Evaluating the effects of regional interventions.

A look beyond current Structural Funds' practice

Warsaw, November 2011

Evaluating the effects of regional interventions.
A look beyond current Structural Funds' practice.

Warsaw 2011

Editors of the book:
Karol Olejniczak, Marek Kozak, Stanisław Bienias

Reviewer:
Professor Jarosław Górniak – Jagiellonian University, Centre for Evaluation and Analysis of Public Policies

Language editors:
Andrew Hinnant: Introduction, chapters 1.2., 1.3., 1.5., 2.2., 3.3., 3.4, 3.6., 3.7. Conclusions.
Ian Milton: Review, chapters 1.1., 1.4., 2.1., 2.3., 3.1., 3.2., 3.5.

ISBN: 978-83-7610-309-9

Publisher:
Ministry of Regional Development
Wspólna Street 2/4, 00-926 Warsaw
www.mrr.gov.pl
www.funduszeeuropejskie.gov.pl

Department of Structural Policy Coordination
Tel. +48 22 461 39 07
Fax +48 22 461 32 63
e-mail: sekretariatdks@mrr.gov.pl
e-mail: evaluation@mrr.gov.pl

Free Copy

NATIONAL COHESION STRATEGY
for Poland's development

MINISTRY OF REGIONAL DEVELOPMENT

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

Table of content

Introduction

Karol Olejniczak, Marek Kozak, Stanisław Bienias

The aim of this book is to enrich current evaluation of the Structural Funds in the field of regional development by connecting it with the advances in contemporary international practices of evaluation. We strongly believe that evaluation studies are one of the key feedback mechanisms in modern public policy. They help us reflect critically on the value and merit of our programmes, learn and improve the utility of new public interventions.

As a new member state in the European Union, Poland has eagerly started to apply evaluation in 2004. By 2010, we collected over 500 evaluation studies. However, as time was passing and the amount of information was growing, we started to look more critically and evaluate this evaluation boom. Soon, we became aware of three shortcomings of our activities.

Firstly, we realized that the majority of studies had been focused on the implementation process. They gave us some ideas on how to spend money faster and smoother, in a transparent way and in line with complex regulations. However, they told us very little on a few crucial issues – the effects of the executed programmes, the real socio-economic change we aimed at and the driving forces behind it.

Secondly, we realized that identifying and explaining the effects of our interventions was a very challenging task from a methodological point of view. We found ourselves with quite a limited toolbox of methods that could allow us to trace, in a sound and reliable manner, how Cohesion Policy contributed to socio-economic change.

Thirdly, we noticed there has been a relatively narrow group of decision-makers interested in using the knowledge provided by evaluation studies. It has primarily been the business of public managers that contract out studies, staff that run the programmes (very often the same people) and contractors – researchers and consultants. The general public, key decision-makers and journalists were not interested and usually not even aware of evaluation results as an information source.

Thus, by the end of 2008 Poland started to look around the European Union for guidance and inspiration. To our surprise, the reality looked different from what we expected. In the field of Structural Funds we found that our colleagues from other countries – both old and new member states – have been facing very similar challenges and shortcomings.

During the search for inspiration we also realised that our field – the evaluation of regional programmes co-financed by structural funds – has been growing in relative isolation both from the general craft of social research and from advanced international practices of evaluation. Out beyond the world of Structural Funds, there have

been a number of exciting ideas, well establish practices as well as new discussions and rising dilemmas.

Therefore, we decided to undertake a trip beyond the frontiers of our evaluation practice. In 2008, the Polish Ministry of Regional Development together with EUROREG – the University of Warsaw, opened an Academy of Evaluation. This was a series of lectures and workshops with national and international experts – academics and practitioners. The meetings were in the form of a one-year study programme for Polish Civil Servants, co-financed by the European Regional Development Fund (see: www.ewaluacja.edu.pl). This book brings together some of the issues, themes and ideas that appeared during those meetings. We hope that it will provide new ideas and encourage the reader to further explore the domain of evaluation of regional development policies. Our trip consists of three stages: looking at the changing context in which we operate, trying to establish where we are in terms of evaluation practice and finally searching for inspiration. The structure of the book follows this division. In the following paragraphs we provide readers with a more detailed look at the content of particular chapters.

We start our book by setting the stage and putting evaluation in the bigger picture of current developments in the field of public policy. We are convinced that we can effectively use evaluation in decision-making processes only when we understand the changing mechanisms and paradigms of the public sector and its regional policy. **Chapter 1.1 {Ferlie}** discusses debates in contemporary public sector management and – in relation to Poland – tries to identify factors that may influence further development of a national public management model, which is in the early process of the evolutionary change. In **Chapter 1.2 {Gorzelak}** the reader will find an overview of key dilemmas of contemporary Cohesion Policy with particular attention given to changes to redirect it towards a new development paradigm. In **Chapter 1.3 {Kozak}** the author discusses the context of contemporary strategic planning of public policy interventions and factors that have influenced the process of building a modern strategic development system in Poland. The gap between declared strategic objectives and their implementation (bringing rather short-term results) is being analysed in the wider context of the cultural and institutional surrounding in which the strategy is being currently implemented. **Chapter 1.4 {Ferry}** explores the possibilities and conditions of using evaluation as a strategic learning toll in managing the Cohesion Policy, and concentrates on benefits, particularly for large, long-term development programmes. The first part of the book ends with **Chapter 1.5 {Bienias, Gapski}**, which analyses the evaluation of public policies as an instrument of establishing a knowledge-based administration in Poland. After an overview of evaluation system and structure of evaluation studies, the authors also formulate recommendations for changes in public administration: the need to concentrate on effects instead of procedures, to introduce a performance-based budget and to properly formulate public policy objectives.

In the second part of the book we try establishing where we are in terms of current evaluation practice regarding the European Regional Development Fund (ERDF is the largest and most pro-development oriented fund and the specificities relating to its evaluation). The opening of **Chapter 2.1 {Bachtler}** is devoted to an analysis of factors that influence the development of the evaluation culture in the European Union. The conclusion stresses the need to adjust evaluation methodology to changing drivers behind Cohesion Policy instruments. **Chapter 2.2 {Olejniczak}** analyses key methodological shortcomings of current evaluation practice in the European Regional Development Fund. Based on the study of 100 evaluations, the authors come to the conclusion that without strengthening the methodological foundations, our knowledge on effects and impacts will remain limited, as a clear reference to regional development theories is lacking in almost all the cases researched. **Chapter 2.3 {Gaffey}** examines the effects of regional policy and the experience gained by the European Commission at the pan-European level (based on recent *ex-post* evaluations). The conclusions point to the need for more realistic expectations as far as thematic coverage is concerned and a need to abandon the imperative to spend money as fast as possible, which is no longer sufficient.

After examining challenges and limitations in the previous chapters, in Part 3 we explore sources of inspiration to make evaluation a better instrument for effective public policies. We compiled a number of the latest approaches and techniques from within and outside of the EU funds practice that aim to evaluate the effects of regional interventions. **Chapter 3.1 {Newcomer}** provides an overview of key issues in designing a strong methodological base for impact evaluations of cohesion policy interventions. It explains how to ensure the credibility of a study and its findings. **Chapter 3.2 {Pirog & Xu}** offers an overview of trends in published evaluation research by reviewing a body of articles published in three leading academic journals. The authors pay special attention to the use of experimental and quasi-experimental approaches in exploring causation. **Chapter 3.3 {Olejniczak}** concentrates on theory–driven evaluation, seen as not only able to provide reliable information on the real effects of the programme but also to explain the reasons why they appear. Despite limitations discussed in this chapter, theory-driven evaluation seems to offer a useful approach to researching the effects of the complex programmes of the Cohesion Policy. **Chapter 3.4 {Martini}** is devoted to an in-depth analysis of two approaches to impact evaluation: counterfactual and theory-based (driven), suggesting finally that objective-oriented Cohesion Policy intervention may benefit from a prospective approach (when operational programme design is accompanied simultaneously by impact evaluation design). **Chapter 3.5 {Widła, Wojtowicz and Wolanski}** tries to push the boundaries of net effect analysis and provides the reader with practical verification of a method proposed for impact evaluation of infrastructural projects in transport. The last two chapters of the book go beyond linear cause-effect analysis into the realm of complex systems behaviour. **Chapter 3.6 {Hummelbrunner}** presents potential benefits from introducing systems theory approach to evaluation of Cohesion Policy effects. It takes the reader through the

conditions and potential benefits of applying systems thinking. Finally, in **Chapter 3.7 {Płoszaj}** the author explores the concept of networks. By showing case studies from different evaluation research, it provides a promising perspective for using network analysis in assessing information flows, cooperation and knowledge accumulation.

We conclude the book by identifying three emerging challenges that, in our opinion, are central to the progress of evaluation practice in the field of regional development. We also try to signal possible solutions to these challenges that could be further explored.

We hope that readers will find this attempt valuable to bridge Structural Funds with the wider tradition of evaluating socio-economic interventions.

‘Some Debates in Contemporary Public Sector Management: An Overview’

Ewan Ferlie

Introduction

What are the choices and challenges facing contemporary public services in terms of their underlying organisational design? Can academic writing address this important public policy question and inform possible responses? This chapter takes a broad overview of academic writing about different organisational forms in the public services, hopefully placing later chapters on evaluation in a broad context.

What might be the underlying organising principles of the public services? This is of course a contested domain, where responses are subject to a heady mix of political, ideological and empirical influences. To provide a clear focus, the chapter will outline and critique three contrasting models or ‘narratives’ of public sector management: the Old Public Administration model, the New Public Management (NPM) model and the Network Governance (NG) model . It will consider some issues which arise in the Evidence Based Policy movement and the rise of the Evaluative State. It will finally consider possible effects of the recent financial crisis on the future scope and nature of public services.

The chapter is mindful of the differences between traditions in different European jurisdictions. This includes the Anglo Saxon tradition which is NPM friendly but also the Continental *Rechtstaat* model with its long standing attachment to rule by a separate body of administrative law. The post Communist trajectory in Eastern Europe may be distinctive. We need to avoid a simplistic convergence perspective which implies that all jurisdictions will eventually on (say) a dominant NPM model. Yet we also need to be aware of supra national bodies and institutions (such as think tanks, management consultancies or academic institutions) that act to diffuse public management ideas across national boundaries. The EU itself may be important as a supra national actor in public management reform.

Old Public Administration: The Weberian Bureaucracy

We start by describing the traditional ideal type: Old Public Administration. Historically, there was a strong divide between the founding principles of public administration and those of the early private firm based on the entrepreneur. As royal courts gave way to the expanding nation state in the nineteenth century (as in Napoleonic France, Prussia and then the United Kingdom), so a permanent government

bureaucracy developed as a substantive organisational form, more elaborate and rule bound than that of the owner managers of the first factories. The bureaucratic form was famously characterised by Weber (1946) (Meier and Hill, 2005, p52) by the following underlying principles:

- Fixed and official jurisdictional areas, ordered by rules, laws or regulations;
- The principle of hierarchy, whereby structures are established with superior and superordinate relationships;
- Management of the office relies on written files;
- Occupation of office is based on expertise and training;
- There is full time employment of personnel who are compensated and who can expect employment to be a career;
- The administration of the office follows general rules that are stable and that can be learnt.

The public administration model emphasised as its underlying principles: due process, the rule of law, probity (civil servants were paid a salary but should not otherwise benefit from their office) and neutral affect (decisions should be taken 'without fear or favour'). The new civil service was permanent, politically neutral, appointed on merit, and technically expert. It replaced earlier patronage based systems in which public offices could be bought by the wealthy. It operated to support elected governments of whatever political hue. The holding of public office was a vocation and a lifetime commitment, including a commitment to ethical behaviour. Public officials were rightly in this view insulated from market forces and were responsible to ministers (and indirectly citizens) rather than to customers or shareholders. The public administration form provided for relatively privileged, high status, and life long careers in the public services but in which there was socialisation into a set of distinct and hopefully protective public service values. It provided for 'rule based on rules' rather than by a sovereign, a charismatic dictator or indeed a vanguardist ruling party.

Old public administration originally developed in the emerging central Ministries of the growing number of nation states, where the permanent civil service provided policy advice to ministers. The policy core was supplemented by a large service delivery periphery. Within the expanding Welfare State agencies (health care, education) of the 1960s and 1970s, public services professionals were established actors in their own right. These settings were distinctive milieux in the operational periphery. They exhibited the 'professionalized public service bureaucracy' form which mixed facilitative 'lay' management, bureaucratised forms of administration (strong sets of rules and regulations) which limited direct line managerial action, but also strong professional dominance (Freidson, 1970) over core working practices. This position was defended ideologically by the professions through the doctrine

of clinical or academic judgement (medicine was the extreme case of a powerful public service profession).

Various criticisms emerged of the 'degeneration' of bureaucratic forms, especially as bureaucratic modes of governance spread into the large private corporations of the twentieth century that become more 'government like' than early family businesses. One criticism was the displacement of substantive rationality by procedural rationality (Crozier, 1964): it became more important to follow the many procedures than to achieve substantively rational goals. The maintenance of organisational routine was more evident than the management of change or innovation. Bureaucratic forms might be functional for high volume routine transactions; yet dysfunctional for niche or innovative production. A further criticism of public bureaux was that they were captured by interest groups of public sector providers and were unresponsive to clients or users.

By the late 1970s, a toxic combination of forces affected the public sectors in many jurisdictions. They had grown strongly through the 1960s and 1970s, financed through higher tax rates. Yet there was increasing political and public scepticism about their ability to provide good quality services for the public at reasonable cost. Public service professionals maintained their tacit empires, but traditional levels of social deference which restrained criticism were in decline. Public sector trade unions were taking more industrial action. There was even talk of 'ungovernability' and the decline of the ability of government to govern.

However, the original advantages of the old public administration form are not trivial (du Day, 2005). It avoided the danger of highly personalised or 'charismatic' forms of rule. It at least tried to ensure probity, avoiding the conflicts of interest apparent in privatised or hybrid forms. It proved highly resilient in practice and examples of post bureaucratic organisational forms may be more apparent rhetorically than in reality (Reed, 2005). The elaborate system of documentation provided an audit trail and some kind of retrospective accountability for decision making. Its stability provided a source of organisational memory (Pollitt, 2009) and therefore (at least potential) organisational reflection and learning.

The New Public Management Reform Narrative

Criticisms of public administrative forms proved influential in the policy domain as traditional public sectors went into crisis in the late1970s/early 1980s. The NPM reform narrative emerged as an academic and political doctrine which advanced reforms which could 'downsize' public bureaux and render them more controllable. It was linked to academic work in the public choice and organisational economics traditions (e.g. Tulloch, 1965; Downs, 1967; Niskanen, 1994). It argued that there was an over supply of public goods and an over expansion of government which needed to be reversed. The theoretical argument was that public agencies would

exhibit a tendency towards perpetual growth unless they were checked, as the rational bureaucrat would try to maximise his agency's budget and jurisdiction. Far from being selfless and ethical, civil servants rationally acted as budget maximisers. The principals (ministers) found it increasingly difficult to control the behaviour of their agents (public bureaucrats), given poor information and difficulty in measuring agency output (what exactly is a 'good' policy?). The prescription was to reduce the role of government and increase that of markets or market-like mechanisms, introduce more transparency, and use strong contracts and incentive structures to align the behaviour of principals and agents.

The NPM was often championed by Ministries of Finance as it helped create a set of policy tools that were designed to control the big spending departments. It was a far reaching and enduring reform wave of international scale and impact (Hood, 1991, Ferlie *et al*, 1996). High NPM impact jurisdictions include the UK, New Zealand and (perhaps surprisingly) Sweden (Hood, 1995); although France and Germany remained resistant. NPM ideas were diffused internationally by agents such as the OECD, management consultants and Business Schools. NPM's dominant values were related to 'hard values' such as productivity, performance management and value for money, rather than 'softer' values of democracy, social dialogue or organisational learning.

What was the core content of NPM ideas? It can be summarised in terms of the three 'M's: (i) markets or quasi markets (ii) measurement and audit and (iii) management: the doctrine was 'management should manage'.

At the macro level, the NPM sought both to reduce the size of the public sector and to make it less distinct from the private sector firm. Privatisation was a key policy development in the 1980s and 1990s as functions moved out of the public sector. NPM doctrine helped answer the second question: what should be done with those functions that remained within the public sector? A core goal was to produce a new framework of governance and incentive systems that would lead to pressure for systematic performance improvement. New corporate governance systems were developed at Board level, based on the private sector model, with more appointed Non Executive Directors from outside and the marginalisation of trade union, elected and staff representatives. Financial control systems (such as audit) were strengthened. Incentive systems were sharpened through performance related funding (rather than historic budgeting) with financial pressure thus being brought to bear on poor performers.

Organisationally, the old vertical bureaucracies of the public administration era were delayered and disaggregated. Traditional line management arrangements were weakened in favour of contracts and regulation, introducing novel 'loose tight' modes of control (Hoggett, 1996). In central government, ministries were separated out into a small strategic core which related through contracts to large operational agencies. In service delivery agencies, commissioners were separated out from

purchasers. This in principle enabled the development of quasi markets and greater pluralism on the supply side with easier entry of non public sector providers. Five year plans gave way to quasi markets where market-like forces allocated resources. Within the operational management units, however, there was a strengthening of the line management hierarchy and a marginalisation of trade union and professional representation. Indeed, there was suspicion of the tacit work practices of professionals and a desire to use stronger management to make them visible and controllable.

The NPM also provoked a shift from the use of *ex ante* rules and regulations to the elaboration of *ex post* audit, performance management and measurement (Power, 1997) systems. This was seen as a way of 'liberating' public sector management, while still providing a framework for retrospective accountability. In practice, audit systems often escalating until they posed a major regulatory burden in their own right, leading to a superficial 'box ticking' culture (McGivern and Ferlie, 2007).

While NPM impact certainly varied by country, it was a reform wave of considerable breadth and indeed persistence, achieving far more than superficial impact. Many of its recipes became deeply embedded in 'reformed' work practices and are now taken for granted.

