmes more effective and more transparent. Indeed, the best results in the evaluation field are those where the evaluator works alongside the partners rather than as a distant neutral observer/technician coming in towards the end of the process to proffer an “independent viewpoint”.

In the end, what really matters is to acknowledge the limits of evaluation and, most of all, to reduce the pitfalls of a “bad” evaluation to the greatest extent possible, through the following steps, enlightened by evaluation theory and in many cases, confirmed by practice:

Commitment to monitoring and devising indicators that is both realistic and sufficiently limited in number to ensure the actual possibility of inputting data in the databases, and monitoring systems. Actually, understanding social complexity needs good indicators but also reflexive or perhaps one could call it intelligent or “smart-monitoring”.

Independent evaluators not reinventing the wheel, rather ensuring that all the available relevant research and evaluation evidence is thoroughly reviewed and synthesised before starting anew. This would improve the scale

and quality of evidence about what works and what does not in particular circumstances.

Regional knowledge of evaluation needs. Knowing what is required from the region or the purchaser in question is crucial for successful evaluation. Otherwise, the evaluator will embark on a lonely journey and nobody will be particularly happy with the outcome. Knowledge from the purchaser’s side also means making good quality information available to the evaluator.

Transparent and reasoned follow-up of an evaluation. This means being reasonably open about why this recommendation was used, and why another was not used.

Mix of qualitative and quantitative approaches. The evaluation needs differ from case to case but both approaches are always needed to generate a useful product.

Examples of good practice or solutions from outside the region or even the country should be included in the evaluations when particular problems are identified. Simply put, a quality evaluation should propose some *new* ideas.

All of this will then help to produce *useful evaluations* for the regions which is, in the end, what policy-makers, theorists and practitioners are all (or at least should be) looking for.

ANNEX – LIST OF EVALUATIONS, ESTONIA 2006

1. Evaluation of implementation system of the SPD and its effectiveness (Riikliku Arengukava rakendussüsteemi ja selle toimivuse hindamine) - commissioned by the Managing Authority and carried out by PricewaterhouseCoopers (finished in July 2006).
2. Evaluation of the structure and implementation of the SPD and project selection criteria (Riikliku Arengukava ülesehituse, rakendamise ja projekti va-

likukriteeriumite hindamine) - commissioned by the Managing Authority and carried out by CyclePlan (finished in July 2006).

3. Evaluation of Priority 1 of the SPD (Riikliku Arengukava 1. prioriteedi hindamine) - commissioned by Ministry of Social Affairs and carried out by Ernst & Young (finished in June 2006).
4. Evaluation of the Fisheries instrument in the SPD, 2006.

References

- Amin, A. and Hausner, J. (1997) *Beyond Market and Hierarchy: Interactive Governance and Social Complexity*. Cheltenham: Edward Elgar.
- Bernstein, R. (1983) *Beyond Objectivism and Relativism: Science, Hermeneutics and Praxis*. Oxford: Basil Blackwell.
- Beutel, J. (2002), *The economic impact of objective 1 interventions for the period 2000 – 2006, final report to DG REGIO*, May.
- Bovaird, A. G. and Gregory, D. (1996) 'Performance Indicators: The British Experience', in A. Halachmi and G. Bouckaert (eds.) *Organizational Performance and Measurement in the Public Sector*. Westport, Connecticut: Quorum Books.
- CEC, (1999), *Orientations pour les programmes de la période 2000-2006, COM (1999) 344 final*.
- CEC, (1999b) *Evaluating socio-economic programmes (MEANS Collection)*, volumes 1-6, Brussels.
- CEC, (2000), *The 2000-2006 Programming Period: Methodological working papers, WORKING PAPER n° 8, The Mid Term Evaluation of Structural Fund Interventions*.
- CEC (2006), *COUNCIL REGULATION (EC) No 1083/2006 of 11 July 2006 laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund and repealing Regulation (EC) No 1260/1999*.
- Clarke, J. and Newman, J. (1997), *The Managerial State: Power, Politics and Ideology in the Remaking of Social Welfare*. London: Sage.
- Conan, M., (1998) *L'évaluation constructive*, Paris, L'aube.
- DATAR, *Fonds Structurels 2000-2006, qualité des programmes, Actes du séminaire de Dijon avril 2003*, Paris.
- Davies, S. Nutley, P. Smith, eds., (2000), *What Works? Evidence based policy and practice in public services*. Bristol: The Policy Press.
- DfEE (2000) *Influence or Irrelevance: Can Social Science Improve Government?* London: Department for Education and Employment.
- Dryzek, J., S. (1990) *Discursive Democracy: Politics, Policy and Political Science*. Cambridge: Cambridge University Press.
- Estonian ministry of finance, *Estonian State budget book*, Tallinn, 2006.
- Gregory A., 'Problematizing Participation: A Critical Review of Approaches to Participation in Evaluation Theory', *Evaluation*, March 2000, Vol 6, no 2, pp. 179-199.
- Gren, J., (1994), 'EG:s Regionalpolitik', in Nisser, C., *Europapraktika*, Stockholm, SNS.
- Gren, J., (1999), *The New Regionalism in the EU: lessons learned from Catalonie, Rhône-Alpes and West Sweden*, Östersund, SIR.
- Gren J., *Regional Benchmarking*, Gothenburg University CERGU working papers n°01:1, 2001.
- Gren, J., (2002), *Den Perfekta Regionen*, Lund, Studentlitteratur.
- Henkel, M. (1991) *Government, Evaluation and Change*. London: Jessica Kingsley.
- Jenkins-Smith, H. C. and Sabatier, P. A. (1993) 'The Study of Public Policy Processes', in P. A. Sabatier and H. C. Jenkins-Smith (eds.) *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder: Westview Press.
- Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. New Haven: Yale University Press.
- Maynard R.A., 'Whether a Sociologist, Economist, Psychologist or Simply a Skilled Evaluator: Lessons from Evaluation Practice in the United States', *Evaluation*, October 2000, Vol 6, no 4, pp. 471-480.
- McVittie, E., Kim Swales, J., 'Regional Policy Evaluation: Ignorance, Evidence and Influence', *ESRC project An Analysis of National and Devolved Economic Policies*, 2003.
- Monnier, E., (1992), *Evaluation de l'action des pouvoirs publics*, Paris, Economica.
- Nutley S and Webb J.J., (2000), 'Evidence and the policy process' in H.T.O. Davies, S. Nutley, P. Smith, eds., *What Works? Evidence based policy and practice in public services*. Bristol: The Policy Press.
- OECD (1997) *In Search of Results: Performance Management Practices*, Paris, OECD.
- Parsons, W. (1995) *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*. Cheltenham: Edward Elgar.
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: Sage.
- Pawson R., 'Evidence-based Policy: In Search of a Method', *Evaluation*, April 2002, Vol 7, no 2, pp. 157-181.
- Sanderson, I. (2000) 'Evaluation in Complex Policy Systems', *Evaluation*, Vol. 6, No. 4.
- Senge, P. (1996), 'The Leader's New Work: Building Learning Organizations', in K. Shaw, I. (1999) *Qualitative Evaluation*. London: Sage.
- Uusikylä P, Virtanen P., 'Meta-Evaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland', *Evaluation*, March 2000, Vol 6, no 2, pp 50-65.
- Valovirta V., 'Evaluation Utilization as Argumentation', *Evaluation*, January 2002, Vol 8, no 1, pp. 60-80.
- Van der Knaap, P. (1995) 'Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation?', *Evaluation*, Vol. 1, No. 2, pp. 189-216.
- Walker R., 'Great Expectations: Can Social Science Evaluate New Labour's Policies?', *Evaluation*, July 2001, Vol 7, no 3, pp. 305-330.
- White, M. (1999) *Evaluating the Effectiveness of Welfare to work: Learning from Cross-National Evidence*, in C. Chitty and G. Elam (eds.) *Evaluating Welfare to Work*. In-House Report No. 67, London: Department of Social Security.
- Wilson D C (1992) *A Strategy of Change*, London, Routledge.

Sustainable Development as a Horizontal Theme in European Evaluations

By Keith Clement

Abstract

Sustainable development evaluation assumes different priorities for different actors. For policy-makers, the concerns include identifying the efficiency and effectiveness of policy implementation, whether economic measures have delivered increased sustainability, and the extent to which policy integration has been achieved. For researchers or evaluators, the focus is wider, ranging from the conceptualisation of sustainable development (SD), the strengths and weaknesses of the methodologies utilised, and whether SD has been carried through to project implementation. For the wider public, the priorities relate more to understanding the concept and its implications, for example whether job-creation has associated costs in terms of loss of heritage, whether the community benefits in broad terms, and whether electoral promises have been delivered.

These differing perspectives lead to conflicts in expectations that raise normative and ethical issues. The main dichotomy is between the economic imperative that generating employment is the immediate and overriding purpose of regional development, and the sustainability perspective that a broader picture is necessary, that long-term implications should be explicit, and that additional account should be taken of natural and social capital. Counter arguments include that maintaining an attractive environment may reduce scope for income generation and result in out-migration.

In practice, client-evaluator confidentiality may present

an ethical dilemma. Not all reports are published, policy-makers can be selective about which 'truths' to reveal, and the client rather than the beneficiaries usually determine the orientation of a report. Moreover, when the client comprises a partnership, the lead (economic) partners may attribute low priority to sustainable development and restrict consultation on this theme, placing the SD evaluator with a choice between objectivity and advocacy.

Nordic regional initiatives approximating to sustainable regional development have been characterised by diversity with no common frame of reference, and with evaluation seeking comparable data on projects that differ in substance, sector, timeframe and funding sources, encompassing bottom-up, top-down, and even middle-down initiatives. In addition, Nordic sustainable development activity has been concentrated at national and local levels, with an absence of specific regional guidance against which to measure progress.

By contract, the Structural Funds are distinctive in the opportunities they present for sustainable development evaluation. Specifically, the Funds have a pre-determined operational framework characterised by committees, conventions, programming documents, territorially-defined assisted areas, and EU guidelines and regulations that indicate the emphasis to be placed on sustainable development. In practice, individual Structural Funds programmes have made considerable progress in defining and performing SD evaluation.

Introduction

As a topic for evaluation, SD has become much more evident in recent years, supported strongly by the European Commission in its economic development activity. Characterised by a range of policy documents, legislation, cross-national studies and thematic networks, the relevance of SD for regional economic development has steadily grown in scale, and clarification of its role within and as evaluation practice has become more urgent. However, there is still an on-going concern with the refinement of SD as a concept, accompanied in some professional contexts by resistance to its legitimacy as an all-encompassing means of appraising and steering development activities.