Network Governance Reform Narrative

The turn to NPM produced new public policy problems and perverse effects in its own right. Firstly, the NPM deliberately stimulated competition and contractualisation, hence eroding traditional public service norms of cooperation. Its vertical reporting lines and strong performance management from above made the development of lateral working across professional and agency boundaries even more difficult. These deficiencies were particularly problematic in the design of responses to so called 'wicked problems' (Clarke and Stewart, 1997) (such as anti-crime or anti-drugs strategies) that went well beyond a single agency and required a system-wide response, as in the case of integrated primary health care (Thomas, 2006). Indeed, an isolated intervention from one agency could destabilise other agencies and reduce overall system capability.

Secondly, the NPM failed to engage many public services professionals with the governmental change agenda as professionals were threatened by assertive general management and crude targets imposed from above. Contractualisation and managerialisation were explicitly designed to challenge traditional professional dominance. In the longer term, however, an underestimated effect of NPM changes was the drawing in of groups of professionals into hybrid professional/managerial roles (such as Clinical Directors in health care) (Ferlie *et al*, 1996), eventually recapturing ground from general managers.

Thirdly, NPM reforms may have been effective in increasing productivity and 'doing more with less', but they were poor at dialogue with civil society (indeed they demobilised civil society by passing functions from elected bodies to appointed central agencies). This led to an increasing concern with the 'democratic deficit' associated with reconstituted NPM agencies, with a loss of legitimacy in public policy decisions. NPM reforms were also poor at creating the space for reflection and dialogue needed for organisational learning.

The NPM could be seen as a project inspired by organisational economics ideas which attempted to reduce public sector implementation deficits found in the 'overloaded state' by designing a set of novel governance arrangements, organisational and managerial reforms and strong incentive structures. The central State devolved operational management but tightened its strategic control. An alternative approach associated with the discipline of political science was to recognise and work with a high degree of policy complexity in the contemporary State. By the 1990s, political scientists were pointing to the 'hollowing out' of the traditional centre of the nation state (Rhodes, 1997) as functions moved upwards (to the EU); downwards (to strong regions) and sideways (to operational agencies). As it was progressively outsourced, service delivery became more complex with an increase in private and independent sector providers. Loose and diffuse policy networks emerged. Public policies had to be negotiated with many more social actors, raising a governance problem (Klijn, 2005). The State now steered through the use of looser policy instruments, such as contracts, alliance building, partnership working, persuasion and public exhortation, rather than the direct use of hierarchy. While the NPM had similarly emphasised the development of contracts, it had also built line management in the strengthened operational agencies responsible for mass service delivery. Such managerialisation was now being unpicked with a shift from hierarchy to network.

The emergent concept of 'multi level governance' was deliberately broader than the old concept of 'government.' It not only referred to network based forms of organising, but also allowed for more balance among the involved actors than assumed in state centric models. There was an increase in the scope for social dialogue and deliberative democracy, reflecting the co production of public policies among more diverse but also equal actors. The growing use of private providers, perhaps in mixed consortia with public providers, in large scale infrastructure modernisation projects (e.g. transport) was one clear trend compatible with the network governance model. There was a blurring of traditional sectoral boundaries and the emergence of hybrid forms (Public Private Partnerships), posing important issues of to whom they were accountable. The growth of policy networks in strong European regions which related directly to the EU as well as to the national capital was a second trend. The growth of citizen and user participation and of influence of social movement organisations on aspects of public policy (e.g. Greenpeace; Oxfam) was a third important effect.

Within the public policy domain, post NPM ideas were developed in the ideology of the Third Way (Giddens, 1997) which informed public management reform strategies in the late 1990s (e.g. Cm 4310, 1999). There was continuity in some areas with the NPM in that there was a stress on strong performance management. Many UK public policy networks were 'managed' from above and expected to deliver central targets, rather than showing strong self steering capacity. But most of the time there was at attempt to distance NG from NPM. Newman (2001) draws out the key characteristics of the NG narrative. Instead of the mix of markets, quasi markets and hierarchies found in the NPM, NG placed greater emphasis on networks. Instead of the vertical silos of the NPM, it promoted lateral working and 'joined up government', including system level interventions. Collaboration replaced competition as a guiding value. The new politics of inclusion brought a greater range of social actors into the policy process, including representatives of civil society and public services professionals more attracted to a policy emphasis on quality (rather than quantity) facilitated in the UK by a decade of buoyant funding for public services. As part of the move from hierarchies to networks, there was a switch from the NPM's emphasis on the formal role of 'management' to a broader development of 'leadership' using a set of interpersonal soft skills: network leaders could not use formal role power to command and instead had to win subtler means of influence.

Network based working was not a purely UK phenomenon but has been long established in European jurisdictions such as the Netherlands (Kickert *et al*, 1997). These networks often show greater capacity for self steering than in the more managed UK case. These 'soft' values of participation and inclusion appear influential in the EU model of the Open Method of Coordination (OMC) (Carmel, 2006).

Evidence Based Policy: Issues and Questions

The NG model included a search for more responsive policy tools which could contribute to more effective policy outcomes and greater policy learning. This included a commitment to Evidence Based Policy (Newman, 2001, pp69–72): the mantra was now 'what counts is what works'. A major theme of this edition relates to policy evaluations of EU cohesion policy undertaken in Poland. Other chapters address the technical and design aspects of evaluation, but this overview chapter should place these activities in a broader political and institutional perspective. Why should ministers and politicians be increasingly interested in policy evaluation activity rather than rely on (as might be the usual case) the opinions of citizens as expressed in elections?

The Network Governance narrative (Newman, 2001) suggests society is becoming more pluralist and complex, where there is less direct connection between government plans and actual effects. So there is a greater need for evidence on the actual effects of programmes which can inform organisational learning and if necessary programme redesign. These approaches were pioneered in America in the 1960s in

the evaluation of the social programmes of the 'Great Society' epoch. This experience is seen as formative in the development of public policy evaluation (Pawson and Tilley, 1997) and sometimes invoked as a role model. It needs to be recalled that there was an exceptionally favourable conjunction of forces in the America of the Great Society period. On the demand side, centre left politicians and policy makers were engaged in large scale social experimentation – making large scale funding available – and were open to feedback about programme design. On the supply side, there was an extensive, sophisticated social science community with an interest in applied as well as theoretical work in major social science orientated universities (e.g. Ann Arbor Michigan) and in corporations such as RAND. Policy evaluation activity diminished in the USA of the 1970s with a move to a more right wing administration sceptical of 'Big Government' and what it saw as the 'scientistic '(rather than scientific) pretensions of policy-orientated social science. It was noted that evaluation studies took too long to report for practical politicians and then often came up with equivocal conclusions. They operated in a highly technical domain far removed from the political forces that ministers have to take account of. So evaluation as a public policy activity can decline as well as grow, reflecting the balance of institutional and political forces.

The 1990s saw a revival of interest in public policy evaluation. Some governments (e.g. New Labour in the UK after 1997) were spending large sums on social programmes (as does the EU) with ambitious objectives but were relatively pragmatic about policy tools. Ideology was less important than performance so that methodological techniques which could reliably assess and even explain programme performance were helpful. The EU has also been an important actor in establishing mandatory regimes for evaluation across member States (Dahler-Larson, 2005).

In part, this revival of interest was initially sparked by the growth of the Evidence Based Medicine (EBM) movement in the 1990s within the vast international biomedical research complex. EBM's well known hierarchy of evidence model places particular weight on meta analyses of well conducted Randomised Control Trials, that is experimental methods. New ICTs made it easier to collect, store and disseminate the results of such studies electronically to an international network of clinical scientists, most of whom worked within the same research paradigm. The Cochrane Collaboration took a lead role in undertaking such synthesis work internationally. EBM spilled over into demands for Evidence Based Policy, evident not only in its heartland of health care but also in sectors such as social care and criminology close to the EBM paradigm.

Some important questions remain. One is methodological. Is the Evidence Based Medicine movement really a role model for Evidence Based policy or are there fundamental differences between the worlds of the natural and social sciences? For example, can experimental or quasi experimental methods be applied in public policy evaluations? If the answer is 'no' or 'only rarely', what other methods should be applied? The objectives of many evaluations may be rather mundane, such as

assessing whether a programme shows good value for money rather than elucidating models of intervention 'that work'. So there are many epistemological and methodological issues to be considered in designing public policy evaluations.

Secondly, what is the relationship between evaluation and the policy process? How does one tell 'truth to power'? Does policy really listen to evaluation (Hunter, 2009), or to put it more precisely: what are the circumstances in which it is more likely to listen? The policy/evaluation linkage may vary by jurisdiction and national tradition, Evaluation is a policy technology largely invented in the USA but now diffusing internationally. As Dahler-Larson (2005, p658) suggests, later adopters may have their own reasons for using evaluation and may develop indigenous styles of evaluation and evaluation use.

Thirdly, how should we balance the rival claims of evaluation and the collective deliberative process normally associated with the world of politics? The perverse effects of adopting a highly technical approach to public management reform may include the demobilisation of political parties and electors. Evaluation cannot provide a compelling political narrative and is largely inaccessible to lay audiences. It can assess empirical effects of programmes but finds it more difficult to balance competing values which is more the terrain of politicians and even ethicists.

Take the not implausible case of a government requesting an evaluation of its decision to privatise the delivery of some health or education services. The finding could well be that such programmes increase personal liberty and consumer choice, but also increase inequality and erode social solidarity. Evaluation can certainly produce empirical evidence which helps address these important policy questions: for example, are middle class consumers really more likely to exercise choice where health and education services have been privatised than working class consumers? If they are, can public policies be refined to correct such effects and increase equity? But evaluation by itself is often poorly equipped to handle trade offs about core values. One response to this limitation is to develop 'softer' forms of evaluation more compatible with critical reflection, learning, dialogue and a deliberate process, such as the fourth generation approach advocated by Guba and Lincoln (1989).

Other Theoretical Perspectives

This brief chapter has examined only three core models of public management for unavoidable reasons of space. There are other important theoretical prisms that could have been used: the democratic public administration model evident in countries which have made a transition from dictatorship (see Bresser-Pereira on Brazil, 2004), a post modernist perspective which draws on ideas from complexity theory to emphasise unpredictability, creativity, and non-linear and adaptive processes in public services organisations (Bogason, 2005) or a 'virtual organisations' perspective (Margetts, 2005) which sees the chance to transform traditional working practices in

public agencies through powerful new ICTs such as e-mail and the WWW. No doubt there are yet other perspectives available: the contemporary field of public management is rich and exciting in the application of social and organisational theories to the study of substantive policy domains.

The 2007–2009 Financial Crisis and the Public Services: Some Possible Implications

How might the recent financial crisis affect the scope and organisation of the public sector? In the short term, government has extended its role in various countries with the nationalisation of banks and tighter regulation of financial services. In the longer term, the implications for public service delivery are more ambiguous. The long period of economic growth based on a model of global and deregulated financial capitalism (mid 1980s–2007) had as one important effect a high tax yield from banks which in turn helped finance buoyant public services. This in turn facilitated a policy shift from a NPM style productivity agenda to the quality agenda more associated with NG ideas. The hegemony of finance capitalism was reflected in the extended influence of the banks in the public sector through the supply of private capital for large scale infrastructure improvements (as in the UK Private Finance Initiative).

This period of buoyant finance capitalism came to a dramatic end in 2007 and many European governments (albeit with significant variation) now face significant fiscal deficits likely to be reined in over the next five to ten years through a difficult combination of tax increases and public expenditure reductions. So the public services across Europe now face a long period of fiscal pressure, after a relatively buoyant decade. The long term implications may possibly include a move back from a quality agenda to a value for money and productivity agenda. The strengthening of budgetary decision making and financial control may re-emerge with as key area for institutional reform, with a shift of power from big spending ministries to Ministries of Finance and auditors. Will this be associated with a move back to a dominant set of NPM values? However, unlike the 1980s, governments are less likely to welcome radical privatisation or an extended financing role being undertaken by the banks.

Within the political domain, it is as yet unclear how forces will now move in terms of what become politically attractive ideas about the role and organisation of the public sector in the new financial landscape. One possibility might be the reassertion of socialist or traditionalist movements that would defend the traditional role of the State ‘against the banks’ and reduce the extent of public/private hybrid forms. However, the EU elections in 2009 showed a strong shift from the centre left to the centre right. So how will centre right forces seek to reshape public services in a period of financial stringency? Clearly there will be an attempt to ensure value for money (thus strengthening audit functions) and possibly withdrawing from non core areas. But core public services also act as an important safety net for all citi-

zens, but especially the less fortunate. Will dominant political forces see the public domain as an important source of social solidarity in difficult and risky times which help obviate strong social tensions and so reinforce its old Weberian values and separateness? Under such circumstances, might we even move back to Old Public Administration?

Discussion

This chapter presented and discussed three ideal types (Old Public Administration; New Public Management and Network Governance) of public services management. Behind them lie different assumptions about the behaviour of public agencies and even the motivations of public services managers. They are of course theoretical ideal types and not analyses of individual national jurisdictions: empirical cases could well suggest a mixed or hybrid pattern in practice. They do helpfully provide a way of thinking about underlying organising principles of public management and how and why they might vary. The Network Governance narrative in particular suggests the EU may be emerging as an important supra national actor in public management, in addition to the member states. The chapter has also hopefully placed the Evidence Based Policy movement in a wider political and institutional context and speculated about the longer term impact of the recent financial crisis for changing patterns of public sector management.

Four main questions emerge from this analysis for further consideration. First, does the presentation of ideal types help clarify what underlying public management reform trajectory Poland as a nation state is on? Clearly it has moved to a post Communist and democratic regime, integrated within the EU. At the same time, it has strong historical and cultural national traditions which may continue to influence political and institutional choices. How do these competing forces play out within the specific arena of public management reform? What are its basic and long term principles of public sector organisation and management reform? Is it adopting a public administration, a NPM or a NG model, or none of these? What does a post communist configuration in an Eastern Europe jurisdiction look like in comparative terms when benchmarked against other European jurisdictions (Pollitt and Bourkheart, 2004)?

Secondly, what are the implications of the discussion for the convergence (vs.) divergence debate within public management (Pollitt and Bourkheart, 2004), specifically in relation to the Polish case? Will the different jurisdictions in Europe – including Poland – converge on one single model of public management or will they retain their separate traditions? Europe contains jurisdictions with very different underlying models of the organisation and management of the public services. There is the Anglo-Saxon NPM rich model in the UK which has developed international traction. There are then the *Rechtsstaat* models of France and Germany, based on a complex body of administrative law. Then there are the post communist countries of Central

and Eastern Europe (including Poland) now building a very different form of public administration, influenced by their membership of the EU. Will the NPM reform wave eventually triumph across Europe? Or will countries continue to show a mix of selective appropriation of reform ideas combined with local tracks and path dependence (as Paradeise *et al* 2009 concluded in their comparative study of reforms to University governance systems across Europe)?

Thirdly, what is the role and impact of the EU as a supra national actor in public management reform processes? Pollitt and Bourckeart (2004) see the EU as self consciously different from the member states, and not particularly receptive to private sector management or NPM based ideas. The EU has promoted the use of certain management techniques to member states, however, often based on 'soft' management ideas such as TQM or a simplified version of the EFMD model of excellence, perhaps reflecting its own multicultural and collegial principles. Its consultation processes appear to emphasise an open mode of communication (Carmel, 2006). Specifically, we ask: will the EU work to produce convergence (Pollitt and Bouckeart, 2004) across Europe on 'soft' management models, or is its implementation capacity indirect and contested at the national level?

Fourthly, what is the wider institutional environment in which the 'Evaluative State' operates? The focus on building methodologically sound evaluations that can supply evidence to the policy domain is surely correct. Yet such evaluations operate within a wider set of political and public institutions and cannot purely be seen as technical exercises. What is being evaluated and why, and who constructs the evaluation agenda? Who listens and who learns from evaluations? Do they remain in a technical research domain or do they feed into and even change the policy process? EU models of management may be most compatible with the development of softer models of evaluation which allow for dialogue and social learning (Guba and Lincoln, 1989). This wider perspective may help us understand the dynamics of evaluation in the specific Polish case.

References

Bresser-Pereira, L.C (2004) *'Democracy and Public Management Reform – Building The Republican State'* Oxford: Oxford University Press

Clarke, M. and Stewart, J. (1997) *'Tackling Wicked Problems: A Public Policy Perspective'*, University of Birmingham: Social of Public Policy Discussion Paper

Carmel, E. (2006) *'Governance and the constitution of a European social'* in (ed) Newman, J. *'Remaking Governance',* University of Bristol: Policy Press, pp39-58

Crozier, M. (1964) *'The Bureaucratic Phenomenon'* Chicago: University of Chicago Press

Dahler-Larsen, P. (2005) *'Evaluation and Public Management'*, in Ferlie *et al*, op cit.

Downs, A. (1967) *'Inside Bureaucracy'*, Boston: Little, Brown and Co.

Du Gay, P. (ed) (2005) *'The Values of Bureaucracy',* Oxford: Oxford University Press

Ferlie, E., Ashburner, L., FitzGerald, L. and Pettigrew, A. (1996) *'The New Public Management in Action',* Oxford: Oxford University Press.

Ferlie, E., Lynn, L. and Pollitt, C. (eds) (2005) *'The Oxford Handbook of Public Management'*, Oxford: Oxford University Press

Freidson, E., (1970) *'Professional Dominance'*, New York: Atherton Press

Giddens, A., (1997) *'The Third Way'*, Cambridge: Polity Press

Guba, E, and Lincoln, Y. (1989) *'Fourth Generation Evaluation'*, Newbury Park: CA: Sage

Hood, C. (1991) *'A New Public Management for All Seasons?'*, Public Administration, 69 (Spring), pp3–19

Hood, C. (1995) *'The New Public Management in the 1980s: Variations on a Theme'*, Accounting, Organisations and Society, 20(2/3): 93–110

Hunter, D. (2009) *'Telling Truth to Power',* in (Eds) McKee *et al*, op cit, pp87–98

Kickert, W. *et al* (1997) *'Managing Complex Networks'*, London: Sage

Klijn, E.H (2005) *'Networks'* in Ferlie *et al*, op cit, pp

McGivern, G, and Ferlie, E. (2007) *'Playing The Tick Box Game'*, Human Relations

McKee, L., Ferlie, E. and Hyde, P. (eds) (2009) *'Organising and Reorganising: Power and Change in Health Care Organisations'*, Basingstoke: Palgrave Macmillan.