In considering SD as a horizontal theme in European evaluation, this chapter addresses issues raised within the academic community and in the practical sphere of policy

instrument design and evaluation. In content, the chapter has four sections. First, the characteristics of sustainable development are described, accompanied by a brief history of its rise and identification of its strategic status within a Nordic policy context. Second, differing perspectives on SD evaluation are considered, introducing the principal actors and normative and ethical issues encountered. Third, experiences of SD evaluation in regional development initiatives are reviewed, drawing examples from two Nordregio projects researching sustainable regional development in the Nordic countries and appraising SD in the Nordic Structural Funds. Each case considers methodology, evaluation challenges and lessons for practice. Finally, conclusions are drawn from the material presented.

Sustainable Development Characteristics and Strategy

Sustainable development has matured from a theoretical concept into practical political guidelines with impact on every level of governance, from the decisions of the United Nations down to the activities of local communities (Kneucker, 1998, p11).

With its present status and scale of activity, SD has a substantial presence and potential. Extending beyond economics to encompass ethical, societal, institutional and environmental dimensions, SD is very difficult to encapsulate in simple terms. It does not represent an attainable fixed state of harmony; instead, it comprises a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development and institutional change are all made consistent with future as well as present needs (World Commission on Environment and Development, 1987, p9).

SD exists as an aspiration or goal that is intuitively understood and approved, but which evades precise definition and has conceptual and practical problems. Its appeal has been attributed to both its breadth and its vagueness, judged as 'palatable to everybody...radical and yet not offensive' (Skolimowski, 1995). However, differing interpretations reflect the wide-ranging variety of visions people have for the future:

Since the Brundtland Report (re-)introduced the concept in 1987, at least forty working definitions of sustainable development have appeared. Consequently, many different projects are furthered under the flag of sustainable development, and quarrels have started to emerge about what sustainable development really is (Hajer, 1995, p1).

No consensus has emerged on a single, practical definition that might bring together economists, ecologists, political scientists and philosophers (Crabbé, 1997), but incremental refinements have drawn useful definitional distinctions. In particular, they have encouraged support for the idea that 'it is both morally and economically wrong to treat the world as if it were a business in liquidation' (Daly, 1992).

Fundamentally, sustainable development is perceived as process-oriented, referring to the process of developing the planet in a sustainable manner (Reed, 1996). The corresponding paradigm is generally considered to have three key components of economic sustainability, environmental sustainability and social sustainability. Economic sustainability encompasses growth, development, productivity and trickle-down effects; environmental sustainability includes ecosystem integrity and attention to carrying capacity and biodiversity; and social sustainability includes variables such as equity, empowerment, accessibility and participation.

However, bringing these multi-faceted elements together requires effective institutional change – sometimes cited as the fourth dimension of sustainable development – as well as interdisciplinary vision. Consequently, realisation of the new paradigm implies a major change in focus for institutions and the development of practical techniques. Several major principles have been identified to support this transition:

- in contrast to the conventional development focus on human-made capital, emphasis should be placed on natural capital as the main limiting factor;

- a sustainable development index should supersede the conventional indicator of gross national product (GNP) to measure development performance in the context of an integrative framework of social, environmental and economic sustainability;
- intergenerational assumptions should feature in the assessment of resource availability;
- waste absorption should be recognised as a major function of the environment and an important limitation to economic growth; and
- methodological tools should be developed to evaluate programmes and projects and to assist investment and planning decisions by giving equal weight to economic, environmental and social variables (Khan, 1995).

The subsequent stage of operationalising SD has significance for decision-makers that would use the concept for policy design and policy evaluation. Distinctions have been drawn between weak, strong and even 'absurdly strong' sustainability, highlighting the scope for conceptual flaws at different levels of integration. Ethical dilemmas have also been anticipated, such as how to decide when trade-offs are feasible between an environmental satisfaction and a human satisfaction (Beckermann, 1995).

After two decades of competing definitions, new questions and alternative techniques of measurement, sustainable development has become ubiquitous, and yet it continues to evolve:

Sustainability still is a somewhat elusive concept midway between politics and science. The discussion about it is old and new at the same time...It is becoming obvious that sustainability is not only a scientific term but also, and prominently so, a political one (Schleicher-Tappeser, 1998, p14).

Acknowledged as probably the most demanding policy concept ever developed and promoted, SD is considered to require a revolution in thinking, a new view on old problems and approaches, and a new mechanism of governance tailored to meet the increasing quality demands arising from the concept (Spangenberg and Giljum, 2005). From this perspective, SD policies require synergies, an emphasis on integration, and a focus on inter-linkages rather than on individual policy domains, and appropriate trade-offs, compromises and compensations within policy packages. Corresponding SD strategies require better government, a sharing of tasks between state, society and the business sector, long-term perspectives instead of short-term activism, and vision instead of pragmatism.

As a global political theme, SD has been compared with concepts such as democracy and human rights, and it has scaled the agenda in national politics and within supranational institutions. In 1997, the Special Session of the

UN General Assembly set the target that all countries should have strategies or programmes for sustainable development by 2002 for the UN World Summit on Sustainable Development in Johannesburg. National responses to this commitment varied in quality and depth, but securing SD was now clearly on the agenda as a governmental task.

Within the European Union, sustainable development became an overarching objective in 1997, when it was included in the Treaty of Amsterdam. Subsequently, at the Gothenburg Summit in June 2001, the first EU Sustainable Development Strategy (SDS) was launched. Whereas the Lisbon strategy focuses on employment, economic reform and social cohesion, the SDS adds an environmental dimension and establishes a new approach to policy-making. In June 2006, the European Council adopted a renewed SDS.

In the Nordic countries, conditions are considered to be especially favourable for sustainable development. This is attributed to political stability, close co-operation between countries, well-functioning societies based on the rule of law, financial markets, business and industry, agriculture, and populations with a high level of education and training and a capacity to adjust (Nordic Council of Ministers, 2001, p10). Demonstrating acceptance of SD at the highest political levels, the Prime Ministers of the Nordic countries and the political leaders of the self-governing areas adopted a declaration in November 1998, entitled *A Sustainable Nordic Region*. Within this document, a set of eleven goals was agreed (see Table 1) to guide the development of the region, while simultaneously steering sustainable development in the adjacent areas.

A strategy to fulfil these goals has been operational since 1 January 2001 and updated in 2004. It focuses on common interests between the Nordic countries where they are well suited to contribute towards SD, and where co-operation can create particular added value. The transition towards sustainable development is expected to bring considerable advantages for economic development, competitiveness and employment, delivering new technologies and competences, and potentially leading to the creation of new markets for Nordic products and services. Realisation of this vision presupposes the active participation of actors at all levels, including local, regional and national governments, business and industry, and NGOs.

The first version of the Nordic Strategy for Sustainable Development contained qualitative targets and measures for the period 2001-2004, as well as long-term goals. The revised version describes the goals and initiatives to which the region is committed between 2005 and 2008 and also contains long-term development goals for the Nordic region for the period up to 2020. A parallel set of indicators, *Focus on Sustainable Development: Nordic Indicators 2006*, was endorsed by the Nordic Council of Ministers in June 2006.

-
- Present and future generations must be assured a life in safety and good health.
 - A sustainable society must be based on democracy, openness and participation in local, regional and national co-operation.
 - A high degree of awareness concerning the measures and processes leading to sustainable development must be created in society.
-
- The principles of sustainable development should be integrated into sectors on an on-going basis.
 - The role of indigenous population should be emphasised when promoting sustainable development.
 - Biological diversity and the productivity of ecosystems must be maintained.
 - Emissions and discharges of pollutants into the air, soil and water must not exceed the carrying capacity of nature.
 - Renewable natural resources must be used efficiently and within their regeneration capacity.
 - Non-renewable natural resources must be utilised in a manner protecting natural systems, and renewable alternatives must be developed and promoted.
 - Over the long term, xenobiotic substances and substances harmful to humans and nature must be eliminated.
 - Appropriate innovative thinking should encourage more efficient utilisation of energy and natural resources.
-

Perspectives on SD Evaluation

Principal Actors

Within a regional development context, the perspectives of different participants generate the range of factors to be included in the evaluation of sustainable development. They determine the focus, such as processes or outcomes, and the priority of the SD evaluation. Three principal actors are considered here, those of regional policy-makers, researchers and the public.

For *policy-makers*, the concerns include identifying the efficiency and effectiveness of policy implementation, including whether economic measures have delivered increased sustainability, the extent to which policy integration has been achieved – potentially offering insights into best practice – and whether the results support existing political aspirations. If integration has not been successful, why did it fail? If the results do not support political desires or electoral promises, are there implications for funding continuity? Should the budget be continued or re-allocated? How effective have the policy instruments been in delivering sustainability – should they be modified – or do the results justify policy changes or re-appraisal of priorities? Are the SD outcomes presented in a manner that emphasises the points that policy-makers want to pursue? To the extent that there is a specific purpose behind an evaluation, this could mean that certain results are under-emphasised.

For *researchers*, the focus is broader, addressing the con-

ceptualisation of SD and whether it has a strategic foundation with theoretical perspectives based on specialist literature or national and international visions of integrated development. What form of political ideology or governmental administration is being imposed through this concept overtly or covertly? Thereafter, research questions may examine the strengths and weaknesses of the methodologies utilised to encompass the elements of the SD concept, and latterly whether SD has been carried through to project implementation. Was there a hierarchical relationship between environment and SD, and were integrative techniques such as a SD index utilised? Researchers may also consider the role of the policy instruments, whether their operation in practice steers or facilitates the process, and whether there are associated implications for outcomes.

For the *wider public*, the priorities relate more to understanding the concept – essentially, what is sustainable development? Does it mostly concern environmental protection, and does job-creation come with associated costs in terms of loss of amenities such as landscape, protected areas and natural heritage? What difference does a SD initiative make, is the community benefiting in broader terms, and how is its impact different across the region or localities? Does this correspond with electoral promises? Overall, if there is clear progress towards sustainable development, the community will be interested in the inclusiveness of the outcome and what it means for the future.

Normative and Ethical Issues

In practice, the differing perspectives lead to conflicts in expectations that encounter normative and ethical issues. The main dichotomy is between the economic imperative that generating employment is the immediate and overriding purpose of regional development, and the sustainability perspective that a broader picture is necessary, where long-term implications should be explicit and additional account should be taken of natural and social capital. Where does the greater responsibility lie? Clearly, job-creation is designed to alleviate economic disadvantage, with measures aimed at stimulating new employment, teaching transferable skills, marketing techniques and innovative thinking, and launching new companies. Within such a scenario, conventional arguments include that investing in the environmental dimension of SD is at best a drain on financial resources – “*substantial environmental protection will have its economic price*” (Neumayer, 1998, p.39) – and at worst a block on growth and income generation that invariably results in out-migration. In this context, environment and SD – often perceived as the same thing – have the image of being a barrier to development.

In efforts to convince economic actors of the merits of following SD approaches, the *language* of sustainable development may need to be supplemented with the *logic* of sustainable development (Clement, 2000). It is easier for regional actors to identify with direct benefits such as financial savings, greater efficiency, sectoral interaction and international competitive status, than to relate to long-term global concerns as a justification for diverting regional financial priorities. The European Commission has already sought to resolve this dichotomy by highlighting mutual benefits arising from environment-economy interaction:

Activities related to the environment can be a substantial source of employment and create major opportunities for SMEs to provide goods and services...(and)...the environment is an excellent justification for diversification as regards both activities (in rural areas, for example) and new skills and qualifications for the workforce (*Commission of the European Communities, 1996, p.20*).

Another normative factor that characterises the scope

of evaluation relates to the distinction between sustainable development and sustainable growth. In some instances, sustainable growth is assumed to be equivalent to sustainable development, whereas on other occasions it is presented as a compromise, with growth being one of the essential aspects of development. However, rather than being restricted to a linguistic distinction, growth of capital in one field is often at the cost of resources in another field, resulting in a condition of unsustainable development, not balanced development. Accordingly, the terminology used by actors can mask very different visions and anticipated outcomes behind phrases that sound similar.

Ethical dilemmas can occur within the sphere of client-evaluator confidentiality. Once an evaluation is complete, the client may have the option to disseminate the results more widely, for example to inform the public. However, not all reports are published, and in those that are published policy-makers can be selective about which ‘truths’ to reveal. From the outset, the clients rather than the beneficiaries usually determine the orientation of a report, and when the clients comprise a partnership, the lead (economic) partners may attribute low priority to sustainable development, restricting consultations, workshops or discussion on this theme. This can place the SD evaluator in an intermediary role – possessing knowledge that is not considered appropriate for wider dissemination – and a choice between objectivity and advocacy, which encourages critical perspectives and pro-active consideration of sustainable development for communities.

For an evaluator in this position, the dilemma might involve interaction on which materials should be withheld, emphasised, or in certain cases given anonymously to beneficiaries or the media. Basically, to whom is the evaluator responsible in divulging the results of the evaluation? The public sector client (which could be local, central or regional government, or a consortium of agencies) or the beneficiaries of a regional development project? Defining the evaluator’s role within this framework may determine what is published, the phrasing and text of reports, and long-term expectations of the process with regard to participation, effectiveness and outcomes.

Regional Development Initiatives: Approaching SD Evaluation

To illustrate SD evaluation in practice, this section draws upon experience of two evaluations carried out by Nordregio. The first case study relates to the evaluation of sustainable regional development projects, and the second comprises an appraisal of Structural Funds programmes.

Both are focused on the Nordic countries (see Table 9, and Appendices 1 and 2). Using a common format, the following text considers the challenges encountered in evaluating SD in these projects and identifies lessons for practice.

Project 1: Sustainable Regional Development: Learning From Nordic Experience

Nordic regional initiatives approximating to sustainable regional development are characterised by diversity, with no common frame of reference. The evaluation method must seek comparable data that allows meaningful insights into projects that differ in substance, sector, timeframe and funding sources, and which encompass bottom-up, top-down, and even middle-down initiatives. Furthermore, Nordic sustainable development activity has been concentrated at national and local levels, with an absence of specific regional guidance against which to measure progress.

Project 2: Assessing Environment and Sustainable Development Integration in the Nordic Structural Funds

The Structural Funds are distinct from other regional development initiatives in the opportunities they present for SD evaluation. They have a pre-determined operational framework characterised by clear budgets, timeframes, deadlines, committees, conventions and programming documents, as well as territorial definition of assisted areas. EU guidelines and regulations indicate the emphasis to be placed on sustainable development, and a recent EU thematic evaluation on the SD contribution of the Funds outlines methods and expectations for current and forthcoming programmes.

Sustainable Regional Development: Learning From Nordic Experience¹

Commissioned by the Senior Officials Committee for Regional Policy within the Nordic Council of Ministers, this project formed part of a series designed to identify and facilitate realisation of sustainable regional development (SRD). It was intended to contribute to a field already characterised by expanding thematic networks, international comparative studies and EU evaluation research, all of which sought to integrate sustainable development principles into regional development practice. The project's three objectives comprised deriving a detailed knowledge of factors impacting on SRD realisation, identifying and highlighting best practice in Nordic SRD, and facilitating a Nordic network for SRD information exchange.

Seeking specifically to learn from Nordic experience, the project looked in depth at seven case studies in Denmark, Finland, Norway and Sweden. In the research design, the criteria for choosing these case studies was based not on similarity, but on distinctiveness. Each exhibited interesting features, for example in the logic of methodology and its scope for wider replication or in the insights gained that could offer lessons for other regions or national contexts.

Methodology

Drawing on the broad horizontal overview established in previous research², the case study survey method involved

interviews with participants from a number of different levels, as well as from parallel organisations. Depending on the specific project, the range of meetings included a combination of the following actors:

- National government officials
- Regional and local government staff
- Steering groups
- Politicians
- Project managers
- Fieldworkers
- Project beneficiaries
- Non-governmental organisations
- Evaluators
- Researchers and academics

Of the seven case studies in the final selection, structured interviews were conducted with 79 individuals, using a questionnaire that addressed the development process, the operational phase of each initiative, and the identifiable achievements. The methodology was qualitative: identifying key actors in the interactive process, distinctive characteristics, strengths and weakness, progress in relation to objectives, barriers encountered and overcome, and lessons learned.

Thereafter, going beyond first-stage evaluation, the detailed analysis of projects formed the basis for a synthesis of significant features. Groupings of positive elements were drawn especially from the identified strengths and

¹ Clement, Hansen and Bradley, 2003

² Clement and Hansen, 2001

lessons in the case studies, offering insights into best practice and allowing scope to create strategic benchmarks (see Table 10). This was intended to serve as a guide to key factors that have characterised, or been critical within, successful projects or successful elements of projects both through realisation (strengths in implementation) and aspiration (project lessons).

This activity corresponded to generic benchmarking – comparing development initiatives that extend into different fields or sectors and rely upon different instruments (Bendell et al, 1998) – identifying leading practices for adoption and improvement. In this context, a benchmark is a standard against which actions or processes can be compared or measured, to understand what to aspire towards in future. Accordingly, benchmarking leads to specific performance targets. Drawing on the project experiences, over 50 qualitative performance measures were derived, corresponding to the phases of project design (feasibility prerequisites), implementation (framework conditions and methodology) and evaluation (impact and outcomes).

Evaluation Challenges and Lessons for Practice

The conduct of the SRD evaluation encountered a number of problems at different stages in the process. This sub-section identifies the difficulties and – where possible – the attempts made to resolve them.

- *Nordic SD activity had been concentrated at national and local levels, and there was no specific*

regional guidance against which to measure progress. Consequently, SRD initiatives were characterised by diversity, with no common frame of reference.

The absence of specific regional SD guidance meant that an external framework to review progress was not available. No models could serve as examples to follow or avoid, and no dedicated funding sources were available to shape project design. Very different cases can offer interesting variety, but they also create difficulties in attempts to impose common structures within fieldwork analysis. The selected projects differed both in their encapsulation of SD within a regional context – varying from an implicit idea to definitions quoting the Brundtland Commission – and in the type of measures adopted for its promotion.

To resolve this incompatibility, a working definition of SRD was defined for the evaluation. Essentially the integration of sustainable development principles into regional development practice, SRD was assumed to encompass regional policy activities, instruments or institutions that both recognised the importance of SD and incorporated economic, social and ecological sustainability into a development context.

- *Projects differed in substance and sectoral orientation.*

Finding projects that met the SRD criteria was challenging, but then achieving direct comparability was even more difficult. Problems arose for various reasons, especially from actors perceiving qualitative issues from different professional or sectoral perspectives.

Table 10. Framework for Deriving SRD Benchmarks³

Action	Phase		
	Data Collection	Data Analysis	Implementing Change
Inputs	Definition of SRD within regional policy instruments, institutions and processes	Analysis of case study characteristics, assimilating strengths, lessons and aspirations	Identification of relevant benchmarks to use as reference points in project design, targeting or management
	Selection of case studies and conduct of interview programme		
Outputs	Overview of European SRD literature and practice in the Nordic countries	Derivation of qualitative SRD strategic benchmarks associated with project phases	Modifying actions in ongoing projects, and designing new projects to meet or exceed benchmarks
	Comparable project data in structured case study format		

³ Source: Clement, Hansen and Bradley, 2003, p95

In practice, direct comparability proved too ambitious and was not pursued. Instead, the approach adopted was to identify the strengths within projects, to build up a catalogue of ideas and initiatives that were considered to have brought success, and to adopt a generic approach that could be useful in generating a modified form of benchmarking.

- *Perceptions of SD progress differed between individuals even on the same project.*

In the multiple-interview method – with between 10-12 interviews for each case study – it proved difficult to obtain a common view on strengths and weaknesses in project delivery. For example, in one project, several participants considered that sustainability had been integrated into the process, with a lasting presence, whereas others maintained that no real progress had been made, and that any evident SD content was already in place prior to project commencement.

In this case, the qualitative approach formed part of the problem. To verify the statements in either direction would require further data gathering with a quantitative character, for example identifying indicators such as financial commitments, new processes or new instruments that could be measured as evidence of SD progress. This was beyond the scope of the evaluation timeframe and budget, and accordingly the reporting process reflected and interpreted the predominant view as portrayed by the majority of meetings.

- *Projects were variously described as bottom-up, top-down, and even middle-down initiatives, with different levels of administration claiming the same positive impacts as their own achievements.*

Tracing the origin of ideas and innovations proved difficult because of the historical perspective, and the competition between different administrative and governmental levels presented an additional evaluation barrier. Events were either remembered differently or reported differently. Who made the intervention to initiate an SRD project, who thought of the innovative approach to SD, and how important was the political leadership to the SD momentum within the project? Opinions frequently differed on such topics.

Again, there was no scope within the project timescale and budget to verify these issues. A larger interview sample would have gained greater numbers, but that may not have been conclusive, as people can recall meetings differently, in some instances even confusing the sequence of events. In other cases, the drive for political kudos may have further distorted the factual presentation of events sought by researchers.

Difficulties in estimating the success of projects were

compounded when no measurable SD objectives had been identified at the outset.

Without external frameworks, projects had to be measured against their own SD objectives. However, these were often vague or expressed in general terms, and without measurable objectives, interviewees had difficulty deciding whether their efforts had been successful. For example, how could they then distinguish between success and ‘partial success’?

In this instance, the general balance and weight of evidence through the interviews was relied upon as conveying the perceived success of projects. If those involved in implementation considered that project achievements had met expectations, overcome problems and derived lessons that informed subsequent or continuing projects, these factors were drawn upon as verification.