Meier, K. and Hill, G. (2005) *'Bureaucracy in the Twenty First Century'*, in Ferlie *et al*, op cit, pp51–71.

Newman, J. (2001) *'Modernising Governance'*, London: Sage

Newman, J. (ed) *'Remaking Governance',* University of Bristol: Policy Press

Niskanan, W. (1994) *'Bureaucracy and Public Economics'*, Cheltenham, Edward Elgar.

Paradeise, C., Reale, E., Bleiklie, I and Ferlie, E. (eds) (2009) *'University Governance – Western European Comparative Perspectives'*, Springer

Pawson, R, and Tilly, N. (1997) *'Realistic Evaluation'*, London: Sage

Pollitt, C. and Bourkeart, G. (2004) *'Public Management Reform'* (2nd ed), Oxford: Oxford University Press

Pollitt, C. (2009) *'Bureaucracies Remember; Post Bureaucratic Organizations Forget?'*, Public Administration, 87(2): 198–218,

Power, M. (1997) *'The Audit Society'*, Oxford: Oxford University Press

Reed, M. (2005) *'Beyond The Iron Cage?'*, in (ed) du Gay, P., op cit

Rhodes, R.AW. (1997) *'Understanding Governance'*, Buckingham: Open University Press

Thomas, P. (2006) *'Integrating Primary Health Care'*, Oxford: Radcliffe Publishing

Tulloch, G. (1965) *'The Politics of Bureaucracy',* Washington DC; Public Affairs Press

Weber, M. (1946) *'From Max Weber – Essays in Sociology'*, (eds) Gerth, H.H and Mills, C.W, New York: Oxford University Press

Cohesion Policy and the Challenges of the Future

Grzegorz Gorzelak

"Cohesion" has become one of the most important phrases of current policies conducted within the EU member states and the European Union as such. It is a relatively new term, first brought into the *acquis communautaire* of the European Communities in the Single European Act of 1986. This Act emphasizes the need to enhance the social and economic cohesion of the European Community with a view to levelling regional disparities and their potential growth, which was an anticipated result of the introduction of the single market. It was when the Maastricht Treaty established the European Union that the notion of cohesion did indeed take root. Since then – especially after the Cohesion Fund was created – it has become one of the leading directives of European Union policies.

"Cohesion" as a policy directive has three dimensions: economic, social, and territorial. The latter one is still a subject of discussion about its meaning.

Since the very beginning, "cohesion" has been simply understood as "convergence". These two terms have been used interchangeably. Cohesion is clearly viewed in terms of its equalising function. To reach a state of 'cohesion' means to eliminate territorial disparities in the level of economic development (economic cohesion) and in the access to labour and income (social cohesion). Such an approach to cohesion coincides with the regional policy of the European Union (formerly the European Community), which allocates about 60 per cent of its funding to support regions with a low level of development (defined as less than 75 per cent of the EU average GDP per capita, using purchasing power parity).

Member States, especially those that benefit the most from Structural Funds and Cohesion Funds, follow the Commission's principles of their own cohesion policies. In most, if not all of them, we also find "cohesion" and "convergence' used synonymously.

Yet, the goal of territorial convergence is hardly, if at all, achievable. This conviction has emerged even within the DG Regio (at that time DG 16), when they wrote in the Third Report on regional policy (*The Regions of the Enlarged Community*), published in 1987, that reaching a state of cohesion within the Community was a task the fulfilment of which was rather distant in time. As time passed, however, these reservations were reduced and a general belief that achieving "cohesion" through implementing "convergence" can constitute the basis for the regional policy of the EU has spread widely among the commission, the governments, regions and localities within the European Union.

Empirical worldwide evidence demonstrates, however, a strong persistence of historical regional patterns, even if these patterns were to be changed by massive external assistance rendered to less well-off regions. Central Appalachia is still the internal periphery of the U.S. in spite of the fact that it enjoyed the greatest share of the Appalachian Program. Mezzogiorno has not entered a path of fast growth, nor has it demonstrated characteristics of a 'Third Italy', emerging all of a sudden without any help from the Italian government and the EU. The most recent example of the former GDR dramatically shows that massive inflow of financial and technical help from the outside leads nowhere, and in many cases may be counterproductive by killing individual motivation and attitudes of self-reliance and self-responsibility (Lenz, 2007).

There are several positive cases of sudden advancement of some less developed countries and regions. Ireland is the one most often quoted – though the recent crisis has challenged some of the economic strengths of this country that appeared to be more "virtual" than real. But it is often forgotten that external assistance from the EU did not appear until 1994, and that it was coupled with massive inflow of foreign (U.S.) investment in computer and pharmaceutical industries. Moreover, Ireland achieved national success at the expense of growing internal regional differentiation, since most of its growth was concentrated in the South, leaving the North behind.

Ireland is a clear example of the polarisation process that is a product of slow growth of lagging peripheral regions and rapid advancement of metropolitan cores. *"Metropolises govern the world"*, as Manuel Castells says. This is because they have become the nodes of the "space of flows", to use another term by this author, which replaced the "space of places". The role the "place" plays in international (usually global) networks and interrelations is more important for its development than the features of the "place" itself, as regarded in isolation.

These processes are clearly pronounced in the countries that undergo accelerated transformation, for example Central and East European new Member States (Gorzelak, Smętkowski, 2010). It should be remembered that their fast growth was achieved before they became EU members, and that membership accelerated this growth not because EU funds began to flow to these countries, but due to the new possibilities of international cooperation, capital and technology transfers and the opening of the European labour market. However, a dramatic collapse of some of these economies (the Baltic States, but also Hungary) during the recent financial crisis (see Gorzelak, Goh, 2009) has proved that external positive influences cannot override domestic socio-economic processes, which, whether positive or negative, are the main drivers for development.

Moreover, in all Central and East European countries there is a positive relationship between the initial level of regional development and its dynamics. Capital city regions "escape" the rest of the country, and the regions that contain big cities fol-

low them. The border regions – those located at the external borders of the EU as well as those located along a country's internal borders – do display slow growth. Yet, these are usually less developed regions in all CEE countries. Thus a regional divergence is a fact that may be attributed to the relatively fast growth that is mostly concentrated in capital and big city regions.

The picture for the regions of the EU is more complicated. As indicated in the 4th Cohesion Report (CEC 2007), there is a negative correlation between the level of GDP per inhabitant and its dynamic of growth. However, the national and regional processes were not separated, and the overall pattern is a cross-product of these two dimensions. Convergence is observed on the national level, and divergence on the regional level. It is doubtful that these two levels will parallel each other. One may even risk a thesis that regional divergence is a condition of national convergence, since poorer countries grow faster the stronger their metropolitan cores are and the better they are linked with other metropolises within the global metropolitan network.

Thus, it appears that we cannot overcome the "equity-efficiency" dilemma, fundamental for the traditionally oriented regional policy.

The body of literature indicating that divergence within the EU seems to be stronger than the convergences is growing. Let me quote two pieces:

> *"Much of the language of European cohesion policy eschews the idea of tradeoffs between efficiency and equity, suggesting it is possible to maximise overall growth whilst also achieving continuous convergence in outcomes and productivity across Europe's regions. Yet, given the rise in inter-regional disparities, it is unclear that cohesion policy has altered significantly the pathway of development from what would have occurred in the absence of intervention"* (Farole et. Al,, 2009, p. 3).

And:

> *"The relative distance in terms of regions "lagging behind" between the new member states (e.g. Hungary, Czech Republic, Poland) and the old member states which have been classified accordingly (e.g. Greece, Spain, Portugal) does not seem to be very significant. A possible conclusion could be that this might be an alarming sign for more than a decade of structural funds spending in these regions which did not apparently achieve their goal in respect of strengthening territorial cohesion. Nevertheless, for some examples this effect seems to be visible – e.g. for Ireland and some regions in Spain* (ESPON 2006, p. 5).

So what's wrong with a cohesion policy that aims at achieving convergence? Several answers were given to this question, and the most important we can find in classical papers by Boldrin and Canova (2001), the Sapir report (2003), Rodriguez-Poze and Fratesi (2004) and the lesser known, but stronger in criticism of the cohesion policy

by Ederveen, de Groot and Nahuis (2006). Their arguments may be summarized as follows (compare Bachtler, Gorzelak, 2007):

- cohesion policy is mostly of social meaning, and in fact does not contribute to growth (Boldrin, Canova);
- traditional cohesion policy that concentrates on "hard" infrastructure and assistance to firms does not increase the competitiveness of lagging regions, and, in the long run, may even decrease their competitive advantages (Rodriguez-Pose and Fratesi);
- this policy may even decrease growth in countries with poor institutions, high corruption and that are not open enough (Ederveen, de Groot and Nahuis);
- only assisting education (Rodriguez-Pose and Fratesi) and institution building (Ederveen, de Groot and Nahuis) may create grounds for long-lasting development of such regions;
- to achieve these goals, EU policies have to be remodelled to increase the outlays for competitiveness and innovation, at the expense of the Common Agricultural Policy and traditional regional policy (the Sapir report).

For much of the past 20 years, the use of Cohesion policy resources has been governed by the assumptions of 'traditional' regional policies of the post-war period originating in Keynesian doctrine and state interventionism in a resource-based economy. Traditional regional policy was both formulated and pursued in what Castells (1997) dubbed the 'economy of places' – an economic reality where specialised economic and urban systems functioned in a way in which they were much more isolated from one another than is now the case. As a result of the shift to an open, knowledge-based economy, and from quantitative to qualitative development factors, traditional approaches have become much less effective. As previously indicated, M. Castells calls the current model the 'economy of flows', as it is a mutually interdependent system with a dominant role for the flow of goods, people, capital and, especially, information. In the current era, countries and regions will only gain a lasting competitive advantage if they can produce innovation on a steady basis. Exerting influence on this new economic model must take different forms than was the case under the previous paradigm.

The paradigm shift has been partly recognized by the Commission in their development of new assumptions for Cohesion policy for 2007–13. Drawing on the Lisbon strategy (Council, 2006): *These strategic guidelines should give priority to ... investment in innovation, the knowledge economy, the new information and communication technologies, employment, human capital, entrepreneurship, support for SMEs or access to risk capital financing'*. It is instructive to note that the Commission's original version of the Guidelines (CEC, 2005) was considerably more definitive in the need for a shift in policy focus but was 'diluted' at Member State insistence.

Maintaining such a direction in reforming Cohesion policy also calls for a reconsideration of the concept of cohesion. Arguably, *cohesion* should be understood in functional terms, and not as an effort to reach *convergence*. Convergence is an approximation of static states, whereas cohesion is dynamic by nature, being the opposite of entropy. Moreover, convergence is difficult to achieve, certainly with the limited resources available at the EU level. Cohesion should be liberated from its 'equalisation' underpinnings and should be understood rather as harmony and collaboration (economy of flows), lack of destructive pressures and irresolvable conflicts, and the possibility for co-existence and cooperation between individual components. Following this line of argument, an alternative understanding of the individual aspects of cohesion would involve a policy focus on three elements: *economic cohesion*, denoting the possibility for effective cooperation between economic agents, lowering transaction costs, and harmonizing relationships between businesses and their institutional environment; *social cohesion*, eliminating barriers to horizontal and vertical mobility through helping to overcome differences in levels of education, career advancement and material status; and *territorial cohesion*, removing constraints on spatial development which restrict the achievement of social and economic cohesion, such as eliminating barriers to transport, connecting the major nodes of European and national space, and developing research and business networks.

In the Green Paper (CEC 2008) on the new, third pillar of territorial Cohesion, some references to such a functional approach can be seen in the chapter entitled: *Connecting territories: overcoming distance*. Here we find statements suggesting that the functional alignment of individual territorial systems is one of the goals of territorial cohesion, e.g.:

> *"Connecting territories today means more than ensuring good intermodal transport connections. It also requires adequate access to services such as health care, education and sustainable energy, broadband internet access, reliable connections to energy networks and strong links between business and research centres. This is also essential to address the special needs of disadvantaged groups….*
>
> *In remote areas especially, the potential of IT to provide access to health care and education through telemedicine and remote learning remains to be developed".*

Unfortunately, these postulates are restricted only to what basically can be regarded as local systems, and leaves out the European dimension.

In addition to the above, the document propounds highly traditional approaches, evidenced whenever such phrases as *balanced development or diseconomies of very large agglomerations* are mentioned (so far, however, disadvantages of agglomerations are much less than the advantages of developing metropolitan functions, which can only be achieved in large cities). Likewise, it is pointed out that *the*

pattern of economic activity, however, is far more uneven than settlements (which is only natural as differences in the productivity of individual sectors are considerable). These assertions are based on a rather old-fashioned (and fallacious) assumption that *bringing economic gains from less congestion and reduced pressure on costs, [which should produce] benefits for both the environment and the quality of life* should still be the order of the day. Such relocation of activity from better to less developed places is only possible for the low and medium segments of the economy, whereas the top segment, which generates innovation and new technologies and has the highest added value, is only concentrated in certain places. So, as these concurrent processes take place, Friedman (2005) is right in his assertion that 'the world is flat' (where M. Porter's "comparative advantage" stemming from low costs of production is the leading driving force of growth), and so is Florida (2008) who conversely claims that 'the world is spiky' (where territorial units enjoy a "competitive advantage" achieved through creativity and innovativeness).

The EU document on territorial cohesion even uses the phrase: *more balanced and sustainable development, implicit in the notion of territorial cohesion*. As we can see, this is a tribute paid to old habits and beliefs, necessary as it may seem in order to move on towards a new and more appropriate approach to cohesion. Nonetheless, these passages smack of the traditional, now historical, approach to cohesion, usually identified with convergence, which – as we see in statistics – seems impossible to achieve, whether we like it or not.

Discussions undertaken during last years have brought new documents, of which the Barca Report (Barca, 2009) certainly is the most significant. The Report is a continuation of the traditional approach to the Cohesion policy, as it analyzes the distribution side (i.e., where and how EU funds should be allocated) and not on the delivery side (which was emphasised in the 2004 Sapir Report). The Barca Report also puts emphasis on "place-based policies" (that is, interventions in specific locations) and makes only a fleeting reflection about linkages, flows, cooperation and territorial interactions in the development processes. Additionally, it does not present a broad picture of the EU-wide dimension of the Cohesion policy, including initiatives spanning the entire EU territory.

The above considerations suggest that should the Barca Report be adopted as the leading guideline for reforming the Cohesion policy, the ideas presented herein are rather unlikely to be put into practice. Apparently, the time is perhaps not yet ripe for more original approaches, which is also corroborated by the repudiation of the Sapir Report in all its practical aspects and by the lack of tangible success in implementing the Lisbon Strategy.

As can be seen, the process of reforming the Cohesion Policy along the requirements of the present and – more importantly – of the future – have begun, but are still weak and need strengthening. The Commission will not introduce these changes by itself, since it depends on the voices of the Member States. Are they

– and mainly the least developed ones – ready to abandon traditional approaches, where “money” is often regarded as the main benefit of membership, and not other “softer” factors? This is a question that cannot yet be answered, since the time of real decisions, which will also be influenced by the current crisis, has yet to come.

References

Bachtler J and Gorzelak G (2007) Reforming EU cohesion policy: A reappraisal of the performance of the structural funds, *Policy Studies.*

Barca F. (2009): *An Agenda for A Reformed Cohesion Policy. A place-based approach to meeting European Union challenges and expectations*, CEC, Brussels. http://ec.europa.eu/regional_policy/policy/future/pdf/report_barca_v0306.pdf

Boldrin M and Canova F (2001) Inequality and convergence in Europe's regions: reconsidering European regional policies, *Economic Policy* 16(32), 205–253.

Castells M. (1997) *The Information Age: Economy, Society and Culture – The Rise of the Network Society*, vol. 2, Oxford: Blackwell.

CEC (2005) *Communication from the Commission Cohesion Policy in Support of Growth and Jobs: Community Strategic Guidelines, 2007–2013*, COM(2005) 0299, Brussels.

CEC (2007) *Growing Regions, Growing Europe, Fourth report on economic and Social Cohesion*, Directorate-General for Regional Policies, Commission of the European Communities, Brussels.

CEC (2008): *Green Paper on Territorial Cohesion Turning territorial diversity into strength*, Directorate-General for Regional Policies, Commission of the European Communities, Brussels.

Council (2006) Council Decision of 6 October 2006 on Community strategic guidelines on cohesion, *Official Journal of the European Communities*, L291/11, 21 October 2006.

Ederveen S, de Groot H and Nahuis S (2006) Fertile Soil for Structural Funds? A Panel Data Analysis of the Conditional Effectiveness of European Cohesion Policy, Kyklos, 59(1), 17–42.

ESPON (2006): *Integrated Analysis of Transnational and National Territories Based on ESPON Results*. http://www.espon.eu/mmp/online/website/content/projects/243/388/file_4022/fr-2.4.2-full-r.pdf

Farole T., A.Rodriguez-Pose, M.Storper (2009): *Cohesion Policy in the European Union: Growth, Geography, Institutions*, Working Paper. http://ec.europa.eu/regional_policy/policy/future/pdf/6_pose_final-formatted.pdf

Florida R. (2008): Who's Your City? How the Creative Economy is Making the Place Where You Live the Most Important Decision of Your Life, Basic Books.

Friedman T.L. (2005): The World is Flat. A Brief History of the Twenty-First Century, Farrar, Stratus and Giroux, New York.

Gorzelak, G., Ch.Goh, eds. (2009) *The Geography of the Financial Crisis in Central and Eastern Europe and Policy Responses*, EUROREG, Warsaw (forthcoming).

Gorzelak G., M.Smętkowski, 2010: Regional Development Dynamics in Central and Eastern European Countries, in: G.Gorzelak J. Bachtler and M.Smętkowski, eds.: *Regions of Central and Eastern Europe: Development Processes and Policy Challenges*, Routledge (forthcoming).

Lentz S., ed.(200f): Restructuring Eastern Germany. German Annual of Spatial Research, Springer, Berlin.

Rodriguez-Pose A. and Fratesi U. (2004) Between Development and Social Policies; The Impact of European Structural Funds in Objective 1 Regions, *Regional Studies,* 38(1), 97–113.