- *Establishing qualitative rather than quantitative benchmarks required a modified approach.*

Benchmarking generally produces numerical or quantitative results that are used to measure and compare performance. In this format, comparability and comprehension are straightforward and the standards that emerge are applicable in a range of regional and national contexts. This level of detail and precise comparability was not possible in a SRD qualitative analysis.

In response, the methodology was adapted to create qualitative strategic benchmarks. This still fulfilled the benchmarking criteria of providing ostensible proof of alternative approaches and a logical basis for prioritising opportunities for such change. It was oriented to identify best practice, allowing organisations to select benchmarks most relevant to their operation, and to set performance targets for future aspirations.

- *Clear boundaries between SRD benchmarks and project management benchmarks were difficult to maintain.*

Characteristics of good project management and successful SRD project implementation overlapped in some instances, meaning that for example a number of benchmarks for feasibility prerequisites, framework conditions or methodology were presented as common to both activities. Rather than offering an exclusive focus on SRD techniques, the resultant benchmarks encompassed broader concerns.

This overlap was accepted as inevitable in the design of guidance for effective projects, not least for SRD, and to withdraw essential criteria would have undermined the foundations and direction set by the range of identified benchmarks. In other words, a partial set of benchmarks would have been inadequate for SRD evaluation, and therefore the range necessarily included both SRD and

project management factors.

Assessing Environment and Sustainable Development Integration in the Nordic Structural Funds⁴

The objective of this project was to review the programming documents relating to the 2000-2006 round of Structural Funds programmes in the Nordic countries, from both qualitative and quantitative perspectives, with regard to the integration of environment and sustainable development.

In contrast to independent regional initiatives, actions within the Structural Funds must conform with EU guidelines and regulations that indicate the emphasis to be placed on sustainable development. Supporting integration, the European Commission published a thematic evaluation on the SD contribution of the Funds, in which it outlined expectations for forthcoming programmes (Commission of the European Communities, 2002).

In practice, the Structural Funds are very distinct from other regional development instruments in the opportunities they present for evaluation. They have a pre-determined operational framework, which is characterised by clear budgets, timeframes, deadlines, committees, conventions and programming documents, as well as the territorial definition of assisted areas.

The study appraised 26 published programming documents, at various stages of development. This encompassed Objectives 1 and 2 in Denmark, Finland and Sweden, and transnational programmes within Interreg IIIA and IIIB.

Methodology

The research methodology combined both qualitative and quantitative analysis. In the qualitative appraisal, the environmental focus considered several features:

- the awareness within programmes of EU environmental policy;
- the incidence, scale, content and analytical scope of regional environmental profiles;
- the scale of environmental integration i.e. identifying the consistency of environment throughout programmes, the involvement of environmental specialists in the design process, whether measures directly promoted environmental gain, evidence of budgetary allocations for environmental factors, and the overall effectiveness in achieving environmental integration;
- in relation to programme tools, the extent to which programmes incorporated environmental objectives, goals and targets (whether qualitative or quantitative), the use of environmental impact assessment, and the inclusion of environmental project selection criteria and associated indicators; and
- the relationship between environment and sustainable

development, especially whether a hierarchical structure was evident.

The focus on SD content considered similar orientations:

- the awareness within programmes of EU SD policy and strategy;
- whether SD was defined within programmes, and whether an overall regional SD strategy appeared as part of the programme rationale;
- the scale of SD integration i.e. identifying the status given to SD (for instance, whether strategic objective or target), whether SD was traceable throughout the document in priorities, measures, project guidance and assessment, if there was evidence of SD budgetary allocations, and what methods were used for SD integration.

The qualitative results for environment and SD were compared across programmes for targets, continuity and integration, and then individually for environmental gain and SD strategies respectively.

The quantitative analysis addressed a selected number of these themes, using a scoring system from 0 to 3. For example, in rating the effectiveness of environmental continuity, no citations would score 0, inclusion at project level would score 1, inclusion at measure and project levels would score 2, and appearance at priority, measure and project levels would score 3. Similarly, in rating the effective integration of SD into programmes, isolated/no integration would score 0, poor integration would score 1, good integration would score 2, and total integration would score 3.

Evaluation Challenges and Lessons for Practice

As a document review, the evaluation was subject to a number of limitations related to the availability of materials and interpretation that could not be supported by interviews with programme managers. Other problems encountered are listed as follows.

- *The programming documents varied in their stage of development.*

Not all programmes had been approved, with first and second drafts amongst the source material, and some programmes were to be substantially revised later. It was impossible to estimate how subsequent drafts would approach SD integration, and therefore the assessment could not represent how the final documents would have performed in terms of overall scores.

The approach adopted was to work with the versions available at the time, accepting that they would reflect a range of stages in the programme formulation process. The alternative of continually updating as each new modified draft became available was not feasible as there was insufficient funding in the budget to support such on-go-

⁴ Clement, Bradley and Hansen, 2004

ing monitoring and evaluation.

- *Determining the status of SD within programme strategies proved difficult.*

Partnerships generally did not include lengthy or insightful descriptions of SD or how it would be secured through the programme. Instead, most documents contained an acknowledgement that SD was considered implicitly, and that it should be assumed that SD was both present and integrated into the design process.

Accordingly, the evaluation made interpretations based on references to SD and appropriate use of terminology, estimating the degree of knowledge through inclusion of current policy and legislation. Indicators were sought that, for example, SD was appreciated as a broad integration process, rather than restricted to an upgrade of environmental input (even though that might represent the starting point).

- *No budgetary allocations were made specifically for SD, precluding any financial priority analysis of this theme.*

In some instances, evaluative methodologies suggest using the budget as a lead to identify SD commitment (Ecotec, 1999), but this system can be undermined by reallocation of funds at a later date through the system known as *virement*. This occurs when individual priorities or measures are seen to be under or over-performing, making any initial budgetary analysis misleading. How-

ever, in the surveyed programmes, no allocations were made specifically for SD, and instead the pursuit of SD was generally attributed to being fulfilled by a range of measures, making it impossible for the evaluator to isolate.

Environmental budget allocations were identifiable, but, apart from comprising only one pillar of SD, they would also be liable to change once project proposals were submitted. To analyse SD funding would have required more precise definition of appropriate activity and ideally be performed as an *ex post* evaluation, once funds had been committed and projects implemented.

- *Project selection criteria were often restricted to the programme complement.*

The partnerships were expected to generate selection criteria at a later date, at which point they would be included within the programme complements. This meant that the programming documents focused on major strategic factors and the financial plan. From an evaluation perspective, this hindered a full assessment of the SD integration methodology.

As a compromise, the evaluation focused partly on the environmental profile, appraising the balance of coverage between economic, social and environmental factors and the extent to which data from the profile were used in the programme SWOT, design and justification of measures. In addition, the use of integration techniques such as environment-economy matrices was accepted as supporting SD dimensions.

Recent Developments in the Structural Funds

Lessons from the Current Programming Period

In the current Structural Funds, for 2000-2006, Regulation No 1783/1999 for the European Regional Development Fund (ERDF) frequently repeats the aspiration of a 'balanced and sustainable development' that combines economic activity, competitiveness, high employment, equality between men and women, and environmental protection (Commission of the European Communities, 1999). The ERDF is expected to provide assistance in the framework of a comprehensive and integrated strategy for sustainable development, but the primary focus is on sustainable jobs, sustainable investment and sustainable urban development. Other SD factors include renewable energy and protection of cultural and natural heritage.

In current practice, programmes are required to consider SD issues in their design, implementation and evaluation as a matter of routine; this approach has now become known as mainstreaming. One aim of mainstreaming is to bring about cultural change on the part of a wide range of

stakeholders by making more aware of SD issues in their programme activities. A second aim of mainstreaming is to deliver practical results that lead to sustainable development.

Useful examples of SD progress can be drawn from Scotland, where four of the five Structural Funds Partnerships have established Horizontal Themes Policy Groups to support the mainstreaming process. In the East of Scotland European Partnership, the Group seeks to ensure that the principles of SD are being followed from project design through to practice. It also monitors progress in mainstreaming SD within the programme and develops indicators to measure the achievement of SD objectives and targets. The Group participates in networks promoting sustainable development across other EU regions.

In the Highlands and Islands Special Transitional Programme, guidance for project applicants has emphasised that SD comprises more than the reduction of environmental damage. Instead, it relates to the wise stewardship

of financial, human and natural resources, supporting projects that maximise and sustain economic, social and environmental benefit. The programme states explicitly that project applications that provide no detail on SD actions will not be approved for funding. Six categories are used as a basis to evaluate effectiveness in SD activity:

- Sustainable use of the region's resources, measured by increased efficiency and new uses of existing resources, and increased use of new and renewable resources;
- Sustaining the viability of regional localities by improving access to essential infrastructure, quality support and core services;
- Sustained local commitment to the region's development by increased public participation in the process, measured by strengthened local involvement in decision-making, private sector investment, and communities capitalising on local social and environmental assets;
- Sustaining the region's biodiversity, encompassing environmental stewardship of important land and marine habitats, improved management of key natural resources, integrated rural development, and a reduction of negative environmental impact from development;
- Sustaining the competitiveness of the region's business and labour force, promoting adaptability, innovation and entrepreneurship, encouraging a culture of life-long learning, addressing equal opportunity issues and supporting the integration of SD thinking;
- Economic sustainability, related to the life-span viability of long-term projects and appropriate exit strategies for short-term projects.

The programme's guidance for applicants also identifies extensive SD project selection criteria and a range of programme-specific SD indicators.

Challenges for the Next Programming Period

For the next phase of Structural Funds programmes, for 2007-2013, the new Regulations incorporate much wider reference to sustainable development. The Regulation for the ERDF, No (EC) No 1080/2006, addresses social, economic and environmental actors, with Article 4 on convergence seeking sustainable integrated regional and local economic development and sustainable jobs (Commission of the European Communities, 2006). Eleven priorities are identified, with research and technological development and the information society being accompanied by a range of environmental investments, sustainable tourism, preservation of the cultural heritage, reducing environ-

mental impacts of transport, energy efficiency and development of renewable energies, as well as education, health and social infrastructure.

Under Article 5, which considers regional competitiveness and employment, the ERDF is again expected to focus its assistance in the context of sustainable development strategies. Article 6 on European territorial co-operation is to be aimed at the development of cross-border activities, comprising joint strategies for sustainable territorial development, with priority areas encompassing innovation, environment, accessibility and sustainable urban development.

Beyond the regulations, a consortium of non-governmental organisations (NGOs) has advanced further suggestions related to SD evaluation (BirdLife International *et al*, 2004). The Commission is asked to review the sustainable development indicators employed by Member States as part of their ex-ante, mid-term and ex-post Structural Fund evaluations, accompanied by new guidance on the most widely accepted and recognised indicators. Themes to be assessed by such indicators would include programme outputs (for example, the actual effect of projects on the environment) and procedures (such as the provision of information, awareness of public consultation, and the quality of partnership involvement).

In addition, Structural Funds managing authorities are asked to take responsibility for evaluating project outcomes. Whereas auditors can be called in at any time to check on the financial management of a Structural Funds project, there appears to be no monitoring process to verify whether a project is meeting its other targets (including environmental and sustainability targets). The evaluation of these targets – and their contribution to the achievement of National Sustainable Development Strategy objectives – would be a valuable addition in the new programming period. Post-project monitoring could help to identify beneficial environmental impacts and assist with the future selection of best projects for delivering SD, with lessons learned being fed back into the project-selection process and potentially into a common EU project database.

A further suggestion relates to rewarding good environmental performance. This relates to the Performance Reserve, currently 4% of Member States' Structural Funds commitment, which may be withheld by the European Commission if a programme does not meet its targets. Under the new Regulations, each Member State may establish a national performance reserve to reward the most successful programmes. This reward could be enhanced by adding criteria related to SD, for example assessing a programme's contribution to the objectives of the EU Sustainable Development Strategy.

Conclusions

Sustainable development has progressed from a vague policy concept to become a central theme and task for local, regional, national and supra-national agencies. In evaluation, it was initially approached as a sub-theme within broader exercises, but SD evaluation now has a separate and independent status, and its significance continues to grow for public and private sectors. Nevertheless, it also exhibits the problems typically associated with such transitions, especially in relation to requiring precise definitions, clarification of concepts in different contexts, and methodologies for accurate measurement. These factors affect the activities of the academic, research and evaluation communities.

A number of challenges remain in addressing SD in evaluation practice, particularly in the field of international comparative projects. For example, beyond broad definitions, there is no single interpretation that suits all occasions. Securing SD can hold different meanings dependent on individual settings, and evaluation must reflect priorities related variously to lack of employment, recession, development that threatens environmental quality, out-migration or social cohesion. Furthermore, terminology can be misleading, as actors involved in SD use similar phrases for different purposes. In such instances, region-specific definitions and clarification of the separate roles of participants may be a prerequisite for tangible progress in evaluation.

With regard to qualitative SD data, there remains the likelihood that insights by participants may be subjective or partial. In tackling this problem, evaluators need to derive practical methods for reconciling ambiguities and different perspectives on the same issues, events, processes and outcomes. In comparison, quantitative data are more robust and commonly understood, but they are often not comparable. Gathered by different sources, with different methods, at different times, compatibility can be limited between dimensions within individual cases, and it can be very low when attempting to compare cases on an inter-regional or cross-national basis.

From a more pro-active stance, SD evaluation offers substantial scope for targeted impact. Designed to indicate where policy initiatives or instruments should be adjusted or reformed, it also allows more fundamental appraisals of value systems and administrative frameworks. This could facilitate wide-ranging and in-depth reappraisals of political goals, initiate bureaucratic and governmental reforms, and encourage greater public involvement in policy formulation and decision-making. Precisely because of the evolving, specialist nature of SD and its varied contextual relevance, it presents opportunities to influence how events and institutions take shape, and evaluation can be a strong mechanism in harnessing and directing this effort.

From the review of SRD evaluation case studies, it is

evident that there is a need for top-down guidance to develop a common methodology that would allow easier comparison across regions and across countries. Such guidance should not be developed in isolation, but rather in consultation with regional actors in a participatory forum, drawing on accumulated experience and parallel initiatives. National SD strategies could act as a logical starting point, for example providing initial definitions and access to working groups for indicators, monitoring and evaluation. In terms of the scope for standardisation, such an initiative could facilitate SD appraisal across economic and regional development instruments, making comparative SRD evaluation in the Structural Funds and other regional initiatives more feasible and more effective.

References

- Beckermann, W (1995) 'How would you like your Sustainability Sir, Weak or Strong? A Reply to my Critics', *Environmental Values*, vol. 4 no 2.
- Bendell T, Boulter L and Goodstadt P (1998) *Benchmarking for Competitive Advantage* Second Edition, Pitman Publishing, London.
- BirdLife International, CEE Bankwatch Network, CEE-WEB, Friends of the Earth Europe, Milieucontact Oost-Europa and WWF (2004) *Delivering Sustainable Development: Environmental NGOs' Common Position on European Regional Policy after 2006*.
- Clement K (2000) *Economic Development and Environmental Gain: European Environmental Integration and Regional Competitiveness* Earthscan, Kogan Page, London.
- Clement (2004) Structural Funds Programmes as Instruments for Sustainable Regional Development: A Review of Nordic Effectiveness, *Innovation* vol. 17 no 1
- Clement K and Hansen M (2001) *Sustainable Regional Development in the Nordic Countries* Report R2001:8, Nordregio, Stockholm.
- Clement K, Hansen M and Bradley K (2003) *Sustainable Regional Development: Learning from Nordic Experience* Report R2003:1, Nordregio, Stockholm.
- Clement K, Bradley K and Hansen M (2004) *Assessing Environment and Sustainable Development Integration in the Nordic Structural Funds* Forthcoming Report, Nordregio, Stockholm.
- Commission of the European Communities (1996) *The*

Structural Funds in 1995: Seventh Annual Report DG Regio, CEC, Brussels.

Commission of the European Communities (1999) Regulation (EC) No 1783/1999 of the European Parliament and of the Council of 12 July 1999 on the European Regional Development Fund *Official Journal of the European Communities* L213, 13.8.1999

Commission of the European Communities (2002) *Thematic Evaluation on the Contribution of the Structural Funds to Sustainable Development: A Synthesis Report*, DG Regio, CEC, Brussels.

Commission of the European Communities (2006) Regulation (EC) No 1080/2006 of the European Parliament and the of the Council of 5 July 2006 on the European Regional Development Fund, and repealing Regulation (EC) No 1783/1999 *Official Journal of the European Union* L210, 31.7.2006

Crabbé, P (1997) *Sustainable Development: Concepts, Measures, Market and Policy Failures at the Open Economy, Industry and Firm Levels* Occasional Paper 16, Industry Canada, Ottawa

Daly, H (1992) *Steady-State Economics*, Second Edition, Earthscan, London

Hajer M (1995) *The Politics of Environmental Discourse: Ecological Modernization and the Policy Process*, Clarendon Press, Oxford

Khan A (1995) 'Sustainable Development: The Key Concepts, Issues and Implications', *Sustainable Development*, vol 3, no 1

Kneucker R (1998) 'Introduction', in Gabriel I and Narodaslawsky M (eds.) *Regions - cornerstones for Sustainable*

Development Conference proceedings, October 28-30, 1998, Graz, Austria.

Neumayer E (1998) Preserving Natural; Capital in a World of Uncertainty and Scarce Financial Resources, *International Journal of Sustainable Development and World Ecology*, vol. 5 pp. 27-42

Nordic Council of Ministers (2001) *Sustainable Development: New Bearings for the Nordic Countries* TemaNord 2001:507

Reed D (ed)(1996) *Structural Adjustment, the Environment, and Sustainable Development*, Earthscan, London

Schleicher-Tappeser R and Faerber (1998) *Towards Sustainable Development: Experiences and Recommendations of Seven European Regions* EURES, Institute for Regional Studies in Europe, Freiburg.

Skolimowski H (1995) In Defence of Sustainable Development, *Environmental Values*, vol 4 no 1

Spangenberg J H and Giljum S (eds) (2005) Governance for Sustainable Development (Special Issue) *International Journal of Sustainable Development* 8: 1-150

World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press

ANNEX-LIST OF EVALUATION PROJECTS USED

Evaluation Project 1: Sustainable Regional Development: Learning From Nordic Experience

Evaluation Project 2: Assessing Environment and Sustainable Development Integration in the Nordic Structural Funds

ANNEX – EVALUATION PROJECTS

Evaluation Project 1: Sustainable Regional Development: Learning From Nordic Experience

Name of the program/initiative to be evaluated	Seven recent initiatives selected from the Nordic countries approximating to the practice of sustainable regional development (SRD).
Budget of the programme/initiative to be evaluated	n.a.
Budget of the evaluation project	DKK 300,000
Timescale of the programme/initiative to be evaluated	Most of the initiatives operated on a two-year timescale.
Timescale of the evaluation	12 months
Methods used in the evaluation	Qualitative only – identifying key actors in the interactive process, distinctive characteristics, strengths and weakness, progress in relation to objectives, barriers encountered and overcome, and lessons learned. Comparison oriented to offer insights into good practice and scope for benchmarking.
Key conclusions/findings of the evaluation project	Elements of best practice were identified from case studies and clustered as qualitative strategic benchmarks. Drawing on project experiences and aspirations, performance measures were derived corresponding to the phases of project design (feasibility prerequisites), implementation (framework conditions and methodology) and evaluation (impact and outcomes).
Key lessons for the organisation/ persons involved in the evaluation project	Whereas diverse cases offer interesting variety, they create difficulties in attempts to impose common structures within survey analysis. As actors perceive qualitative issues from different professional or sectoral perspectives, clarity in interview discussion and appraisal is essential. In the absence of specific regional (SRD) guidance against which to review progress, derivation of strategic benchmarks may offer a feasible alternative means of obtaining constructive output.
Other remarks	Nordregio Report 1:2003 Sustainable Regional Development: Learning from Nordic Experience was published June 2003.

Evaluation Project 2: Assessing Environment and Sustainable Development Integration in the Nordic Structural Funds

Name of the program/initiative to be evaluated	Structural Funds programmes in the Nordic Countries - 26 programming documents encompassing Objectives 1 and 2 and Interreg IIIA and IIIB.
Budget of the programme/initiative to be evaluated	n.a.
Budget of the evaluation project	SEK 140,000
Timescale of the programme/initiative to be evaluated	Design phase average of 1 year, operational phase of 5 years.
Timescale of the evaluation	12 months
Methods used in the evaluation	Qualitative analysis to identify rationale for environment and sustainable development integration, utilisation of goals, objectives or targets, and tools such as estimates of impact, project selection criteria and indicators. Quantitative analysis based on scoring system from 0 to 3, rating effectiveness of SD integration and environmental gain, amongst other factors.
Key conclusions/findings of the evaluation project	SD well integrated in only three programmes (for Interreg 3B). SD criteria for project selection appeared in three programmes, but only one programme specifically cited SD indicators for monitoring (although nearly all had environmental indicators). Environment was introduced mostly at priority level, with a subsequently high level of continuity, but less than half the programmes identified environmental objectives and targets.
Key lessons for the organisation/persons involved in the evaluation project	As SF programmes are continually revised and improved, the difficulty in assessing relative progress or merit must be acknowledged. Partnerships attach different value to SD, depending on regional contexts, and this impacts on integration. There is a subjective element in allocating scores, and a narrow range (0-3) may minimise distortion or misrepresentation.
Other remarks	Nordregio Report 7:2004 <i>Environment and Sustainable Development Integration in the Nordic Structural Funds</i> was published in December 2004.