Sapir A, with Aghion P, Bertola G, Hellwig M, Pisani-Ferry J, Rosati D, Viñals J and Wallace H (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. http://www.euractiv.com/ndbtext/innovation/sapirreport.pdf

Strategic planning of regional development

Marek W. Kozak

Evaluation of regional development always requires a fair amount of knowledge not only about the results and impacts of operational programmes but also about the wider assumptions made by those planning, the procedures of planning, and the social, political, economic and institutional circumstances under which the whole process of planning and implementing takes place. Therefore, the main aims of this chapter are to:

1. discuss external factors determining the strategic foundations of regional development management (including planning)
2. analyse the legal framework of regional development in the context of ongoing European integration processes
3. discuss the strong and weak points of the solutions adopted.

For centuries the term "strategy" referred exclusively to the process of planning and running wars (in line with its original meaning in Greek). Since the 1960s, strategy has become a magic password to the successful management of companies and – increasingly – the public sector. The first publications from the sixties concentrated on the linear process of achieving long-term goals while over the next decades this picture was significantly changed towards a much more complex vision (thanks in a large extent to sociological contribution) (Pettigrew *et al*. 2002). For obvious reasons relating to the differences in main goals and functions of the private and public sector, a direct transfer of commercial experience to the public sector has to be limited. This leads Ferlie (2002:279) to the notion of "quasi-strategies" being typical for the public sector. In spite of all these differences, strategic approach (of various quality, as usual) has recently dominated public development policies all over the world. The European Union Cohesion Policy is probably the world's most complex and advanced example. There are quite different ways of defining the strategy or strategic development. Here we will use a fairly general definition of strategy as an instrument of conscious and systematic actions aimed at achieving specific long-term development goals. It makes a strategy part of a system of planning documents where, most typically, it is associated with medium-term and short-term plans (horizontal, territorial or sectoral). In the case of EU Cohesion Policy system, specific medium-term plans were renamed into programmes in order to stress a) their concentration on achieving priorities as described in the strategy; b) their detailedness and executive character. The aim of this text is to explain the reasons for adopting strategic management tools by the public sector rather than discuss the term itself. It is important to keep in mind that the separation of these reasons is for analytical needs only, as they are closely interlinked.

Factors of economic development and a shift of paradigm

The post-war period was characterised by the wide influence of Keynes' doctrine based on the assumption that regional development can and should be promoted using top-down, exogenous factor-based activities. Already in the 1960s it turned out that the initial successes of this supply-side based approach had been phased out by a series of economic turmoils and social unrest in localities benefiting from previous external public investments.[1)] As the global economy started to move from industrial towards a post-industrial, knowledge intensive economy, the interventionist doctrine, which was visibly not delivering, had been slowly crowded out by the neoclassical approach, underlying the meaning of demand rather than supply – endogenous and qualitative factors of growth. Today, an explosion of different approaches to key factors of growth can be observed. As there is much rich literature on the topic, here are just a few examples. A growing number of social scientists stress the role of culture in development (Landes 2005; Harrison, Huntington 2003). For many, social capital is considered a leading factor. The term was introduced by Bourdieu in 1972, proliferated by Coleman (1988) and Putnam (2000). Human capital has so many proponents that it is difficult to name them. Camagni (1991) introduced the term "innovative millieu". Florida (2005) proposes that the key is the existence of a creative class which grows when the 3T's are present (technology, talent, tolerance). Castells (2008) underlines that the contemporary economy is not about location, but flows linking key development centres and forming global network of cooperation and competition. The concept linking development with the quality of the institutional system is strong (Keating *et al*. 2003; Amin, Thrift 2000). Globalization, competitiveness, and innovativeness are the three factors that are said to determine (in interaction) the development chances of different regions and places (Gorzelak 2007). What do all of these approaches have in common? They all concentrate on qualitative factors and their interactions. What about the role of technical infrastructure in development, particularly in countries with both underdeveloped and outdated infrastructure when compared to affluent societies? The answer to this question should be defined not in terms "either-or" but rather in time sequence: for instance, a high level of mobilization (social capital) leads to infrastructural improvements (water, transport, sewage systems), but not the other way around (Hryniewicz 1998). When it comes to contemporary types of infrastructure, there is only one that is considered a prerequisite of development: a broadband network.[2)]

---

1 Factories built and later subsidized under the postwar policy of industrialization of less developed regions (eastern Poland) were the first "victims" of market economy introduced in Poland in 1990.

2 Even here one may ask whether the only reason that people do not extensively use ICT technologies to improve their lives lies exclusively in the lack of access to broadband. Studies on social exclusion stress that in the case of e-exclusion the most difficult problem is about personal competence to use the net. And that has nothing to do with the technical aspect. "You can bring a horse to the watering hole but you can't make it drink" – says British proverb.

Features of the new economy:

Firstly, a global nature manifested in both, decisions regarding the company location as well as the extended competition covering the IT and services sectors along with technology mobility and labour market globalisation.

Secondly, considerable acceleration of the following processes is a characteristic attribute of the new economy: the life cycle of a product has been significantly shortened and the process of technological change has gathered speed. Likewise, professional qualifications need to be permanently upgraded, and sometimes actually changed completely.

Thirdly, the modern economy is based on knowledge, which means not only increased employment in the sectors involved directly in information processing, but also steadily growing requirements with regard to employee qualifications.

Fourthly, the increasing importance of network. An ongoing process of specialisation results in companies being even more inter-dependent (linked) with other companies, institutions or suppliers than ever, while at the same time, networking brings about an increased flexibility and growth capacity (Blakely, Bradshaw 2002:5)

Globalization

Globalization is a multifaceted process and state of creation and existence of global subjects of economic, political, and cultural activities. It has a number of important consequences in various fields, of which we shall mention just a few.

Among the key features of globalization is the strength and flexibility of economic links between global old and emerging growth poles. Another feature is a phenomenon known as “shrinking time and space”. Product life cycle is short (particularly in cutting edge industries). The results of decisions can be spotted almost immediately (in seconds in the case of stock exchange information) everywhere in the world. Global problems (including environmental) call for global solutions. And indeed, never before has the world seen so many activities looking for global regulation executed by supranational bodies. On the other hand, people look for protection against the negative impacts of globalization in “territorialisation” of development (Bauman 2000). The changes have never been so fast and often dramatic.

The industrial era was already characterised by growing development disparities (before it was estimated at 1:2; Landes 2005), but nowadays the time of globalisation gives examples of rapid changes in disparity levels. Concentration of new development factors in metropolises has led to the de-industrialization as well as the depopulation of more peripheral, less developed areas (Blakely, Bradshaw 2002). Most territory of the former GDR and large parts of Poland’s eastern provinces may serve as examples. On the other hand, however, there is a visible catch up process on the global level: a number of states (China, India, Malaysia, Brazil and others)

have significantly improved their position. China alone has quadrupled its GDP in less than 30 years. World Bank studies report the visible growth of the share of middle class all over the world (Markandya *et al*. 2005). Not all countries are that successful. It should also be mentioned that given the conditions of accelerated structural changes, the position of socio-economic groups tends to change fast, which has obvious implications on their notions and behaviour (Gardawski 2008). These changes are particularly intense at the local level, where the restructuring processes can be turbulent.

Among the new features are: an increasing segmentation of the labour market and new forms of poverty related to the existence of so-called marginal jobs, which hardly require any skills and are poorly remunerated (see Jarosz 2008).

Postmodernist/information society

The changes taking place in the economy have had a significant impact on societies. New technologies call for better human capital. An increasing demand for a highly qualified workforce leads to the development of new sectors of the economy where teaching institutions are no longer a luxury but an industry closely linked with the business sector and its changing needs. Such phenomena mean that new society is not only better educated, but also expects more influence on public matters through increased participation and decentralization. Growing affluence makes people more mobile and better informed – having the firsthand experience of various countries, regions and places. Some also suggest that new generations are more tolerant and open minded; this, however, is based to a large extent on anecdotal evidence or research limited to certain millieus. Nevertheless, societies tend to be more fragmented than ever, undergo significant demographic changes, which in turn impacts lifestyles, public policies, and the economy (increasing the number of socially active people in retirement; increasing demand for health services for the elderly; a shortage of labour force etc.). One of the important aspects of the postmodern society is also its ability to organize people around common goals, starting from local environmental protection or lobbying for specific regulations. All these features make representatives of society an important stakeholder, who (collectively) is fully aware of his/her civic rights and possesses all the necessary knowledge and instruments helping to influence or control public administration. Knowledge and ICT are the main resources in the hands of new citizens (Krzysztofek, Szczepański 2002). The growing differentiation in the ability to utilise ICT becomes one major factor leading to the marginalization of significant parts of society.

Development challenges

In the context of globalization, the growing demand for effective instruments of coping with uncertainty, increasing competition, technological change, and dimin-

ishing distances reducing the "own space" has turned the attention of the public sector towards strategic planning and management. It is a natural response to the changes described by Castells as the rise of a network society (2000). For regional and local governments, globalization means, among others, the weakening of the state as a key actor in the development field and the increasing role of supranational organizations (such as the European Union or international financial institutions) and multinationals, which de facto have no national flag and move their operations around the world. It leads to defensive responses often described as part of "glocalization". Glocalization (a term introduced by Robertson in 1995) refers to parallel processes of globalization (which is of an economic nature) and the predominantly social response against negative impacts of globalization on local/regional communities. Recent decades have witnessed fast progress in community development, endogenous growth, and other concepts that are supposed to make communities less vulnerable to external shocks. It is also of growing importance in the perspective of new growth theories linking development with metropolises as hubs of commodities, services, and information production and exchange. According to Pike, Rodriguez-Pose, and Tomaney (2006:7), in the contemporary world there are three types of regions that may be successful:

- "Large metropolitan regions: Large urban agglomerations in both the 'developed' and the 'developing' worlds are where many of the high value-added service activities are concentrated. (...)
- Intermediate industrial regions: (...) This type of area often combines labour cost advantages with respect to core areas, with human capital and accessibility advantages with respect to peripheral areas, making them attractive locations for new industrial investment (...)
- Tourist regions: Among the regions in the 'developing' world that have managed to find their market niche in a globalised economy are the tourist areas. Places like Cancun in Mexico or Bali in Indonesia have thrived thanks to their capacity to attract large number of tourists from all over the world."

This raises the question about what is going to happen with other areas? This is a challenge to their public authorities and business. A growing number of researchers suggests that their future, to a large extent, depends on their ability to define and develop products that would make their economy functionally connected to better developed areas, mainly metropolises. A recent debate initiated by the European Commission on the territorial aspect of cohesion policy (EC 2008), Fabrizio Barca's report (2009), the Fifth Report on Economic, Social and Territorial Cohesion (2010), and many other documents confirm the significance of the problem and difficulties with proposing such instruments of development that would satisfy the needs and expectations of those less lucky regions.

Changes in public administration

Public administration is not an Ivory Tower splendidly isolated from its environment. It exchanges resources with and adjusts to the outer world. The increasing complexity of economies and societies, the development of new technologies (ICT first of all) push toward the necessary adaptations. In practice the highly rational Weberian model of bureaucracy (public administration) turned out to be neither as rational as promised nor able to cope effectively with the fast changing social and economic environment (Mazur 2003), Particularly when totally new and complex fields were introduced into the public domain (environment management, R&D, etc). The main reasons that forced the public administration and authorities to take a more active role through the development of a number of sectoral and horizontal policies were related to the internationalization and globalization of development and, on the other hand, to the process of making societies more mobile, fragmented, and diversified (OECD 2000). De facto national administrations have had to cope with the phenomena and effects of decisions of exogenous nature. European integration process also gave strong stimuli to modernize public administration in the context of both globalization and decentralization. Institutional response mostly takes the form of New Public Management (NPM), introducing elements of managerial rather than typically bureaucratic decision making procedures and accepting the fact that authorities are nothing more than one of the stakeholders in development (OECD 2007; Osborne, Gaebler 2004). The concept of Multi-Level Governance (MLG) is interrelated in that it takes into account not only a horizontal, but in particular a vertical dimension as well (OECD 2007; OECD 2008; Kozak 2009). The coordination of motives, interests, plans, and actions of numerous and diversified socio-economic partners turns out to be a major challenge for public authorities. The complex process of adjustment in administration functioning to the needs of contemporary economy and society should not be considered finally defined. Cultural differences significantly influence the mechanisms of functioning. Formal adoption of structures that accommodate solutions recommended for MLG or NPM does not guarantee immediate change. The Polish case shows that under the pressure of a set of factors (institutional reforms, needs of strategic programmes' management, civic society development, ICT proliferation) the process of change is irrevocable.

Functions of strategies

The aforementioned factors characterising contemporary society, economy, and institutional framework explain to a large extent the increasing interest in strategic management of development. The world and our socio-economic environment become more and more complex and interconnected, and public authorities are under the growing scrutiny of society. Change and uncertainty are daily facts of life. Strategy seen from a broader perspective is nothing more than an attempt to cope with the growing uncertainty through long-term planning. And strategy is consid-

ered a valuable instrument to mobilize tangible and intangible resources around shared goals and visions.

More specific functions of a strategy cover a number of aspects:[3)]

a) strategy helps understand changes and trends taking place in the global environment and their actual or potential impact;

b) strategy gathers in one place all the relevant information about the region, its strengths and weaknesses;

c) strategy informs the citizens and investors about the plans, priorities, and instruments, thus providing valuable information required for rational long-term decisions (type, location or conditions of planned investment, etc);

d) strategy can be a valuable instrument of mobilizing people around long-term development goals;

e) strategy serves as a basis for operational programmes elaboration and a benchmark helping to ensure coherence between different programmes and activities;

f) strategy changes the reality; and finally

g) strategy enables access to external funding (be it a public or private source).

In practice, the value of strategy depends on the main motives. A number of local and regional strategies in the EU Member States skyrocketed immediately when possessing a strategy was introduced as a formal requirement to apply for grants within the framework of the Cohesion Policy. On one hand, it contributed to increasing the awareness of strategic planning, while on the other hand, it often reduced the issue to a pure formality.

Polish strategic system and regional policy

After the first years of economic and monetary stabilization at the beginning of the 1990s, the need for a better planned development was generally accepted. Despite fierce debates it was only the preparations for accession and the accession itself that helped make a decisive step towards the introduction of a systemic set of planning documents in Poland.

The process was also facilitated thanks to the completion of the last phase of national territorial reform, decentralisation of the state, and the implementation of administration areas (including regional development), which laid the institutional foundations for the planning system development. The adoption of Principles of Support to Regional Development Act of May 12, 2000 (Journal of Law of June 14, 2000) was one of the most important regulatory initiatives, as it introduced a system of regional contracts providing the basis for the financing of regional development

[3] based on analysis of publications by Klasik, Markowski, Kot and others.

with central government funds, which was an important novelty element introduced at that time. Despite the ongoing work to prepare for the upcoming European integration, the Act was not fully compatible with the already available European law on cohesion policy in the member states.

The first stage was also marked by the elaboration of a document called the Concept of the Spatial Development Policy (kppzk). Finally adopted in 2001 along with a set of 6 sectoral and regional strategies[4)], the Concept and 6 strategies were expected to provide a starting point for the development of a National Development Plan (NDP), which, according to the requirements set out in Framework Regulation 1260/1999, was to create the basis for the elaboration of Operational Programmes (and their Complements) within the framework of the EU Cohesion Policy. From 2001–2006, as in other years, the National Strategy of Regional Development focused on the distribution of the European funds. With the exception of the Concept of Spatial Development Policy (kppzk), the system was obviously subordinated almost completely to meet the needs of the preparatory process indispensable to ensure access to European funds.

Th excessively detailed acquis on the Cohesion Policy planning and programming in the years 2000–2006 resulted in an unnecessary red tape burden, thus becoming an object of heavy criticism in the member states. The problem of bureaucracy was particularly acute in the countries with a poor quality of institutional systems. In Poland, for example, the high legal status given to both OP and OP Complements, which, combined with excessively restrictive and detailed law provisions has consequently led to an extremely time consuming and complicated procedure if changes to simplify and amend the system were to be implemented, as such changes proved to be essential in practice. All this created additional difficulty in the process of implementing cohesion programmes and projects.

The programming period immediately preceding the accession exposed the weaknesses of the legal system regarding the implementation of development activities as they were to meet the requirements of Polish law, and, quite soon, that of the European Union as well. Besides, they had to ensure the effectiveness and efficiency of the entire process as well. In an attempt to put some order into the existing system, the National Development Plan Act of 20.04.2004 (adopted 11 days prior to accession) determined the system of plans and their respective objectives (Journal of Laws, No 149, item 1074). In particular, the following structure was envisaged:

1. The National Development Plan (covering the programming period indicated in the plan) is identify the goals of socio-economic development, including the regional development objectives to be supported (Art. 3.1);
2. long-term regional development strategy (for 25 years);

[4] Strategies: the number of 6 strategies or their subject matter were by no means accidental. They reflected exactly the number of structural funds (4) and the type of measures financed under the Cohesion Fund (2).

3. National strategy of regional development (Covering the same time as the programming period).

Interestingly, the National Development Plan (NDP), being a document equivalent to the Development Plan, i.e. the document required in the aforementioned Framework Regulation, had to, pursuant to Art. 5, take into account the goals and objectives included in the planning documents of lower rank. The elaboration of sectoral, regional, and other OPs (Art. 8) was envisaged with a view to implementing the Plan. Each OP was to be supplemented by a Programme Complement, spelling out the details of the implementation system (Art. 11).

The provisions set out in the Act marked a step forward to improve the system structure, but they preserved the acutely felt absence of more comprehensive systemic solutions that would comprise not only the cohesion policy measures under Polish and European funds, but also the development planning in the areas excluded from the coverage of European policies or the cases when Polish funds were engaged independently of the European funds resources.

The document called "National Development Plan for the years 2007–2013" marked the first attempt to identify and discuss all the public funds available to meet the development needs. Prepared and discussed by the government in September, 2005, the draft was immediately put away on a shelf by the new government elected the same year. However, this was actually not the only reason, though: with the document preparation still underway, it soon became apparent that the European Union intended to introduce a new planning system for the 2007–2013 planning period. The document called Development Plan and prepared at the member states level was to be replaced by National Strategic Reference Framework (which also meant the clearly failed Programme Complements was to be eliminated as well). Thus, it was decided late in autumn 2005, that the work on an entirely new document would be launched, i.e. on NSRF. At the same time, work on Operational Programmes started. It needs to be added here that another document was ready in September, 2005, i,e. Updated Concept of Spatial Development, which was a far reaching modification of Spatial Development Policy from 2001. This document was also put away on a shelf and has not been replaced by a new one (however, the project of a new document called "Concept of Spatial Development 2030" has been under consultation since January 2011).[5)]

Generally, one can agree that despite certain difficulties, the foundations of the system to meet the needs of development policy and its implementation were laid in a successful and timely manner. Unfortunately, right from the very beginning it was flawed by certain shortcomings due to the hasty drafting of the law and documents. Furthermore, the system focused on spending European funds and achieving Community policy goals, which, right from the beginning, brought about some

[5] The text of the 2005 document is available at: http://www.funduszestrukturalne.gov.pl/informator/npr2/dokumenty%20strategiczne/kpzk.pdf (in Polish) (downloaded on 27.10.2009)

criticism (Kozak 2006). The imperfections were due to insufficient drafting time, and, consequently, numerous institutional changes followed in the years 2000–2005. They resulted in the transfers of administration areas dealing in regional policies amongst subsequently established and liquidated ministries. The only public agency (PARR – Polish Regional Development Agency) reporting to the minister of regional development and specialising in the implementation of regional development programmes also fell the victim to the institutional changes process (see EPRC 2009). As a result, the first period following Polish accession to the EU was a time of great chaotic effort to reorganise the system in Poland. Work to correct and improve the system was also undertaken at a strategic level. In Autumn 2005, the idea of creating a ministry of regional development (MRD) had a comeback (a similar ministry used to operate in the years 2000–2001) because of the well-justified expectation to gain benefits as a result of keeping the development policy in one ministry (being the Chairman of Coordination Board, the minister of regional development can in theory also influence the entire development policy).

Soon afterwards intense work was undertaken to elaborate the Principles of Implementing Development Policy Act of 6.12.2006 (Journal of Laws, No 227, item 1658), which was soon amended (only several months after its enactment).

Article 15 defines the relations between strategies and operational programmes. Thus, OPs are to serve the purpose of implementing the national development strategy, sectoral, and supra-regional strategies as well as regional ones in the form determined in the Act. The time a given OP is in effect cannot exceed the effective date of a respective strategy (except for the supra-regional strategy). The Act also envisaged that an implementation plan should be devised as a framework programming document to be prepared if the development goals were pursued in two or more OPs. Otherwise, they should be thematically coordinated operational programmes (see Art. and Chapter 4). When defining the requirements for the OP structure and the manner it should be developed (Art. 17 and the subsequent ones), a reservation was made that such programmes have to meet the requirements of the European law as well in the case of projects be compliant with the requirements set out in Art. 87.1 European Community Treaty or those in line with the notion of de minimis aid.

The planning period of the European cohesion policy for the years 2007–2013 has once again created the basis for planning Poland's development in general (even though European funds account for only one-fifth of the structural investments in Poland in the years 2004-2007; MRD 2009: 168). Despite everything, these funds expanded significantly the investment freedom Polish authorities had at a national, regional, and local level and created opportunities for accelerated growth or just the improvement of life quality. To a great extent, the difficulties faced during the preparatory period resulted from the fact that the elaboration of planning documents took place at the same time as the implementation of the projects and programmes of the 2004–2006 planning period, when their effects were still unknown and the

findings as to the experience confirmed by the *ex-post* evaluation were still to be gathered. The example of Eastern Poland shows that the planning work has not been fully coordinated and, as a consequence, a higher rank document (Strategy of Eastern Poland Development) was created later than the Operational Programme of Eastern Poland Development, which formally should have followed from the Strategy instead.[6)] Generally speaking, some solutions simplifying the system were set in place according to European guidelines. However, the peculiarity of Polish situation is that despite the negative evaluation of OP Complement and its rejection in Cohesion Policy, these documents survived under a slightly changed name (Detailed Description of Priorities instead of Complement) as national documents accompanying Operational Programmes. My hypothesis to explain the preservation of such a bad solution in the entire system is the following: the societies at a really low level of social capital development and, therefore, lacking in mutual trust, develop a natural tendency to enhance the safety of social contacts by providing excessively detailed legal solutions in their attempt to foresee all the possible social situations. Poland belongs to those countries known especially for their poor social capital and extremely over-regulated legal framework.

Given this context it is worth underlining that some more effort to give a more orderly structure to the national development management system was undertaken in 2008 with a view to making the system more transparent and resilient to the changes in, for example, the European legislation. The initiative was launched by the Ministry of Regional Development. The work resulted in the drafting of the Act on the amendment of certain acts in connection with the implementation of structural funds and Cohesion Fund of 7.11.2008 (Journal of laws No 216, item 1370). In Art. 9, the Act provides that the development strategies are as follows:

1) long-term national development strategy – a document outlining major trends, challenges and scenarios of the country's socio-economic development as well as guiding its spatial management while taking into account sustainable development principles to cover the period of at least 15 years;
2) mid-term development strategy – a document specifying basic conditions, goals and directions of national development in the social, economic, regional and spatial terms within the 4-10 year period to be implemented through the development strategy and programmes while taking into accoun the European Union programming periods;
3) Other development strategies – documents specifying basic conditionalities, goals and directions of development in certain areas defined in the mid-term national development strategy that refer to the development of regions, spatial development, that of sectors or fields and implemented thanks to programmes.

[6] The European Commission approved the Programme on 1.10.2007, and the Strategy was adopted by the Council of Ministers on 30.12.2008.

The arrangement of strategic documents partly reflects the actual situation with regard to the planning work that had started before drafting the Act. Therefore, the Act is to some degree reactive rather than creative, as the document covers the time perspective up to the year 2030 was elaborated under the leadership of Michał Boni, Secretary of State in the structures of the Prime Minister's Office early in 2008, before the Act was adopted. In its current form, the document presents undoubtedly an innovative approach to strategic planning based on, *inter alia*, the formula of challenges Poland is facing (Polska 2030).[7)] In the middle of 2011 it was followed by the long-term strategic document Polska 2030. Trzecia fala nowoczesności (Poland 2030. Third wave of modernity) ready for Cabinet approval. According to national regulation this long-term strategy will be accompanied by a medium-term strategy (in its final stage of preparation) and supplemented by eight sectoral and one regional strategy[8)].

Krajowa strategia rozwoju regionalnego 2010-2020 (National Regional Development Strategy 2010-2020)[9)] is of crucial importance for both regional policy and the aforementioned 8 sectoral strategies, as it is the only of a more horizontal/regional nature (and the only ready) (MRD 2010b). The document was subject to public consultations process and finally adopted in mid-2010. It presents a new approach to development as it puts more attention than ever on efficiency rather than equity through attempts to apply a polarization and diffusion model based on the assumption that diffusion of metropolitan development benefits to the surrounding area is plausible. The main objectives of the strategy are: 1. Support to regional competitiveness; 2. Strengthening of territorial cohesion and counteracting marginalization; and 3. Creating conditions for efficient, effective, and partnership-based implementation of territorially oriented development activities.

Koncepcja Przestrzennego Zagospodarowania Kraju 2030 (Concept of Spatial Development 2030) is another document of great importance, especially in terms of the planning process completeness. The document was adopted by the Minister for Regional Development in January 2011 and presented for consultations.[10)] The document defines the strategic aims of the spatial development policy[11)] and pro-

---

7 A forgotten predecessor of the long-term strategy is a document called "Poland 2025. Long term strategy of viable and sustainable development," RCSS and NFOŚ, Warsaw – which the Council of Ministers adopted in 2000.

8 These are the strategies under preparation: 1. Transport development; 2.National security; 3. Energy and environmental safety; 4. Social capital development; 5. Economy innovativeness and efficiency, 6. Sustainable rural areas and agriculture development; 7. Human resource development; 8. Effective state. From the methodological point of view it is important to notice that the National Regional Development Strategy 2010-2020 was ready one year before the long-term strategy.

9 It is to be pointed out with satisfaction that the adjective describing this document has finally been translated into Polish as that concerning a country, not a nation. Earlier, in the case of NDP for example, the English word 'national' was automatically translated into Polish as 'narodowy' – an adjective concerning a nation, not a country, even though 'krajowy' is the only adjective in Polish describing the documents concerning the country, and not the nation . Unlike in English, "national" in Polish has two different meaning.

10 Available at: http://www.mrr.gov.pl/aktualnosci/rozwoj_regionalny/Strony/Konsultacje_KPZK_2030.aspx, [6.02.2011]

11 The main aims may be described in short as follows: 1. Increased competitiveness of main urban centres; 2. Increased internal cohesion of the country; 3. Improved internal accessibility at various levels through infrastructure development; 4. Creation of spatial structures able to improve and protect high quality environment and landscape;

poses a hierarchy of planning documents needed to reduce territorial tensions and ensure controllable and sustainable territorial development. It is clear that before the Polish Presidency institutions in charge of development, in particular MRD, are gaining momentum at developing new systems and a set of modern, high quality planning documents.[12)]

The description of changes in the strategic planning system has so far focused on institutional aspects, particularly with regional policy. Legal and institutional solutions determine the limit values of development policies but do not affect their contents. By the way of a summary one needs to underline both the strong determination to reach the goals, as well as the serious initial problems encountered due to the quality of institutional framework. However, while evaluating this field of regional policy one has to bear in mind its being nothing more than an integral part of the general institutional system in Poland, which is considered one of the worst in Europe according to many international studies. Therefore, one has to value the achievements even more (such as the full utilisation of European funds in the years 2004–2006, to say the least), but it must be realised that no one can expect a high quality institutional system of regional development to be created unless a comprehensive reform of the overall institutional system is carried out in Poland (EPRC 2009). Nevertheless, even today one can say the current regional policy implementation system is the most advanced public intervention system in the country.

The picture would not be complete if one omitted the changes in the development assumptions, factors determining the development, or, in a word, the new development paradigm applied in practice. Given this viewpoint, the Polish system is of a dual nature, probably not the only one of its kind in Central and Eastern Europe. As with the Community Strategic Guidelines, the key documents now in force with regard to the cohesion policy (National Development Strategy; National Cohesion Strategy) refer to the Lisbon Strategy (or its newer version – Europe 2020) assumptions, or in wider terms, to the paradigm of economic development based on the knowledge and information society. At a general level, the reference is reflected to some extent in the operational documents of the cohesion policy, as manifested, *inter alia*, by the fact that 64% of all cohesion policy funds were earmarked voluntarily for the Lisbon Strategy goals (New Member States are known to be exempt from this requirement in the years 2007–2013). The readiness of central level administration to satisfy this earmarking criterion did not meet with much enthusiasm at the regional level, where, under the operational programmes managed by regional authorities, on average only 40% of the total funds were intended to be used for attaining Lisbon Strategy goals. The analysis of projects implemented in the framework of the cohesion policy in Poland after the day of accession (May 1, 2004) led the Ministry of Regional Development to conclude that the funds had

5. Improved territorial situation in relation to environment disasters, energy safety and defense. 6. Regaining and strengthening of spatial order.

[12] It does not refer only to EU-related activities. In line with the regional policy development pillar, the new Rozwój miast w Polsce (Cities and Towns Development in Poland) report was published in 2010 (MRD 2010 c).

been too heavily dispersed amongst numerous small projects of hardly any importance for reaching the cohesion policy goals (as they met only the social needs "at a local level and that of local communities''), which calls for some action to counterweigh this situation by, *inter alia*, incorporating such projects into 'well thought out comprehensive activities' (MRD, lipiec 2008:18; MRD 2010a).[13] Referring back to dualism, it can be said that the closer one gets to operational activities such as the selection and choice of projects to be implemented, the weaker the readiness to accept sensible spending to achieve the goals of the Lisbon Strategy because the long-term development goals would often lose when competing with the current 'needs' of local communities. The more strategic and flexible approach to the planning and management of the cohesion policy program has turned out to be too difficult to adopt, which is manifested by the fact that Program Complement (of Operational Programme) has been maintained as a planning document under a slightly changed name even though it was repealed in the EU regulations. On the other hand, the principles of monitoring and evaluation[14] have been implemented into Polish practice,[15] which, in the long run, can make a more substantial contribution into the modernisation of the country than financial resources.

Strategies and their impact: preliminary comments

Having a strategy is not the only important element, but ensuring its ability to achieve goals is more important still. The assessment of effects, the evaluation of implemented programmes, and, thus, the evaluation of whether the strategic aims have been achieved is one of the basic keys to the rationalisation of development policies. To what extent can the question about the effects (impacts) of the implemented activities be answered? So far, *ex-post* evaluation of the National Development Programme 2004–2006 proves that in the period 2004-2009 the GDP only grew thanks to European Funds by 3% (MRD 2010a: 13). Econometric models suggest that European Cohesion policy resources contribute yearly by 0.4 to 0.7 percentage points to both employment and GDP growth (ibidem). As it turns out, somewhat surprisingly, the impact on economic structure was the most positive in terms of industrial GVA production rather than services (ibidem: 15). The visible concentration of spending on infrastructural, often small, local projects may contribute to explain another conclusion found in two out of three of the presented econometric models that positive impact on GDP will end together with the end

[13] The dispersion of projects financed from the cohesion fund results largely from the fact that municipalities have great financial freedom in operating their budgets, and thus, the use of European funds depends largely upon their interest in various Operational Programmes priorities. The infrastructure shortcomings and weak points direct their attention to investment in technical infrastructure. However, management poses a problem. With a tremendous pressure on spending, EU funds have obviously made the project of relatively low effectiveness acceptable (it follows indirectly from the MRD report; July,2008).

[14] One important outcome is a considerable growth of interest in the methods of public policy management, including the laying of foundations of the evidence based policy.

[15] Though almost exclusively under the Cohesion Policy for the time being.

of the influx of European money. For citizens of most affluent member states it can be of interest to know that money spent in Poland significantly contributed to an increase of Poland's imports from them. Most importantly, almost 60% of these imports relate to products of upper medium and high technology (ibidem: 19). Without going into detail, the evaluation leaves no doubt about the efficient implementation of the Cohesion Policy in Poland in the period after accession and also shows the benefit to net-payers. Simultaneously it stresses that to a large extent future benefits will be determined by the ability to activate demand side effects and ensure the durability of projects.

Regarding the financial aspect, the resources allocated under European structural funds for the years 2004–2006 were used up in Poland by more than 100% (exactly: 102%) (MRD 2009:3). The surplus resulted mainly from exchange rate fluctuations.

Any discussion on the effects and impacts has to take into account that in practice those stemming from the financial perspective 2004–2006 overlap with those from the resources of 2007–2013. Most Poles highly regard European Union funds though without necessarily knowing the specific source of funding.

The interpretation of results extending beyond the 2004–2006 period calls for some caution: the notable growth of employment after accession and the increase of income or level of investment activities is the result of many factors coming into play (such as economic growth trends worldwide and opening markets thanks to European integration processes); they cannot be attributed only, or mainly, to the cohesion policy. Apart from the analysis of changes currently underway in Poland, in a breakdown by regions, the publication 'Regional development in Poland. 2009 Report' indicates that as much as 70% of the total spending under structural funds in the years 2004-2007 was allocated for the development of basic infrastructure, support of manufacturing sector (16%), and human resources development (14% respectively) (MRD 2009 May: 168). Another report shows that transfers from the EU budget in 2009 have reached the level equivalent to 2.3% of Poland's GDP and 14.1% of total Polish investment outlays (MRD 2010 d: 31). However, not much convincing data is available to confirm that modernisation has been achieved thanks to structural changes in the economy or employment, or spending on research and development, expenditures on innovation, or the high share of advanced technologies in production or exports (Kozak 2010). This proves that after 5 years, the laying of the foundation of a new economy is less advanced than expected, and the good results that the Polish economy has achieved (including *inter alia*, the dynamic growth of export) follow primarily from comparative, not competitive (qualitative) advantages, but final opinions should be withheld until comprehensive next generation *ex-post* analyses are completed.

Conclusions

To sum up, the development of strategic thought in Poland over the last two decades should be associated mostly with preparations for accession and participation in the implementation of EU policies, especially the cohesion policy. The association was particularly notable during the first period after accession, when the entire planning system was actually subjected to European policies. During the next stage, which is still in progress, work has been undertaken to create a comprehensive and more flexible system capable of adjusting to the changes in the national and European institutional framework. Given a more holistic nature (i.e covering divergent and yet important aspects of social life), it is also to ensure the capacity to focus on truly strategic goals. So far, only the foundations of strategic planning system have been laid and the indispensable skills have become an important element in training, among other things, the Polish administration. Generally, the key strategic documents are acknowledged to be of high quality. Apparently the process of system maturing has not been completed, so it still remains to be assessed in terms of its usability and quality. The system is not only expected to ensure a clear division of responsibility and accountability for the efficient preparation of strategic documents that would meet the development needs of the country and be based on a new paradigm of development, but, at the same time, they should not be contradictory to the institutional framework and legal system of Poland and the European Union. Moreover, it is primarily to ensure the capacity to interface the strategic goals onto the operational level and ensure the cooperation of all stakeholders so that they can be achieved. Otherwise, strategic documents might just turn into papers to be put away on shelves, or elaborations created just to meet the needs of European and Polish law, and be void of any significance for development guidance and support. At the end of the day, the measurable impact strategic planning has on development is the ultimate test of its quality.

The process of pursuing and achieving the strategic goals has not advanced enough yet to be subject of a final assessment. However, it can be preliminarily stated that the planned structure of intervention has not been strictly maintained during the first planning periods. The share of funds used to finance basic infrastructure development was much higher than originally planned at the expense of investment into the production environment, and human resources in particular. The closer one gets to the reality at a grass roots level, the less readiness one finds there to finance projects referring directly to the elements of a new development paradigm[16]. Furthermore, this thesis has been additionally confirmed by the dispersed funding of numerous and unrelated local projects of small value and significance. Even considering these factors, the delays and difficulties encountered in the course of many infrastructure projects (motorways, railways etc.), implementation came as

[16] All in all, about 64% of all cohesion policy funds in Poland have been allocated for pursuing the Lisbon Strategy goals in the years 2007–2013; which, in the case of national programmes means about 90% of cohesion funds, whereas it is only 40% at the regional programmes level.

a surprise and exposed the poor quality of institutional framework and the need to opt for radical changes. The implementation of innovation-focused projects turned out to be a problem: both beneficiaries and implementing institutions lacked experience on how to define and evaluate innovation parts of projects. Many problems were caused by excessively complicated management systems with too much red tape. Given this, we should appreciate all the more the fact that none of the EU structural funds allocated for the 2004–2006 programming period were allowed to remain undisbursed.