The European Commission's Approach to Evaluation in the 2007-2013 Programming Period

By Kaisa Lähteenmäki-Smith¹

Evaluation framework of the Structural Funds: stability and dynamism

The European Commission has worked extensively to develop and influence evaluation practice through the way in which the Structural Funds (SF) evaluation framework has been set up and managed, particularly in the context of the logical models it entails, as well as with the methodological tools and guidance available in relation to the guidelines and the methodological working papers published by the European Commission. This short overview has the very practical motivation of providing an introduction to the new aspects of the EU's current approach to evaluating regional development interventions for the 2007-2013 programming period.

It is clear on reviewing the numerous evaluation reports compiled in connection to the SF programmes that the evaluation community has taken to heart many of the policy issues, methodological concerns and innovations put forward by the Commission, while it could also be said that the evaluation community may also, in turn, have influenced the development of new methodological support for evaluating SF interventions. This in itself is evidence of an implicit learning process. The issue of 'learning' has attracted ever greater attention, both in terms of facilitating the development and dissemination of 'best practice' within evaluation, and by so doing, through improving the quality of monitoring and adding value to the main strategic programme objectives within the Structural

Funds.² Learning has thus emerged as both an *instrument* and an *outcome* in itself. Joint learning and process evaluation is thus increasingly emphasised in addition to the traditional more cost-benefit-based and quantitatively set goals relating to accountability, efficiency and impact assessment (e.g. Hummelbrunner 2005, 20). This has been supported by both the policy community (including the European Commission) and the research and evaluation community.

In one sense learning has always been an integral part of territorial policy and the Structural Funds, in particular as evaluation has been an important element of the Structural Funds support system ever since the first reform of the funds in 1988. When it was first introduced, the objective of accountability (showing that the financial resources were used correctly and efficiently) was the main Commission driver in its approach to evaluation, with a particular focus on *ex post* evaluation, undertaken after the intervention. Gradually however greater emphasis has come to be placed on the learning aspects of evaluation, with an accompanying development towards incorporating the additional *process approach to evaluation*. The focus on learning as a cumulative process of participants of different types in a given policy intervention, resulting in improvement and sustained action has been promoted and local institution-building leading to increases in the capacity of people to initiate action on their own has been defined as a central goal (European Commission 2004, 2). This does not exclude the control functions and the necessary monitoring involved in accountability and performance approaches to evaluation nor does it replace them

¹ This article is based on an original written contribution describing the successor to the MEANS collection, first referred to as GUIDE and subsequently EVALSED, drafted by Ms. Veronica Gaffey (in 2004), as well as an oral presentation and *Power Point* presentation made at a *Nordregio* evaluation seminar on the 19th of December 2006 by Ms. Anna Byrulo, both from the European Commission. The article's interpretation of the Commission's approach and the presentations made herein are however those of the current author alone.

² On the nature of the policy-learning that this has entailed, see Gløersen, Lähteenmäki-Smith and Dubois 2007.

with a process focus, rather it reflects the different interests that the various parties to Structural Funds management and partnership have. It is obvious that the 'governance' issues relating to the quality and nature of delivery mechanisms and ways of doing things are relevant only in so far as they deliver effective and efficient policy interventions. The quality of implementation and learning may be important preconditions for efficient policy delivery, but they are not in themselves sufficient. (On the varying interests of different parties to the evaluation, from commissioners to objects; see also Jørgen Gren's article elsewhere in this publication)).

The standardised evaluation cycle of the Structural Funds provides a stable yet adaptable policy framework which can be seen to provide a suitable basis for policy learning to take place. It has also been argued that it is this stability, together with the dynamism of learning that has provided a framework for evaluation practice that has in itself helped to institutionalise the 'learning reflex' as part of the routine delivery of regional economic development policy (Bachtler and Taylor 2003, vi). The standardised programme cycle model, with its subsequent stages of learning has also provided a source of inspiration and best practice across the evaluation community. The stages of evaluation have, in the 2000-2006 period, entailed three types of evaluations in the temporal scale, i.e. *ex ante*, mid-term and *ex post*, as well as numerous thematic evaluations. Other types of evaluations and methodological developments of relevance here, undertaken in 2000-2006 included the mid-term up-date (to verify the findings and check the following recommendations of the Mid-term evaluations undertaken), as well as a number of other methodological developments in the new Member States on a voluntary basis (e.g. drafting evaluation plans, ongoing evaluations focussing on the quality of implementation mechanisms in particular, as well as thematic evaluations).

For the 2007-2013 period, *ongoing evaluation* is introduced as a specific kind of evaluation, drawing on the experience of the mid-Term evaluation update in 2005. The motivation here was that the update allowed Member States to narrow the evaluation scope and to focus on areas which were not sufficiently examined in the MTE (especially in relation to the results achieved), rather than to deal again with all the components required by the 2003 evaluative exercise. In line with the Commission's commitment to *subsidiarity* and proportionality, minimum core content was specified in this respect, with managing authorities identifying the additional evaluation needs they wished to have addressed in the update and which could add value to better delivery of operational programmes. (CEC 2006c, 5)

In order to determine the way in which interventions work, and how effective and efficient they are etc., the question of measurement is central. As the overall aim of the Structural Funds is to support better socio-economic

development which will contribute to the greater convergence within the EU as well as better global competitiveness, it is the methods and indicators used in assessing the effectiveness and efficiency of the interventions in question that are mostly connected to this theme that become of particular interest. Quantitative aspects have thus dominated, though the assessment of notions of 'effectiveness' and of long-term impacts have also necessitated an increasing reliance on qualitative analysis.

The logic of intervention developed in the SF context illustrates the main aspects to be considered in an evaluative perspective.³

The main evaluation criteria within an *ex ante* process are summarised in the figure below (CEC 2006a, 6).

Evaluation in the SF context is very much a question of creating feedback loops and qualitative programming support for those involved in programme planning and implementation, as well as creating the means by which all the involved authorities receive the necessary information in order to ensure both the accountability and the quality of programming.

Ex ante evaluation can be seen as a type of in-built quality check mechanism and as such it is relevant for policy contexts beyond the Structural Funds. The purpose of *ex ante* evaluation is to optimise the allocation of resources and to improve the quality of programming. The process of *ex ante* evaluation should be interactive and iterative in nature, i.e. it should support the establishment of links between programme management and external experts on the content of programmes drawn up and the recommendations of the experts should be taken into account by the planners in subsequent drafts of different parts of their programmes.

In addition to the programming stages described in the cycle below and their sub-sequence evaluative dimensions, the Commission also has at its disposal the ability to commission a number of rather more *ad hoc*, thematic evaluations, exemplified for instance by the evaluations of the Lisbon strategy (Danish Technological Institute 2005), Sustainability (GHK et al. 2002) or employment (Centre for Strategy and Evaluation Services 2006), to name but a few. These have been planned in an annual evaluation plan drafted by the European Commission (e.g. CEC 2005). This, together with the methodological working papers, provides a firm basis upon which to develop evaluation practice.

The programming structure in 2007-2013 is slightly altered from that of the 2000-2006 programming period, which has important implications for the evaluation activities undertaken. In 2007-2013 the programming process entails the following steps:

Community Strategic Guidelines: Strategic Guidelines were published following the approval of the SF regula-

³ The specific indicators foreseen for each of these dimensions are not analysed more closely in this paper, though some aspects of indicator work are referred to in other chapters of this publication.

Figure 5. Logical framework (Source: CEC 2006a, 6)

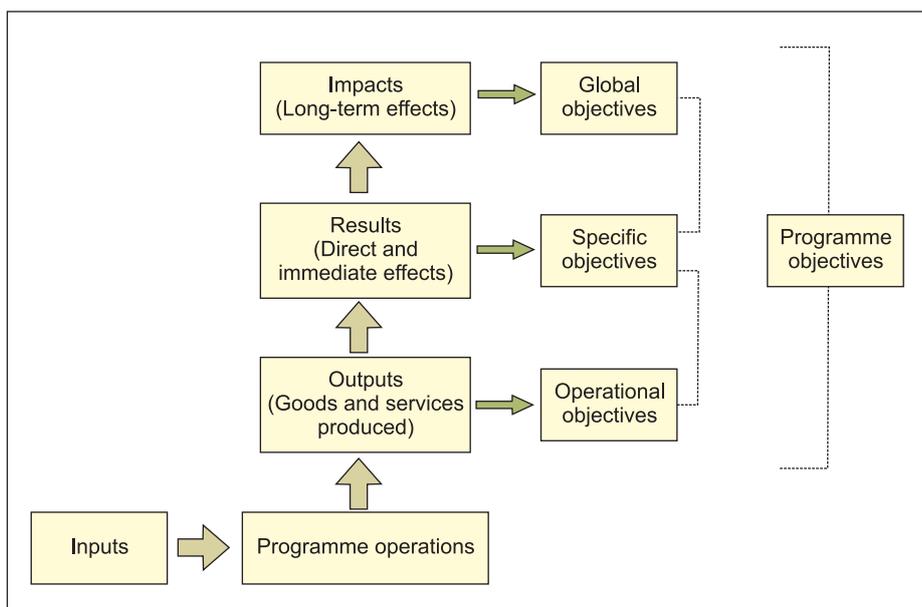
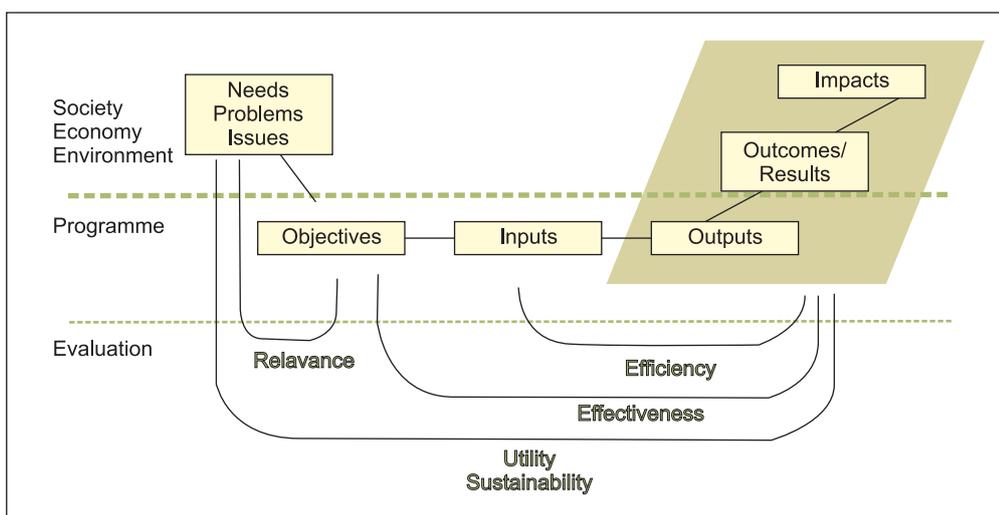


Figure 6. Evaluation criteria (Source: CEC 2006b, 4)



tions (in July 2006) in October 2006, with the objective of guiding the process of drawing up national strategies and Operational Programmes. In the new programming period a particular aim here relates to strengthening the linkages between the Structural and Cohesion Funds and the Lisbon Agenda.