The information presented here can be used to discuss a wider context, mainly of a cultural and institutional nature. However, strategic documents are neither created nor implemented in a vacuum. Their implementation depends directly and indirectly on many aspects of institutional order, which is not highly regarded, as many international studies indicate. A relatively poor quality of human capital is a notable weak point with a very low level of social capital in particular, which brings about the lack of trust in mutual relations amongst people and implementing institutions on one hand, and the tendency to over-regulate the system on the other (the new version of Complement – 'Detailed Description of Priorities' – is an excellent illustration of the problem at hand).[17)] As many stakeholders focus their attention on the easily noticeable civilisation gaps, the possibilities of implementing a new development paradigm are thus postponed even further. While not overestimating the importance of strategic documents, they can become important tools to move the Polish way of thinking about the priorities and factors of development into a different direction. Undoubtedly the effects of the Cohesion Policy, the major source of funding development in Poland, show an increasing efficiency of the system. With a newly constructed strategic planning system one may also expect that the long-term effect shall improve significantly.

[17] According to many opinions, this is the heritage of the farm estate and landed gentry culture dominant in Poland until the end of the inter-war years, which is found so very difficult to discard.

References

Barca F., 2009, An Agenda for a Reformed Cohesion Policy. A place-based approach to meeting European Union challenges and expectations. Independent Report.

Bauman Z., 2000, Globalizacja, PIW, Warszawa

Blakely E.J., Bradshaw T.K., 2002, Planning Local Economic Development. Theory and Practice, Sage Publications Inc.

Camagni R., 1995, The concept of innovative millieu and its relevance for public policies, in: Papers in Regional Science, 74/4

Coleman J., 1980, Foundation of Social Theory, Harvard University Press, Massachusetts, London

EC (European Commission), 2008, Green Paper on Territorial Cohesion. Turning territorial diversity into strength, {SEC(2008) 2550}

EPRC (European Policy Research Centre), 2009, *Ex-post* Evaluation of Cohesion Policy Programmes co-financed by the ERDF (Objective 1 and 2), Work Package 11: Management and Implementation Systems for Cohesion Policy, Draft Final Report, Metis, Vienna and University of Strathclyde, Glasgow

European Commission, 2010, Fifth Report on Economic, Social and Territorial Cohesion, http://ec.europa.eu/regional_policy/sources/docoffic/official/reports/cohesion5/index_en.cfm [5.02.2011]

Ferlie E., 2002, Quasi Strategy: Strategic Management in the Contemporary Public Sector, in: Handbook of Strategy and Management, SAGE

Florida R., 2002, The Rise of the Creative Class: And How It,s Transforming Work, Leisure and Everyday Life, Basic Books

Gardawski J., 2008, Ekonomiczny wymiar struktury społecznej, w: Gardawski J., Gilejko L., Siewierski J., Towalski R., Socjologia gospodarki, Difin, Warszawa

Gorzelak G., 2007, Rozwój – region – polityka, in: G. Gorzelak, A. Tucholska, Rozwój, region, przestrzeń, Ministerstwo Rozwoju Regionalnego, Warszawa

Gorzelak G., Kozak M.W., 2008, Poland, in: M. Baun, D. Marek (eds.), EU Cohesion Policy after Enlargement, Palgrave, MacMillan, New York

Harrison L.E., Huntington S.P. (red.), 2003, Kultura ma znaczenie, Wydawnictwo Zysk i S-ka, Poznań

Hryniewicz J., 1998, Wymiary rozwoju gospodarczego gmin, w: Gorzelak G., Jałowiecki B. (red.), Koniunktura gospodarcza i mobilizacja społeczna w gminach, EUROREG, Studia Regionalne i Lokalne No 25, Warszawa

Hryniewicz J., 1998, Wymiary rozwoju gospodarczego gmin, w: Gorzelak G., Jałowiecki B. (red.), Koniunktura gospodarcza i mobilizacja społeczna w gminach, EUROREG, Studia Regionalne i Lokalne No 25, Warszawa

Jarosz M. (ed.), 2008, Wykluczeni. Wymiar społeczny, materialny i etniczny, Instytut Studiów Politycznych PAN, Warszawa

Keating M., Loughlin J., Deschouwer K., (2003), Culture, Institutions and Economic Development. A Study of Eight European Regions, Cheltenham UK, Northampton MA, USA: Edward Elgar

Klasik A., Kuźnik F., (red.), 2001, Zarządzanie strategiczne rozwojem lokalnym i regionalnym, Wyd. Akademii Ekonomicznej w Katowicach, Katowice

Kot J., 2003, Zarządzanie rozwojem gmin a praktyka planowania strategicznego, Wyd. Uniwersytetu Łódzkiego, Łódź

Kozak M.W., 2006, Factors influencing absorptive capacity of new member states, w: Markowski T., Turała M. (red.), New members – new challenges for the European regional development policy, in: Studia Regionalia vol.18, Komitet Przestrzennego Zagospodarowania Kraju PAN, Warszawa

Kozak M.W., 2006, System zarządzania europejską polityką regionalną w Polsce w pierwszym okresie po akcesji, in: Studia Regionalne i Lokalne, Nr 2

Kozak M.W., 2009, Kompetencje i rola poszczególnych poziomów realizujących politykę regionalną (w tym kraj, region, instytucje i ciała wspierające) z uwzględnieniem aspektu zależności hierarchicznych i multi-level governance", luty 2009, ekspertyza dla Ministerstwa Rozwoju Regionalnego

Kozak M.W., European cohesion Policy and the development of Poland, 2010, in: P. Churski, W. Ratajczak (eds), Regional development and regional policy in Poland: first experiences and new challenges of the European Union membership, Part II, Polish Academy of Sciences, Committee for Spatial Economy and Regional Planning, Warsaw

KPRM, 2011 (lipiec), Polska 2030. Trzecia fala nowoczesności. Długookresowa strategia rozwoju kraju, Warszawa

Krzysztofek K., Szczepański M.S., 2002, Zrozumieć rozwój, Wyd. UŚ, Katowice.

Landes D.S., 2005, Bogactwo i nędza narodów, MUZA SA, Warszawa

Markandya A., Taylor T., Pedroso S., 2005, Tourism and Sustainable development: lessons from recent World bank experience, in: Lanza A. Markandya A., Pigliaru F. (red.), The Economics of Tourism and Sustainable Development, Edward Edgar Publishing Ltd., Cheltenham

Markowski T., 2008, Teoretyczne podstawy rozwoju regionalnego i lokalnego, in: Strzelecki Z. (red.), Gospodarka regionalna i lokalna, Wyd. Naukowe PWN, Warszawa

Mazur S., 2003, Historia administracji publicznej, in: J. Hausner (ed.), Administracja publiczna, Wyd. Nauk. PWN, Warszawa

MRD (Ministerstwo Rozwoju Regionalnego), 2008 lipiec, Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004-2007, Warszawa

MRD, 2009 maj, Rozwój regionalny w Polsce. Raport 2009, Warszawa

MRD, 2009 wrzesień, Wykorzystanie środków z funduszy strukturalnych ii funduszu spójności w ramach narodowego planu rozwoju 2004–2006 oraz narodowych strategicznych ram odniesienia 2007–2013. Informacja miesięczna za wrzesień 2009 r.

MRD, 2009a, Krajowa Strategia Rozwoju regionalnego 2010-202. Regiony, miasta, obszary wiejskie, Warszawa

MRD, 2010a, Ewaluacja *ex-post* Narodowego Planu Rozwoju 2004–2006, http://www.ewaluacja.gov.pl/Ewaluacja_ex_post_NPR/Documents/Ewaluacja_ex_post_NPR_07_12_10.pdf, [12.01.2011]

MRD, 2010b, Krajowa strategia rozwoju regionalnego 2010-2020, Warszawa, http://www.mrr.gov.pl/rozwoj_regionalny/polityka_regionalna/ksrr_2010_2020/informacje_podstawowe/strony/ksrr_2010_2020_inf_podstawowe.aspx [10.01.2011]

MRD, 2010c, Rozwój miast w Polsce, Warszawa

MRD 2010 d, Wpływ funduszy europejskich na gospodarke polskich regionów i konwergencję z krajami UE. Raport 2010,Warszawa

OECD, 2000, Government of the Future, OECD Paris

OECD, 2007, Linking Regions and Central Governments. Contracts for Regional Development, OECD Paris

OECD, 2008, Poland. Territorial reviews of OECD, OECD Paris

Osborne D., Gaebler T., 2005, Rządzić inaczej. Jak duch przedsiębiorczości przenika i przekształca administrację publiczną, Media Rodzina, Poznań

Pettigrew A., Thomas H., Whittington R., 2002, Strategic Management: The Strengths and Limitations of a Field, in: Handbook of Strategy and Management, SAGE

Pike A., Rodriguez-Pose A., Tomaney J., 2002, Local and regional development, Routledge, London and New York

Putnam R., 1995, Demokracja w działaniu, Wyd. Znak, Kraków

Rada Ministrów, 2001, Polska 2025. Długookresowa strategia trwałego i zrównoważonego rozwoju, Published by RCSS i NFOŚ, Warszawa.

Robertson R., 1995, Glocalization: time-space and homogeneity-heterogeneity, in: Featherstone M., Lash S., Robertson R. (red.) Global Modernities, SAGE, London

Evaluation as a learning tool in public policy: The case of cohesion policy

Martin Ferry

Introduction

One of the basic justifications for the widening use of evaluation, including under Cohesion policy, is that it can serve as a vital learning tool for improving public policies and programmes (FURUBO *et al*., 2002). The ability to assess and draw lessons concerning the impact and utility of public policies has been prioritised as part of moves towards rationalisation and performance-oriented management under the New Public Management and New Public Governance models (OSBORNE S., 2006). In the current climate of financial crisis and constrained public expenditure budgets, the learning opportunities offered by evaluation studies remain a crucial tool for policy-makers.

The European Union's (EU) requirement that interventions co-financed from its budget are evaluated is seen as a significant contributor to what commentators have termed an evaluation 'boom' in the past decade. Over successive programming periods, the EU Commission has introduced more sophisticated, rigorous and comprehensive approaches to Member States' evaluation of its policies and programmes (POLVERARI *et al*., 2007). An explicit motivation behind this is the generation of knowledge and information that can be used in the design and delivery of Cohesion policy programmes. Used effectively, evaluation results can offer a range of learning opportunities: for improved programme design and delivery; more efficient deployment of resources; improved management and implementation of programmes; greater understanding of 'causality' or the factors determining the success of programmes; and, broader scope to assess the value and costs of interventions. In fact, evaluation is seen as one of the clearest examples of policy learning between Cohesion policy practice and domestic policy systems. In several cases, Cohesion policy has been credited with growth in the status attached to evaluation, reflected in the number and quality of studies carried out and the expansion of overall evaluation capacity (BACHTLER & WREN, 2006: 143–53).

However, analyses of the extent to which Cohesion policy evaluations serve as an active instrument for policy learning are limited (FERRY & OLEJNICZAK, 2008). Who takes part in the learning process (individuals, organisations, Contracting Authorities, evaluators themselves, policy stakeholders, the wider community)? What do they learn (theoretical or conceptual understanding or practical knowledge)? How

does learning actually occur (e.g. through participation in the evaluation process, through dissemination, mentoring etc.) (SVENSSON *et al*. 2009: 32)?

There are significant challenges to the effective use of evaluation as a learning tool. In several Member States, evaluation is a new feature of policy environments and capacity for learning is still being developed. More generally, there are inherent tensions between the formal evaluation obligations of EU Cohesion Policy and more voluntary learning processes; between the desire to produce information on how well a programme is being implemented and the aim to produce knowledge of how well programmes actually 'work' in different contexts. There are also difficulties associated with generating knowledge for learning in a field such as Cohesion policy, which increasingly cuts across a range of sectors and draws on various actors situated at different territorial levels.

This chapter assesses how evaluation can serve as a strategic, learning tool in managing Cohesion policy. It is divided into three subsequent sections. The following section identifies the key learning opportunities that exist for policy makers, based on a brief review of evaluations Cohesion policy programmes in different Member States. Section 2 highlights some of the challenges currently facing those commissioning and conducting Cohesion policy evaluations in creating learning opportunities. The final section explores some emerging responses to these challenges.

1. What learning opportunities are offered by Cohesion policy evaluation?

Cohesion policy has become one of the most intensively evaluated policies in Europe (BACHTLER & WREN, 2006). To a certain extent, increased emphasis on evaluation has been driven by concerns about accountability and performance. The amount of Cohesion policy funding available has increased considerably over time. Moreover, the shift from a project-based approach to multi-annual programming and the decentralisation of increased programming responsibilities to Member States and regions highlighted the need to assess the performance. At the European level, the Commission uses the information generated by evaluations to monitor the efficiency and effectiveness of programming and to find out whether the promised targets are being achieved (RAINES & TAYLOR, 2002).

However, recent analyses of evaluations of EU-funded programmes have highlighted their potential contribution to other forms of policy learning. In theory (or in political discussions), Structural Fund evaluation has been interpreted more widely as a learning opportunity rather than simply a reporting exercise. Part of the rationale for this is that Structural Fund programmes are designed and delivered on a partnership basis, drawing on the knowledge and expertise of a wide range of vertical and horizontal partnerships. Different types of knowledge can be produced and,

drawing on examples from a range programme evaluations, there is some evidence of learning taking place (BACHTLER, 2001).

1.1. Improved strategic planning

First, evaluation can produce knowledge concerning the strategic orientation of a programme. This provides an opportunity to improve planning and learn more about the rationality and justification of programmes by verifying their internal coherence and relevance to existing needs. Evaluations of programmes can be particularly important in assessing whether resources are being efficiently deployed and whether adjustment of initial programme plans is necessary. This type of knowledge can provide a learning opportunity for programmes that are still underway or in planning for future programmes. For instance, a mid-term review of the Objective 2 Programme in North Rhine Westphalia 1994–99 presented a number of results, including: proposals for the orientation and prioritization of measures, the need for more systematic and comprehensive monitoring, and the institutional setting in which the responsible managing authorities were operating. The results of this were evident in the reprogramming of the Objective 2 programme in 1997–1999. An update of the mid-term evaluation and various thematic case studies were commissioned. On the basis of this, evaluators put forward proposals for the programming period 2000–2006. Following debate amongst economic and social partners, local authorities, and implementing bodies, several of these proposals became part of the NRW Objective 2 programme for 2000–2006. Major changes in this programme included: a re-orientation from infrastructure to business and innovation-related measures; the implementation of a systematic, computer-assisted project selection and monitoring scheme; and the setting-up of a technical secretariat to assist the managing authority in areas such as project selection, monitoring, reporting, and information. These were all strongly influenced by the preceding evaluation activities (JAKOBY, 2006: 281-4).

1.2. Improved management and delivery

Second, evaluations can produce knowledge to improve programme management and delivery. This refers to management structures and procedures and the quality of delivery. This is less strategic and more oriented towards efficiency than effectiveness. Nevertheless, this type of knowledge can also provide learning opportunities for future programmes. For instance, in Finland an updated Mid-term Evaluation exercise covering all if its 2000–2006 programmes was carried out in 2005. The use of these evaluations to improve the performance and quality of the 2000–2006 programmes was limited: the results of the original MTEs were still valid and the programmes were moving towards completion anyway. However, the updated MTEs made several recommendations that were more long-term. Hence, they become a useful source of learning for improving programme management structures and

procedures for the 2007–2013 period. For instance, the updated evaluations recommended the simplification of the management system and the integration of administrative procedures across different government departments. As a result, for the 2007–2013 period the administrative structure for OP management were rationalised.

1.3. Improved understanding

Third, evaluations can provide broader understanding by testing and verifying the theories and assumptions behind policies or programmes and drawing lessons for other interventions through the promotion of good practice. For instance, Scotland's Objective 2 programmes 2000–2006 were notable for the emphasis placed on sustainable development as an integrating or 'horizontal' theme. With the prospect of the mid-term evaluations of the 2000–2006 period, it was decided to incorporate benchmarking and international comparisons (with Nordrhein-Westfalen) into evaluations of sustainable development in the Scottish programmes. The aim was to achieve a deeper understanding of the issue in both countries through collaboration. This involved the organisation of workshops as part of a mutual evaluation exercise for evaluators and steering groups in both programmes. It also included the development of a common methodology to enable cross programme learning and comparison, drawing out the strengths and weaknesses of the approaches taken to the horizontal themes in different contexts. (RAINES & TAYLOR, 2002; DOWNES, 2003).

1.4. Improved coordination

Finally, evaluation can be used to broaden learning opportunities amongst those participating in a particular programme and strengthen practical knowledge about partnership working. This can relate to consolidating links or networks for cooperation between participants. In Sweden, for instance, national authorities undertook overarching reviews of all of the regional mid-term evaluations for the 2000–2006 period. The aim was to draw comparative lessons and disseminate these to programme managers and more widely in the policymaking community. The national agency NUTEK prepared a brochure containing an executive summary highlighting the key conclusions across all regional evaluation reports. On the basis of this, a national conference on Structural Funds experiences and regional development was organised. This was followed by regional seminars and conferences, which provided a platform for learning based on exchange of knowledge between regions and the national level. Evaluations can support learning in organisations that have previously played only a peripheral or marginalised role in the programme. For instance, one of the objectives of the interim evaluation of the Rural Development Plan 2000–2006 in the Basque Country was to include as broad as possible participation from stakeholders. This was deemed to have increased utilisation of evaluation

results (ZQUIERDO, 2006). There is growing interest in the role that evaluation can play in strengthening governance in general (GORE & WELLS, 2009:158-67).

2. What are the Current challenges?

By generating different types of knowledge, therefore, Cohesion policy evaluations can potentially provide learning opportunities for policy makers or programme managers as they consider the strategic orientation of programmes and arrangements for their delivery. Evaluations can also contribute to broader learning on particular policy themes and on partnership working. However, academic and policy assessments of Cohesion policy evaluations conclude that their use as tools for learning has, thus far, been limited (BATTERBURY, 2006:179-88; ESER & NUSSMUELLER, 2006:249-58). The following challenges to evaluation learning can be identified.