National Strategic Reference Framework: Member States are required to draw up a National Strategic Reference Framework for the Funds, outlining the strategy to be adopted for the different objectives. It also provides a list of Operational Programmes and indicative financial allocations. This strategic document seeks to establish a link, at the national level, between the Community Strategic Guidelines and the national reform programme of the Lisbon Strategy. The process of developing the National Strategic Reference Frameworks varies (e.g. Bachtler and Polverari 2006), as can be expected based on the differences in

administrative cultures, the degree of centralisation or decentralisation of Structural Funds work, partner inclusion etc. Most were drafted in 2006 and are still in the process of being accepted in early 2007.⁴

Operational Programmes: Programmes have also become more strategic and focused more strongly than in the past on the 'priority axis level'.⁵ There is no requirement for a Programme Complement which was one of the main components in the planning for 2000-2006. (CEC 2006c, 4)

4 It has been estimated by the Commission that all 27 NSRFs should be adopted by the end of March 2007.

5 'Priority axis' refers to one of the priorities of the strategy in an operational programme comprising a group of operations which are related and have specific measurable goals, as defined in Article 2 of COUNCIL REGULATION (EC) No 1083/2006 of 11 July 2006 laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund and repealing Regulation (EC) No 1260/1999.

Strategic processes have thus also been changed to some extent, though even more interesting changes from our perspective are involved in the area of evaluation itself. When compared to the 2000-2006 programming period, the different methodological documents, guidelines and information sources used indicate some interesting changes for the 2007-2013 programming period. One of the central issues here is the increased focus placed on learning and on the dynamic nature of evaluation, as reflected for instance in the fact that the previously central *mid-term* evaluation has been complemented by a process of *ongoing* evaluation. As argued in the methodological working papers and SF regulations (e.g. General Regulation 1083/2006), one of the main lessons from previous programming periods has been that the nature of SF interventions necessitates a shift from the concept of a mid-term evaluation driven by regulatory imperatives towards a more flexible, demand-driven approach to evaluation throughout the programming period: and thus to, ongoing evaluation, as this best serves the general aim of evaluating cohesion policy, i.e. to improve the quality, effectiveness and consistency of the assistance from the Funds (CEC 2006c, 4). The shift from regulative and accountability-driven evaluation to ongoing and needs-driven evaluation is also accompanied by changes in the instruments supporting evaluation, as will be demonstrated below.

This shift can be seen to have emerged in the context of the Mid-Term Evaluation (MTE) experience gained during the 2000-2006 period, where improved evaluation quality and maturity across the Member States was clearly discernable. Nevertheless, the rigidity of deadlines and the generally broad nature of the requirements remained a limiting factor restricting flexibility: potentially too little focus was directed to results or to specific management needs, when the evaluation commissioners wanted to ensure that they covered all the bases in terms of what was required from the MTEs. The MTE update process undertaken in 2005 already attempted to address some of these limitations by placing a sharper focus on the areas where *added value* could be identified and by addressing additional questions drafted on a more *ad hoc* programme-basis. In 2005 additional thematic and strategic evaluations were also undertaken, and a similar approach was undertaken in the new Member States, all of which was deemed very useful and also supported the view that a more differentiated and flexible evaluation model should be implemented in contrast to the relative rigidity of 2000-2006.

The area of EU evaluation remains highly dynamic while at the same time we can now increasingly see the gradual attainment of a degree of maturity across the field. The emergence of renewed EU practice in the field of evaluation has been necessitated by the fact that evaluation practice needs to reflect the changing policy environment, as well as the needs of the evaluation community itself. This community has become mature in the sense that eval-

uation capacity in the EU has developed both within national and European administrations and among the evaluators themselves. This is evident in the better integration of evaluation in planning and implementation across the board, as well as in higher quality evaluations which use a variety of methods and are based more on interdisciplinary and pluralistic approaches. Such evaluations also increasingly reflect a shift from *accountability-driven* evaluation to a *learning-driven* evaluation, i.e. while perhaps in the early days of evaluating European policy interventions, evaluations were undertaken as a regulative necessity and as a response to the need to *legitimise policy intervention*, today they are increasingly (also) driven by real needs to *improve policy delivery and relevance*, as well as a means towards qualitatively better policies.⁶ In this respect they now more than ever have an aspect of learning imprinted into them: evaluation is expected to be a means of improving, not only the absorption and impacts of an intervention, but also the quality of policy more broadly.

The changes in this regard can be summarised as entailing a shift from a regulation-driven to a more needs-driven and flexible evaluation approach, which is less compartmentalised into the various sections of the cycle from ex ante to mid-term and ex post which has become a well-established and useful part of the evaluative practice within the SF context. The model to be implemented in 2007-2013 builds on the experience of the previous programming periods and seeks to provide stronger links between monitoring and evaluation, as well as the links between these two exercises and decision-making. In parallel with programming, an evaluation plan (Art. 48.1 of the General Regulation) is recommended in connection to all objectives, at the national and programme levels, also seeking to integrate the two aspects closer together. Some of the key aspects of the evaluation plan can be summarised as *improved co-ordination* between monitoring and evaluation, to be achieved through the periodicity/regularity of these exercises and through added focus on human resources and capacity building. The evaluation plan is also expected to include a revision mechanism, allowing for more lessons to be drawn and implications for practice implemented as an ongoing process. Some changes are also foreseen for the role of the Commission and its responsibilities (Arts. 47 and 49 in the General Regulation). An “early warning mechanism” is also foreseen, for situations where the monitoring system reveals a significant departure from the goal initially set. The intention of this system is to allow for timely adjustment to the programme though it has

⁶ As in evaluation theory and practice more generally, the balance between the accountability and legitimacy dimensions in relation to evaluation should not be seen as mutually exclusive – rather both motivations and rationales are present, though their centrality may vary from time to time and from one policy context to the next. In some cases the distinction may seem artificial, while in most cases the distinction between the more formal and quantitative against the more substantive and qualitative aspects are illustrative of different motivations; see for instance van der Meer and Edelenbos (2006, 202-203).

to be negotiated on a case-by-case basis with the Member States. Moreover, the Commission's future role is likely to be central in providing a broader comparative strategic perspective, e.g. through strategic thematic evaluations, which can support the Member States in their capacity building in

areas where a broader comparative approach is required.

All in all, it could be argued that this is a more forward-looking approach to evaluation best reflected both in the new forms of organisation and of the evaluation process, as well as in the forms of methodological support.

Methodological support: towards greater flexibility?

The increasing flexibility of the evaluation approach implemented in the Structural Funds context was referred to above, e.g. in connection to the evaluation plan approach, which can provide a useful framework for the ongoing evaluation while ensuring that it is effectively used as an integrated management tool during the programme implementation phase. While defining the co-ordination mechanisms for the ongoing evaluation, the evaluation plan leaves the Member State considerable room for manoeuvre in identifying areas where evaluative interventions may be necessary, as well as deciding on individual evaluations (thematic scope, time frame etc.).

The methodological guidance provided by the EU is another important resource in this respect. In addition to the evaluation cycle model and the subsequent stable and predictable evaluation framework set down in the Commission documents, it is the methodological guidance in its different forms that has given the EU such a central role in the evaluation field. In this regard perhaps the most influential single instrument across the evaluation community in the field of regional development has been MEANS Collection, developed as a state-of-the-art approach to methodological guidance by the European Commission in the 1990s. These guides were drafted for the purpose of setting the working practice recommended for the socio-economic development programmes and their evaluation, clarifying and standardising the concepts, indicators and technical solutions used, as well as introducing alternative methodologies and providing some thematic frameworks of analysis (e.g. within the fields relating to the competitiveness of SMEs, equal opportunities and the environment). The Commission has naturally also provided support for evaluation in the Structural Funds context in a number of rather more indirect ways, through the evaluation practice that has emerged around the EU-supported socio-economic interventions and in particular SF evaluation, thereby also having important indirect influence on programme and project evaluation more broadly, and also beyond the SF, in wider European or territorial policy contexts. This has been visible through the influence of methodological and regulative guidance, but also through the gradual emergence of a *European evaluation community*, which has emerged in close contact with the EU (e.g. EU-sponsored conferences, publications etc.) The two issues that will be further addressed in this chapter involve the methodological support provided by the EU, and the

changes in evaluation practice expected in the 2007-2013 programming period.

The need for more flexible and dynamic evaluation tools is also relevant here. In presenting the successor to the MEANS collection, Commission representatives have thus argued that the aim has been to create a dynamic resource that will be an 'evolving object' rather than something always about to become obsolete and requiring revision. The resources in the form of the Guide and the sourcebooks are made accessible by providing them free online and they are intended to be an interactive resource⁷, which can provide guidance on the evaluation of socio-economic development, by promoting exchanges of experience and good practice, as well as increasing the relevance of evaluation to decision-making. Methodological support is developed in order to ensure that evaluation is both methodologically and theoretically 'fit for purpose', while reinforcing and supporting the management and policy-making systems such that they are better suited to take full advantage of the evaluations.

The MEANS Programme

MEANS was a four-year programme financed by the European Commission involving external evaluation experts, which aimed to improve methods for the evaluation of structural policies. At the end of the programme in 1999 six volumes of methodological guidance were published and these have since formed an invaluable contribution to raising the quality and incidence of the evaluation of structural policies. The 'MEANS collection' has in fact had an influence beyond the Structural Funds community, as it has been used as a resource for evaluation in a broader context.

As with any type of policy practice within a multi-actor community based on broad partnerships and constant capacity building and learning, also within the SF evaluation community a process of learning was observed. At the fifth conference on the evaluation of the Structural Funds, held in Budapest in June 2003, the Commission representatives were thus able to discern how far the evaluation capacity had in fact developed within the EU⁸. The key aspects were summarised as follows:

⁷ The interactive part consists of examples and external references that the policy community is invited to submit to it. This part of EVALSED was first introduced in October 2006.