2.1. Capacity issues

Developing the evaluation capacity to optimise the scope for learning is an ongoing challenge in an increasingly complex policy field. The extent to which Member States can generate and utilise knowledge as a result of evaluation depends on the resources it can dedicate to the process. This can concern the quality of human resources of the Contracting Authority (CA) that is contracting and supervising the research. Once staff have gained experience in evaluation, they better understand the specific benefits of evaluation and they are aware of how it can be used (BOYLE & LEMAIRE, 1998). In some cases, the willingness of senior officials to use evaluation as a learning tool is closely linked to the fact that their university and professional training have raised awareness of this function (THOENIG, 2000:217–229). The stability of an institution can also be important. Once an institution is stable and well-established, with good experience in dealing with interventions similar to the evaluated case (i.e. once it has an 'institutional memory') it can accumulate knowledge from evaluations and refer to it in new situations. Finally, the position of the institution in the policy system can be important: the ability to allocate or channel resources, to promote results and make changes can have an obvious impact on the extent to which an evaluation is used.

The degree to which institutional frameworks are in place to promote evaluation learning varies across EU Member States (and even across regions within EU Member States). Research has identified a 'north-south' divide (BACTHLER & WREN, 2006). In countries such as the Netherlands, the UK, Germany, France and the Nordic countries, the evaluation of regional policy has a stronger tradition. Regular evaluations of national regional policy interventions take place and the knowledge produced by evaluation is an integral part of the policy-making process. By contrast, in Italy, Spain, Portugal and Greece policy evaluation has not been well-established within the public administrative culture, evaluations have been seen primarily as

a periodic reporting exercise and have been used for broader learning only intermittently and in an *ad hoc* way (CASAVOLA & TAGLE, 2003). A further group of countries consists of newer Member States, where evaluation capacities are growing rapidly but from a very low base, driven largely by Structural Funds requirements. These variations also apply to the 'supply side'. The strength and characteristics of evaluation expertise outside of public administration can influence the extent to which studies are used for learning purposes. On the one hand, there has to be sufficient expertise and experience on the supply side to produce useful knowledge. Analysis of some of the MTEs carried out in the 2000–2006 period noted that the supply of evaluators was stretched in some Member States and that this had a strong impact on the quality of the reports and thus on their potential utilisation (EPEC, 2005). On the other hand, external evaluators can include functions that educate or facilitate learning as part of a participative evaluation research process (FERRY & OLEJNICZAK, 2008).

These variations notwithstanding, it is important to note some common challenges. First, regardless of existing strengths and weaknesses, all EU Member States continue to prioritise evaluation capacity building in different areas to improve the scope for learning. Second, even with strong evaluation capacity (internal and external experts, a robust database etc.) the use of evaluation as a tool for policy learning relies on political will and support. The use of good quality evaluations for learning may be limited where only positive conclusions are presented or, when the report is manipulated, shortened or held back for political reasons (ALKIN & TAUT, 2003:1-12).

2.2. Characteristics of the policy being evaluated

Characteristics of the policy being evaluated can have a direct impact on how an evaluation is used. The features of a given policy field, the size and importance of the intervention (in terms of resources, geographical scope and time-scale) can all be influential. Evaluations of interventions covering fields that are high on the political agenda are clearly of greater potential interest to managers and policy-makers as well as the wider public. EU Cohesion policy is funding important interventions, particularly in newer Member States. Moreover, its role is currently the subject of increasing debate across the EU. However, up to now those wishing to learn from Cohesion policy evaluation have had to overcome limited understanding of the effectiveness or impact of interventions. The evaluation of Cohesion policy programmes has had a tendency to concentrate on the question of "why it works", often based on qualitative methods rather than "what works", or what effects particular interventions produce (STRYCZYŃSKI, 2008). After over two decades of Cohesion policy, the very limited information on "effects" makes it extremely difficult to learn about "how" an intervention works or does not work, or about how to draw lessons for the implementation of other programmes in other contexts (BARCA, 2009). There are specific challenges associated with generating evaluation knowledge for policy

learning under Cohesion policy. Current Cohesion policy interventions can cover a large number of components, crossing the boundaries of regional and sectoral policies and including activities that can have 'soft' or 'intangible' impacts, at least in the short-term (e.g. skills development, innovation support, technology transfer, advisory services). For the purposes of evaluation, it makes it more difficult to establish lines of causality between different activities and impacts. The devolution of powers of policymaking and delivery to multiple assemblies and agencies, under the Cohesion policy principles of partnership and subsidiarity, contribute to this complexity and create problems with accessing consistent, comparable data. How do you measure subtle, complex or intangible policy impacts? How can the individual contribution of a variety of institutions and initiatives from various levels and policy fields be identified and evaluated? As a result, it is increasingly difficult for evaluators and policy-makers to draw lessons about the external impact and effectiveness of programmes and the factors that decide this (MCVITTIE & SWALES, 2003).

2.3. Research timing

Timing evaluations to maximise learning opportunities is also challenging. Research timing can influence the use of evaluation for learning in two ways. First, the timing of the research in relation to the stage of the programme cycle can be crucial. It is widely recognized that *ex ante* evaluations focus on planning functions. This concerns strategic issues such as the a priori assessment of a plan's rationality. It also relates to operational considerations, notably the design of the implementation system. On-going evaluation address mainly the management, partnership and learning of good practices issues that is things of operational nature. Moreover, in advanced interventions they can draw more strategic lessons concerning initial impacts. *ex post* research focuses on strategic topics, accountability and assessments of the theories and assumptions on which the intervention was built. In all of these cases it is vital that the evaluations are timed to feed into the relevant stage of policy planning and delivery. Second, the timing of the research in the overall policy cycle can be important. If a policy field, such as Cohesion policy, is in transition between old and new approaches or undergoing reforms or is high on the political agenda for some other reason, related evaluation studies may receive particular attention.

2.4. Evaluation focus and approach

The focus of an evaluation and the quality of the process is another determinant of its use for learning. A distinction can be made between the specific learning opportunity offered by reporting on programme progress (generating knowledge on the performance of programmes for the immediate use of programme managers), and broader learning opportunities (undertaken to improve understanding, quality and effectiveness) (RAINES & TAYLOR, 2002). Choices of evaluation focus will impact on the weight given to objective analyses of programme performance and processes

of interaction, reflection and learning in the evaluation approach OWEN & ROGERS, 1999). In Cohesion policy, as in other policy fields, the focus on performance management in evaluation has, up until now, predominated: it is usually value-for-money oriented and in, it is easier to generate reasonably robust information on a programmes physical and financial progress in a short time frame. However, an inherent danger is that the prominence of efficiency and accountability functions – demonstrating how programme implementation has progressed – can lead to the evaluation being seen as simply a report for the benefit of higher authorities, notably the European Commission. This can diminish the potential for alternative uses of the evaluation for learning at the programme level and beyond. Moreover, efforts to combine the twin aims of accountability and learning can affect the focus and ultimate utility of an evaluation. For instance, the mid-term evaluations (MTEs) of the 2000–2006 programmes aimed to review the implementation of programmes at their mid-point and makes proposals for future learning and improvement. According to some authors, in practical terms, the MTEs covered a function between accounting and learning, resulting in a lack of clarity and focus that in some cases detracted from the utility of the exercise (ESER & NUSSMUELLER, 2006).

2.5. Dissemination

A final challenge is ensuring that the results of an evaluation are properly disseminated. In order for learning to take place, insights and knowledge generated by evaluations must reach the relevant organisations and individuals. This refers to the relevance, clarity and accessibility of the evaluation. Beyond the danger that some evaluation results can be held back for political reasons, the challenge is to produce sound conclusions and concise, realistic proposals and to communicate these to the key audiences. To a certain extent, more efficient dissemination has become a necessity for Cohesion policy, given the devolved system and the volume of studies produced. For instance, rather than individual reports, DG Regio produced a synthesis of evaluation results for the Objective 1 and 2 MTEs in the 2000-6 period. However, as with other aspects of the evaluation process, interpretations of this framework vary across Member States. According to recent analyses, although the dissemination of evaluation reports is relatively good, it could be improved (ESTEP, 2007:76).

3. Drawing learning opportunities from evaluations: current practice

How can we anchor evaluation knowledge and experience in institutions and thus ensure that evaluations contribute to strategic learning? How can evaluation processes be improved to support a process of learning and encourage the effective use of the knowledge that they produce? The final section of the paper makes some recommendations based on current practice.

3.1. Strengthening the evidence base for learning

A clear message from this review is the importance of a strong base of programme information in determining the credibility and therefore utility of evaluations as opportunities for policy learning. This relates particularly to the generation of knowledge on the impact or effectiveness of programmes. In this context, it is of note that DG Regio is aiming to institutionalise a more strategic and results-oriented approach by introducing a flexible evaluation framework for the 2007–2013 period. Mid-term evaluations are now optional, replaced by on-going, needs-based evaluations to assess programme implementation and react to changes in the external environment. Evaluations may also be triggered by actual or potential difficulties revealed by the monitoring system. According to their own needs, Member States can decide what level of evaluation is required (whether based on programmes, themes or Funds). At the same time, the European Commission continues to work on different methodologies to address this challenge, including piloting studies, case studies and "counterfactual" approaches (STRYCZYŃSKI, 2008). Several Member States are placing particular emphasis on impact evaluation, identifying, measuring and aggregating different types of gross effects. For instance, for the 2005 Mid Term Evaluation in France, a national framework for the quantification of programme effects in the labour market was developed. The aim was to overcome inconsistent approaches to data collection at the regional level.

3.2. Strengthening capacity for learning

There are ongoing efforts to build capacity for learning across Member States. In several cases, there is a strong focus on the organisation of staff training courses, workshops and consultations. In some cases, particularly newer Member States, institutional flux and frequent staff rotation make it challenging to anchor knowledge and expertise gained in evaluation processes when the people directly involved move on. One solution is to ensure public databases of all evaluation studies are available as a source of learning (Poland has already started the process of collecting and storing all evaluation reports). Thus, in cases where staff or institutions have changed, their replacements could have a point of reference and at least partial memory of the past programme, its problems, the solutions that have been found, etc.

Evaluation systems in several Member States are increasing their coordinating capacity in order to encourage cross-programme learning. In some cases, this is being achieved through increasing the role of national-level organisations in coordinating evaluation work across programmes and regions. In Italy, the National Evaluation System involves representatives from the evaluation units of the regional and national administrations. Another frequently-cited example can be found in Austria. Following accession to the EU, the Austrian Federal Chancellery established the so-called 'Checkpoint EVA' platform. This supported exchange of experience and

learning in regional policy evaluation. The network included evaluators and national as well as regional policy makers. 'Checkpoint EVA' collected and distributed information on evaluation issues and organised workshops and annual conferences. The platform continues to play an important role in the current programming period. In some cases, efforts to raise evaluation knowledge and awareness have included the creation of international networks. For instance, there is a network of civil servants from the Visegrád countries in the field of Structural Funds evaluation. The aim is to promote learning and exchange of experience and it has contributed to the dissemination and follow-up of evaluation results. In some Member States, such as Sweden, coordination through meta-evaluation is key to the evaluation strategy. This involves a sophisticated system that coordinates different project, programme and thematic evaluations and integrates their results (BRULIN, 2009). There are ongoing efforts to improve general levels of evaluation knowledge and learning outside of public administration. Academic participation in evaluation can be part of this, and this can be strengthened where evaluation is studied as a discipline in universities and research centres. We have already noted a broad tendency to invite the participation of the evaluation community in conferences, workshops and seminars organised by public authorities. Evaluation societies can contribute to this.

3.3. Designing an evaluation approach to encourage learning

Activities under this heading can be grouped according to those aimed at the Contracting Authority (CA) and those aimed at the evaluator. For the CA, deciding the aims and structure of the evaluation process before a study is contracted out can be vital to improving its potential as a learning tool. The CA is required to make some important decisions concerning: the focus of the study (e.g. on process issues or effects); the methodology (e.g. interactive, theory-based etc.); and, practical issues such as timing, costs and data availability. In fact, the close involvement of the CA at all stages of the evaluation research supports learning in a number of ways: it provides a 'champion' for the project, ensuring that it continues to be seen as a priority in the institution; it boosts evaluation experience in the CA; it ensures that the evaluation research meets the needs and interests of the institution; it can help to ensure the participation of other stakeholders and beneficiaries in the evaluation process; and, it can ensure that evaluation findings are disseminated properly. In several cases, CAs are building on past experience and the European Commission's new guidelines on evaluation to play an increasingly active role. For instance, the Northern Periphery Programme 2007–2013 has initiated a number of activities to support learning as part of the on-going evaluation process. First, the *ex ante* evaluation of the 2007–2013 Programme raised a number of issues and highlighted potential challenges that are relevant to the on-going evaluation of the programme (MCMASTER *et al.*, 2006). Second, an Evaluation Steering Group was established to facilitate and monitor the ongoing-evaluation process. The EVA-group consists of representatives of the Programme Managing Authority (MA), the Joint

Programme Secretariat (JPS), Regional Contact Points (RCPs), and national representatives. Third, an Evaluation Strategy was agreed for the Programme, which sets out an overall framework for on-going evaluation and effective quality management, including linkages between monitoring and evaluation activities. The Programme has completed a series of internal evaluation exercises, including 'touchstone' surveys involving the project lead partners, Regional Advisory Groups (RAG), Regional Contact Points (RCP), and the JPS. Finally, a strategic programme overview has been prepared, addressing thematic priorities. These pre-evaluation activities facilitate learning by identifying the key areas for knowledge generation in the subsequent external evaluation of the programme (NPP:4).

Similar issues, (concerning evaluation objectives and scope, resource issues, data constraints) face the evaluator. Evaluators can perform a range of functions depending on the orientation of evaluation: as a relatively detached consultant; as a scientific expert; as an animator or facilitator of stakeholder participation; and, as a mentor or educator, encouraging a process of learning amongst stakeholders. There is no optimal model for the evaluator's role. Several hybrids can exist within this broad categorisation. However, if learning is a priority, evaluators should follow the role of animators (where studies focus on processes) and educators (where studies focus on the assessment of effects or impacts). Educating and facilitating learning as part of the research process (by making the process interactive) is crucial.

3.4. Creating learning opportunities through dissemination

A final point relates to the role of dissemination in ensuring the use of evaluation for learning purposes. Innovative approaches have emerged in recent years to communicate evaluation findings to relevant audiences in an efficient way. For instance, the MTE of the Highlands & produced a series of thematic Working Papers throughout the course of the evaluation. These covered different sectoral themes and management and implementation issues, such as rural development, the horizontal themes, and the programme's contribution to the European Employment Strategy. Rather than wait till the end of the study, each of these reports were disseminated as soon as they were ready. This ensured that the material gathered could be adapted to feed in to the programme context but also into the broader debate in Cohesion policy and academic circles. All of these papers were compiled in a Technical Report that was published alongside the Final Evaluation study. Targeted dissemination events were organised, dedicated to specific themes. This made evaluation findings more accessible and increased opportunities for learning to take place. Interested parties could focus on the output that was most relevant to them. This approach was also dynamic: dissemination events allowed the evaluators to engage in a dialogue with interested parties and stakeholders as the evaluation progressed.

Conclusions

The learning function of evaluation is emphasised in current strategies for public policy, including regional economic development. As development policy becomes more sophisticated and complex (drawing in multi-level and multi-sectoral networks and partnerships, based on programmes with comprehensive and diverse objectives) the learning opportunities provided by evaluation are seen as key to guiding current and future interventions. This is reflected in the evaluation requirements for Structural Funds programmes in the current programming period which encourage flexible, ongoing, needs oriented approaches. For the purposes of improved learning, the aim is to generate relevant knowledge and transfer it to the appropriate audiences as efficiently as possible. Generating learning opportunities from evaluation studies is challenging and certain conditions must be met: the establishment of institutional capacity in public administration and in the evaluation community; access to a robust evidence base; proximity to and interaction between Contracting Authorities, evaluators and policy stakeholders in the evaluation process; initiation in the earliest stages of the programme or project; and, commitment to ongoing and iterative feedback channels and dissemination provisions. Meeting these conditions requires a significant investment of time and resources, and, crucially, a commitment to learning on the part of public administrations and stakeholders that is likely to go beyond the timescale of a given programme. Nevertheless, evaluation learning has a range of potential benefits, particularly for large, long-term development programmes: it can produce more strategic and effective programmes and projects; it can improve the quality of policy delivery; it can contribute to broader theoretical knowledge of specific policy fields; and, it can strengthen relationships between different parts of the policy community.

Notes

This chapter is based on research carried out by the author and Dr. Karol Olejniczak, Ferry, M. & Olejniczak, K. (2008) *The use of evaluation in the management of EU programmes in Poland*, Warsaw: Ernst & Young – Program “Sprawne Państwo” and the IQ-Net research consortium organised by EPRC at the University of Strathclyde.

Bibliography

ALKIN M. C., TAUT S. M. (2003) 'Unbundling evaluation use', *Studies in Educational Evaluation* 29.

BACHTLER J. (2001) 'Quod erat Demonstrandum? The Evaluation of Regional Policy' in Eckey HF, Hecht D, Junkernheinrich M, Karl H, Werbeck N and Wink R (Editors) *Ordnungspolitik als konstruktive Antwort auf wirtschaftspolitische Herausfoerderungen – Festschrift zum 65. Geburtstag con Paul Klemmer*, Lucius & Lucius Verlag, Stuttgart.

BACHTLER J., WREN C. (2006) 'Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges', *Regional Studies* 40(2).

BARCA F. (2009) *An Agenda for a Reformed Cohesion Policy A Place-Based Approach To Meeting European Union Challenges and Expectations*, independent report prepared at the request of Danuta Hübner, Commissioner for Regional Policy.

BATTERBURY S. (2006) 'Principles and Purposes of European Union Cohesion policy Evaluation', *Regional Studies* 40(2).

BOYLE R., LEMAIRE D. (1998) *Building effective evaluation capacity: lessons from practice*. Transaction Publishers, New Brunswick, NJ.

BRULIN G. (2009) 'Multilevel On-going Evaluation Approach: Experiences from the Swedish Approach', conference presentation at E*valuation of EU Structural Funds: Reinforcing Quality and Utilisation*, Vilnius, March 2009.