⁸ The papers delivered at the conference and a summary of the key messages are available at: http://europa.eu.int/comm/regional_policy/sources/docconf/budapeval/index_en.htm

- *A greater maturity of evaluation:* Evaluation capacity in the EU has developed both within administrations and among evaluators. This is evident in the better integration of evaluation into planning and implementation, higher quality evaluations which use a variety of methods and are based more on interdisciplinary and pluralistic approaches and an approach where evaluations are increasingly undertaken because they are needed rather than because they are an obligatory requirement.
- *Evaluation as an aid to decision making:* The most important contribution of evaluation is as an aid to decision making. Evaluators do not take decisions, but they can play an important role by asking the right questions, which may highlight contradictions in policies or policies which are inappropriate, too complex or too broad. The more evaluation is integrated into the system of governance, the more useful its contribution will be.
- *Complexity of Policies and Complexity of Evaluations:* Structural policies are becoming ever more complex, being decided upon through systems of multi-level governance and involving a wide range of eligible activities as well as transversal themes. This can lead to wide ranging and complex programmes which require complex evaluations in response. An important contribution of any evaluation is to highlight any inherent contradictions in such programmes and assist in the process of prioritisation.
- *Evaluation for Learning and Accountability:* Evaluation has a dual role, to contribute to learning and to demonstrate accountability. Learning should involve all partners and should take place at all stages in the evaluation life cycle. Accountability should also involve all partners and be an ongoing process, rather than being the exclusive competence of the Commission, discharged only at the *ex post* stage. The contribution of evaluation to learning is generally accepted, while progress is still needed in terms of accountability.
- *Learning - the Need for Explanation:* Evaluation needs to explain why things happen. In this way, lessons can be learned and improvements made to current and future programmes both within a region and across regions and Member States.
- *Building Evaluation Capacity Takes Time:* The objective of building evaluation capacity is to improve the quality of programmes supported by the Structural Funds rather than to produce good evaluations as an end in itself. This process takes time and requires learning in administrations and among evaluators. The approach should be based on taking practical steps on an incremental basis, rather than trying to put in place complex evaluation systems all at once.
- *Balance in Evaluation:* Good evaluation of the Structural Funds balances issues of efficiency, opportunity cost and market failure with ownership among the wider stakeholder partnership, taking account of the different perspectives on efficiency and cost issues. Good evaluation balances quantitative information with qualitative explanations. Impact evaluation needs to become more accessible to users and should exploit the complementarity in approaches, for example through the use of micro studies (cost benefit analysis and surveys of beneficiaries).
- *Monitoring and Evaluation are Interdependent:* Without good quality monitoring, it is difficult to have good quality evaluation. Good quality monitoring, which is the responsibility of the managing authority, minimises the need for evaluation to reconstruct non-existent information on programme performance and allows it to focus on its true function – providing evidence based judgements on performance and explanations for that performance. Progress in the production of more high quality evaluations depends on the improvement of data provided by the monitoring systems.
- *Proportionality is Required:* The skills required for good evaluation include economic analysis skills, people and institutional skills and management and financial control skills. These skills are required in evaluators and to some extent in the people using the evaluations. While democratic evaluations which are fully engaged and sensitive to the complexities are well worth the effort, the issue of the cost of such evaluations and the need to actually deliver the programmes must be borne in mind.

All of these aspects reflect the importance of evaluative information as a cognitive basis for policy development, though at the same time ‘relativising’ the status of evaluative information and evaluation processes. Evaluation should never become an objective in itself, but rather should remain an instrument to achieve more fundamental policy goals. As one source in building a cognitive base for better and more evidence-based policies it remains however a valuable asset.

From MEANS to EVALSED

In order to meet the challenge of this changing situation, the EVALSED interactive guide was created as the follow-up to MEANS.⁹ The new EVALSED guide is both more open and accessible and more interactive than the previous MEANS model, as it is freely available to all through the internet. The expressed intention of the Commission here is to update it regularly and explore the possibility of de-

⁹ The European Commission and its Directorate General for Regional policy have an important role to play in integrating the changes in evaluation needs with the methodological guidance provided. These are reflected in the form of the Means Collection and its successors, the “GUIDE” to Socio-Economic Development developed by the European Commission and the external experts used in developing these tools, e.g. the Tavistock Institute, GHK Consulting and IRS (Istituto per la Ricerca Sociale). The “GUIDE” was first published Online in 2003, with the revised version made available online in the form of EVALSED (2006-2007), focusing increasingly on evaluation capacity and the development of an interactive ‘community of practice’.

veloping an interactive element of the website which can promote the exchange of experiences and good practice among evaluators and those who commission evaluations. This multi-actor perspective and the fact that evaluative information is increasingly perceived as useful, not only for the evaluators, but also to policy makers who have an interest in evaluation, such as programme managers who wish to use evaluation in programme management and planning, public sector managers and civil servants who commission evaluations and programme partners involved as stakeholders. One of the themes through which these different types of 'user groups' can be addressed is the increased focus on examples and good practice.

The methodological support provided by the Commission (first through MEANS and subsequently in its EVALSED-instrument) can thus be seen as itself reflective of the gradual development of evaluation practice and increased maturity of the broader evaluation community. The main aspects of change are the interactive and flexible nature of the guide, as well as the focus on good practice, thus also reflecting the enhanced focus on evaluation as a gradual learning process. In this respect one could conclude that the changes in methodological guidance are consistent with the changes in intervention philosophy across the field, reflecting a gradual shift from a linear or straightforward *input-output model* to a more *multi-faceted* picture, reflecting the interactive and process nature of interventions, as well as a more systemic and evolutionary picture of the policy reality, as is apparent also from the other contributions to this volume.

References

Bachtler, J. and Taylor, S. (2003) *The Added Value of the Structural Funds: A Regional Perspective*. Glasgow: The European Policies Research Centre.

Bachtler, J. and Polverari, L. (2006) *The National Strategic Reference Frameworks: Towards a New Generation of Structural funds strategies*, Strategic Planning for Structural Funds 2007-2013 series, Glasgow: The European Policies Research Centre.

CEC (2004) Sourcebook 2 - Techniques & Tools in the "Evaluation of Socio-Economic Development: The Guide" available at http://www.evalsed.info/downloads/sb2_participatory_methods.doc.

CEC (2005) *DG REGIO Evaluation Plan, 2005*. Electronic document available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/info/evaluation_plan_2005.pdf. The various thematic evaluations commissioned by the DG REGIO at the European Commission are available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/rado_en.htm

CEC (2006a) *Indicative Guidelines on Evaluation Methods: Monitoring And Evaluation Indicators*, Working Document No. 2 in a series of guidelines for The New Programming Period 2007-2013.

CEC (2006b) *Indicative Guidelines on Evaluation Methods: Ex Ante Evaluation*, Working Document No.1 in a series of guidelines for The New Programming Period 2007-2013.

CEC (2006c) *Indicative Guidelines on Evaluation Methods: Evaluation during the Programming Period*, Working Document No.5 in a series of guidelines for The New Programming Period 2007-2013.

Centre for Strategy and Evaluation Services (2006) *Study on Employment Effects of the Structural Funds*, Electronic document available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/empleffecto6.pdf.

Danish Technological Institute (2005) *Thematic Evaluation of the Structural Funds' Contributions to the Lisbon Strategy. Synthesis Report*, February 2005. Electronic document available at: http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/lisbon2005.pdf.

GHK et al. (2002) *The Contribution of the Structural Funds to Sustainable Development. Final Report to the European Commission*, DG REGIO, December 2002. Electronic document available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/sustainable_vol1.pdf.

Gløersen, E., Lähteenmäki-Smith, K. and Dubois, A. (2007) « Polycentricity in transnational planning initiatives: ESDP applied or ESDP reinvented? », *Planning Practice and Research*, Forth-coming 2007.

Hummelbrunner, R. et al. (2005) *Process Monitoring of Impacts. Towards a new approach to monitor the implementation of Structural Fund Programmes*. Austrian Federal Chancellery, Division for Co-ordination of Spatial and Regional Policies

Van der Meer, F-B. and Edelenbos, J. (2006) "Evaluation in Multi-Acactor Policy Processes", *Evaluation*, Vol(12), 2: 201-218.

Appendix 1: The official regulations outlining the framework for the 2007-2013 Structural Funds programming period

(available at: <http://eur-lex.europa.eu/en/index.htm>)

- General Regulation 1083/2006, laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund and repealing Regulation (EC) No 1260/1999, 11/07/06, OJ 31/07/06.
- REGULATION (EC) No 1080/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 July 2006 on the European Regional Development Fund and repealing Regulation (EC) No 1783/1999/ERDF Reg., 05/07/06, OJ 31/07/06.
- COUNCIL REGULATION (EC) 1084/2006 establishing a Cohesion Fund and repealing Regulation (EC) No 1164/94, 11/07/06, OJ 31/07/06.
- REGULATION (EC) No 1081/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Social Fund and repealing Regulation (EC) No 1784/1999, 05/07/06, OJ 31/07/06.
- REGULATION (EC) No 1082/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a European grouping of territorial cooperation (EGTC), 05/07/06, OJ 31/07/06.

Appendix 2: Methodological working papers from the Commission

(available at http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm) including thematic guidance on issues such as...

- The Ex-Ante Evaluation of the Structural Funds interventions
- Indicators for Monitoring and Evaluation: A practical guide
- Guidelines on the calculation of public or equivalent structural spending for the purpose of *additionality*
- Guidance on the methodology for carrying out cost-benefit analysis
- Indicative guidelines on evaluation methods: evaluation during the programming period; see the link: http://ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_en.pdf
- Innovation in the national strategic reference frameworks (working document)

Kaisa Lähteenmäki-Smith (ed.)

Learning through Evaluation: the Nordic Experience

Nordregio Report 2007:3

Evaluation activity has become a normal part of policy development and implementation across all policy sectors, not least within regional development, where the European Union has supported the formalisation and methodological development of evaluative practice. As a research- and policy development activity the aims of evaluation are manifold. Functions are however primarily dependant on context. In addition to addressing the important function of accountability within the public sector, evaluations also fulfil an important systematization function, as they contribute to the rendering of raw information in a more comparable and commensurate manner. For organisations implementing and designing policy, functions of evaluative practice can range from that of a legitimising function to the role of a communicative or management instrument. In this respect evaluation has important system effects in the development of policy influence and diffusion and in seeking to manage the social complexity of various policy contexts.

Learning through Evaluation: the Nordic Experience highlights numerous examples of the process of learning through evaluation as it takes place in the Nordic countries. One of the key conclusions of the anthology is the need to identify different sources of methodological inspiration and comparative grounds for testing and assessing the usefulness and validity of various evaluation methods in order to better understand policies and their impacts. Even in areas of policy intervention where evaluation practice is firmly established and systematised (e.g. Structural Funds, R&D or the environment), there remains ample room for learning across geographical scales, as well as across policy sectors and disciplines.



NORDREGIO
Nordic Centre for Spatial Development

Nordregio
P.O. Box 1658
SE-111 86 Stockholm, Sweden
nordregio@nordregio.se
www.nordregio.se
www.norden.org



norden

Nordic Council of Ministers

ISSN 1403-2503
ISBN 978-91-89332-64-5