CASAVOLA P., TAGLE L. (2003) Building Capacity for evaluation: Lessons from Italy, *Fifth European Conference on Evaluation of the Structural Funds "Challenges for Evaluation in an Enlarged Europe"*. European Evaluation Society, Budapest, 26-27 June 2003.

DOWNES R., TRENKLER-FRASER R .AND TAYLOR S. (2003) *Mainstreaming the Horizontal Themes in the Objective 2 programmes of Nordrhein Westfalen, East and West Scotland: Comparative Report,* Report to the Scottish Executive and the Land Nordrhein Westfalen (NRW), Ekos and European Policies Research Centre, Glasgow.

ESER T. W., NUSSMUELLER E. (2006) 'Mid-term Evaluations of Community Initiatives under European Union Structural Funds: a process between accounting and common learning', *Regional Studies* 40(2)

ESTEP (2007) *Final Report on the Framework to Analyse the Development of Evaluation Capacity in the EU Member States*.

FERRY M. & OLEJNICZAK K. (2008) *The use of evaluation in the management of EU programmes in Poland*, Warsaw: Ernst & Young – Program "Sprawne Państwo".

FURUBO J.-E., RIST R. C., SANDAHL R. (2002) *International atlas of evaluation*. Transaction Publishers, New Brunswick, U.S.A.

GORE T., WELLS, P. (2009) 'Governance and Evaluation: The case of EU regional policy horizontal priorities' *Evaluation and Program Planning*, 32.

JAKOBY H. (2006) Evaluation as part of the regional policy life cycle: the example of North Rhine-Westphalia, *Regional Studies* 40(2).

MCMASTER I., VIRONEN H., MICHIE R. (2006) *Ex Ante Evaluation of the Northern Periphery Programme,* Final Report to the Managing Authority, November 2006.

MCVITTIE E. AND SWALES J.K. (2003) 'Constrained Discretion in UK Monetary and Regional Policy' paper presented at the *Regional Policy in the UK* seminar, LSE, November 2003, organised by the ESRC Urban and Regional Economics Seminar Group.

NPP PROGRAMME 2007–13, *Evaluation Strategy*.

OSBORNE S. (2006) 'The New Public Governance?' *Public Management Review* 8(3).

OWEN J. M., ROGERS P. J. (1999) *Program evaluation: forms and approaches*. SAGE, London.

POLVERARI L., MENDEZ C., GROSS F, BACHTLER J. (2007) 'Making Sense of European Cohesion Policy: 2007-13 Ongoing Evaluation and Monitoring Arrangements', *IQ-Net Thematic Paper*, 21(2) EPRC, University of Strathclyde, Glasgow.

RAINES P., TAYLOR S. (2002) Mid-term Evaluation of the 2000–06 Structural Fund Programmes, *IQ-Net Thematic Papers* 11(2).

STRYCZYŃSKI K. (2008) 'Rigorous impact evaluation: the magic bullet for evaluation of Cohesion Policy?' presentation at European Evaluation Society conference, *Building for the future: Evaluation in governance, development and progress*, Lisbon, October 2008.

SVENSSON L., BRULIN G., JANSSON S. and SJÖBERG k. (2009) *Learning Through Ongoing Evaluation* Studentlitteratur, Lund, sweden.

THOENIG J.C. (2000) 'Evaluation as Usable Knowledge for Public Management Reforms' *Evaluation* 6(2).

ZQUIERDO B. (2006) The Use of Evaluation in multi-actor Context: A Participatory Experience in the Basque Country, *Easy-Eco conference Improving the Quality of Sustainable Development Projects*. Easy-Eco, Saarland, October 2006.

Evaluation of public policies as an instrument of establishing knowledge-based administration in Poland

Stanisław Bienias, Tomasz Gapski

1. Evaluation of public policies as an instrument of establishing an efficient State

The most important mission of public authorities covers establishment of an efficient State, which should aim at setting up conditions for improving the quality of life for citizens and possibilities of running economic activity. In order to properly fulfils its role the efficient State must strive for efficient and effective management of resources and needs well-qualified public administration. This management is based on clear definition of strategic and operational objectives, as well as on development and implementation of action plans aimed at their achievement. The plans and actions of public authorities as well as results thereof, should be clearly communicated to the public –that public which is the beneficiary of public policies and the customer of public sector and which actually finances the actions of those authorities.

Targeting the actions of the public administration at achieving objectives and results, and thereby departing from archaic methods of operation which focus the attention and efforts solely on procedures, is a prerequisite for smooth execution of the above process.

The entire cycle of running public policies from the moment of identifying the problem and forming the objectives and methods of action until the moment of achieving the assumed effects is not carried out in a void. It must refer, in a continuous manner and under all its aspects, to the current socio-economic context, as well as to the effects of previous actions (evidence-based policy) and theoretical knowledge within a given field (theory-based policy). In other words, in order to fulfil their role properly, public authorities must **know how to manage knowledge** for the actions they undertake. Hence, in the contemporary world one of the major challenges for the public authorities is to **establish knowledge-based administration**.

The knowledge which is necessary in the process of formulating and implementing public policies may be generated internally within the administrative frameworks through accumulation of knowledge and experience. It may also be generated externally – through skilful use of knowledge, practice and experience of entities other

than public administration. The knowledge management system may, therefore, be based on the administrative capacity by establishing centres, which are to provide to the public authorities the knowledge necessary for its smooth operation (administrative study centres, internal evaluation units, experts outside the main hierarchical structure, etc.) or it may benefit from the services of external entities (academic centres, think tanks, external evaluators, consultancy companies, non-governmental organisations, etc.). These processes, which take place in the public sector, correspond to the process of knowledge and technology transfers and growth of the R&D, which take place in the private sector. The process which today constitutes one of the basic conditions for socio-economic development. The public sector is subject to similar socio-economic conditions of global dimension, which impose on it an innovation requirement and a need to "govern" in an efficient manner, while keeping the ability to respond quickly and efficiently to the unexpected challenges in the national and international policy.

Evaluation of public interventions enables to place the implemented policies in the context of objective facts, data and socio-economic scientific theories and allows for efficient and effective knowledge management in the administration.

What does evaluation mean in practical terms? Evaluation uses socio-economic studies and analyses to seek an answer to the question: what can be done to improve the implementation of public interventions (projects, programmes, strategies, policies). Evaluation studies aim at providing reliable information, which are subsequently used in the process of taking the key decisions that optimize the expenditure of the public resources. They can be implemented at each stage of implementing public policies (*ex-ante* evaluation, ongoing evaluation and *ex-post* evaluation). The basic idea is the search for answers to previously formulated evaluation questions, which refer to the effects of the actions carried out by public authorities and their assessment with the application of relevant evaluation criteria (efficiency, effectiveness, accuracy, usability, sustainability of public interventions). Each evaluation study should lead to conclusions, as well as useful and applicable recommendations targeted at public authorities.

The literature on this topic covers numerous characteristic features of the evaluation function and objectives that it should implement. On this basis, as well as on the grounds of the experience of Polish administration gathered so far and the context of its operation, it is possible to define three major strategic objectives, to which implementation evaluation of public policies in Poland should contribute:

First: establishment of knowledge-based administration, which is at the same time targeted at objectives and effects instead of procedures

Second: improved process of spending public resources by focusing the resources on efficient actions and resigning from actions, which fail to bring the expected results or waste public resources

Third: setting up a communication system for the public authorities and society in reference to objectives, strategies and effects of the undertaken actions

Evaluation as an instrument for collecting, processing and submitting information on the effects of implemented policies contributes to the establishment of modern knowledge-based administration. Furthermore, it constitutes one of the basic instruments of management targeted on achieving objectives. Popularization and promotion of evaluation culture in public administration fosters the process of transforming the "procedures-oriented administration" into the "results-oriented administration." The process of ongoing evaluation of undertaken actions makes the public aware of the primary role of strategic and operational objectives which in turn results in a "more flexible" functioning of administration through more active search for the most efficient instruments of implementing the planned objectives and departure from fixed procedures. Adopting the objectives-oriented perspective instead of limiting oneself to the procedures and the "scope of competencies" gives a greater freedom of action and allows to rapidly and efficiently respond to the unexpected challenges that appear in the national and international policy.

Effective expenditure of the public resources is one of the basic conditions of establishing the efficient State. Inefficient structure of spending public resources is one of the major reasons of a weak State, leading to inefficiency and simultaneously increasing the tax burdens for the society. Therefore, one of the most significant missions of evaluation is the efficiency assessment of public interventions in order to formulate a justification to modify or liquidate inefficient instruments used within the frameworks of implemented public policies. Hence evaluation is a tool to "release" misallocated financial resources and their optimal reallocation, thus providing suitable conditions for the establishment of an efficient development policy.

Democratic legitimacy of public authorities is the main pillar of an efficient State (apart from undertaking of effective and efficient actions). Authorities are responsible for meeting the transparency rules of performed activities with special emphasis on the need to communicate the assumptions and results of implemented policies to the society. Dissemination of evaluation results is a useful instrument for executive authorities to communicate with the public. It also makes possible to incorporate the the public's opinions into the process of forming and implementing public policies.

2. Cohesion Policy evaluation system in Poland

From the perspective of other European countries, where the evaluation has been functioning for many years as an indispensable part of the management process in the public area, Polish experience in this regard is limited. Before Polish accession to the European Union no principles related with management targeted at results were introduced on a wider scale and evaluation was not used as an instrument to assess the efficiency and effectiveness of public expenditure. It followed both from the heritage of the former socio-economic system, under which the meaning of the terms efficiency and effectiveness were unknown, as well as from the fact that at political level there was no vision of the model of modern State operation, and thereby issues associated with the organisation of the public area management system were still underestimated.

However, the situation has changed for better after 2004. The membership in the European Union and Polish participation in Cohesion Policy forced the application of the EU principles within the scope of evaluation following from the regulations on Structural Funds and preparation of relevant national provisions (*inter alia*, the Act of 20 April 2004 on National Development Plan[1)]). The implementation of the National Development Plan (NDP)/Community Support Framework (CSF) 2004–2006 became a turning point in the Polish history of evaluation. It was the beginning of a systematic, coordinated and based on an uniform system of indicators, efficiency assessment of the instruments used within the framework of Cohesion Policy, efficiency assessment of administrative activities in that regard, as well as a global assessment of the impact of Structural Funds on the socio-economic situation[2)].

Despite the fact that Poland has begun to use evaluation in a systematic manner only since 2004, it now becomes one of the leaders in this respect among other Member States. The involvement of Polish administration in the establishment of a broadly-conceived evaluation culture under Cohesion Policy was recognized by devoting to the issue a separate "good practice" case study under the *ex-post* evaluation of the 2000–2006 period implemented on the request of the European Commission.

---

1 Dz.U., No. 116, item 1206.

2 Bienias S., Sudak S. /ed./ *Proces ewaluacji polityki spójności. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość [Cohesion Policy evaluation process. Summary of the current experience. Plans and challenges for the future],* National Evaluation Unit, Department of Structural Policy Coordination, Ministry of Regional Development, Warsaw 2008.

2.1 Evaluation in Poland – historical overview[3)]

(Before 2004) – Evaluation of pre-accession funds

Evaluation in Poland is strictly connected with the EU accession process. At the very beginning evaluation activities were related to the pre-accession instruments such as: Phare, SAPARD and ISPA and undertaken only on an *ad hoc*, quite rare basis. There were only few institutions that touched upon the issue of evaluation in the form of contracts, training or publications. However, their activities were totally isolated from each other and little information was exchanged between the partners involved in the process. In 2000, the Polish Evaluation Society was founded, but it was strongly associated with the field of education and human resources programmes run by the Office of the Committee for European Integration[4)].

Most evaluations of pre-accession programmes were contracted by the European Commission directly. Between 1999 and 2004, Polish authorities contracted only around 10 studies – all of them were *ex-post* in nature (assessing the effects of executed programmes). The studies sprang from the programme life cycle, not from a systemic (policy) point of view. That is why, they aimed at fulfilling the current information needs of the managing units rather than putting the findings in the wider context of policy strategic development[5)]. To conclude – in the pre-accession period the evaluation was not used in an systematic and coherent way.

(2004–2005) Entering the EU – preparatory stage

In May 2004, when Poland joined the EU, evaluation system in Polish public administration had been virtually non-existent. The first task was to build an appropriate evaluation capacity in respective institutions and to start launching first evaluations.

The following model of structural funds evaluation was applied in Poland: 1) all evaluations were outsourced to independent external evaluators selected in accordance with the Public Procurement Act, 2) evaluations within NDP/CSF 2004–2006 were conducted at two levels: the NDP/CSF level and the level of operational programmes (the NDP/CSF evaluations being a responsibility to National Evaluation Unit, while evaluations within operational programmes (OP) were commissioned by six evaluation units located within the structures of managing authorities), 3) evaluation results were discussed (from the perspective of their accuracy and usefulness) on evaluation steering groups' meetings and then handed over to decision

---

3 This chapter is based on the publication S. Bienias, I. Lewandowska ed. *Evaluation in the Visegrad countries, MRD,* 2008 and presents only the most important information .

4 1 K. Olejniczak, *Mini-Case Study: Building the evaluation system in Poland Ex-post Evaluation of Cohesion Policy Programmes 2000-06 Co-financed by ERDF*. in: *Work Package 11: Management and Implementation System for Cohesion Policy*, 2009, p. 6–7

5 Ibid, p.7.

makers for further use, 4) all evaluation reports were made available on a website devoted to evaluation.

All evaluation researches that were planned for the 2004–2006 programming period were included in "Evaluation Plan of the National Development Plan for 2004–2006"[6] – a document, which was aimed at providing for effective management of the evaluation process in Poland. The Steering Group for the NDP/CSF evaluation (which aimed at coordinating horizontally all evaluations carried out at different levels of EU funds implementation) was appointed, and sector steering groups to manage the evaluation process within the Sectoral Operational Programme Human Resources Development and the EQUAL Initiative were set up.

Methodological guidelines for structural funds evaluation were formulated, namely "Evaluation of the National Development Plan and Operational Programmes in Poland – Guide"[7].

Due to a limited staff capacity (evaluation units functioned only within the Sectoral Operational Programme Human Resources Development, EQUAL Initiative, Sectoral Operational Programme Improvement of the Competitiveness of Enterprises and National Evaluation Unit), procedural problems in the field of public procurement and delays in implementing technical assistance funds, only few ongoing evaluations were conducted in 2005. These were mainly low-budget evaluations of small scale, focusing on selected fields of the EU funds' interventions. In many cases, they were pilot evaluations and the usefulness of their results did not prove sufficient.

(2006) *Ex-ante* evaluations of 2007–2013 perspective

The external assessment within the *ex-ante* evaluation covered the National Strategic Reference Framework and 22 operational programmes, including 16 regional programmes (ROP). The process of the actual evaluation was supported by macroeconomic impact forecasts and strategic environmental impact assessments. The *ex-ante* evaluation of the National Strategic Reference Framework 2007–2013 and the operational programmes was the largest undertaking of this kind. In less than a year, an analysis and assessment of the plans of spending a sum amounting to more than € 65 billion was conducted.

---

6 After extensive counsel with administration structures and European Commission services, the document was accepted by the NDP Monitoring Committee on December 7, 2005. However, it turned out difficult to carry out numerous actions strictly connected with the evaluation processes, due to the fact that the scope of particular researches was defined too precisely, which, as it turned out later, did not quite meet the actual current information needs of the institutions managing the EU funds. However, the experience gathered during the Plan implementation was made use of when creating the current on-going evaluation system.

7 The guide was worked out by National Evaluation Unit in 2005. Results of external surveys outsourced by the former Ministry of Economy and Labour and conducted by B. Ciężka, J. Ratajczak, K. Olejniczak, T. Skierniewski and B. Ledzion were used (full texts are available on www.ewaluacja.gov.pl under: Dokumenty ewaluacyjne/Ekspertyzy, analizy i opracowania).

The main determinants of the process of the *ex-ante* evaluation, were the scale and complexity of research process, considerable time pressure and low evaluation capacity of Polish administration (insufficient especially at regional level). Having taken into consideration the above aspects, a decision to centralize the evaluation process at the level of the Ministry of Regional Development was made. The competence in the field of *ex-ante* evaluation was concentrated in Managing Authorities for operational programmes, managed by central institutions, and in the National Evaluation Unit – for Regional Operational Programmes. Moreover, in the case of regional programmes, a decision was taken to commission evaluation of all programmes (including the Operational Programme Development of Eastern Poland) to one evaluator. In order to coordinate the whole process, the Coordinating Committee for Evaluation Process of Operational Documents 2007–2013 was appointed. Research guidelines and their results were presented to the Committee and conclusions and recommendations were discussed by that Committee. Also, a team of researchers from different fields covered by the research was appointed at the National Evaluation Unit to provide support in coordination and management of the evaluation process, including the receipt of the results[8)].

The usefulness of the *ex-ante* evaluation process, varied considerably and depended mainly on the quality of the evaluators' work. Due to the time pressure and advanced stage of elaboration of operational programmes the recommendations that were used were usually of the technical, and not strategic nature. The quality of final reports was assessed by the commissioning institutions on a scale of 2 to 5 (5 – highest); most of the reports were assessed as good, with the average grade around 4.

The experience gathered during the *ex-ante* evaluation, which was a big organisational challenge for the Polish administration, allowed to reorganize the Polish evaluation system and redefine its major principles. In 2007, an on-going evaluation system was launched which integrated the evaluation processes of National Development Plan for 2004–2006 and National Strategic Reference Framework for 2007–2013. The basis for the decentralised evaluation system was an organisational and institutional structure that would guarantee compliance with the partnership principle in the broadest possible scope.

2.2. Cohesion Policy evaluation system

Polish evaluation system, which has been created in 2007 is strongly decentralised. Competence for commissioning evaluation researches has been granted to institutions most interested in conclusions and recommendations from the evaluation at the appropriate level of implementation. This organisation allows evaluation to be the closest to the real problems and at the same time – through involvement of

[8] One of the effects of the Team's work was a publication summing up the process of the *ex-ante* evaluation "*Ex-ante* evaluation – summing up the experience of Polish administration" available on www.ewaluacja.gov.pl.

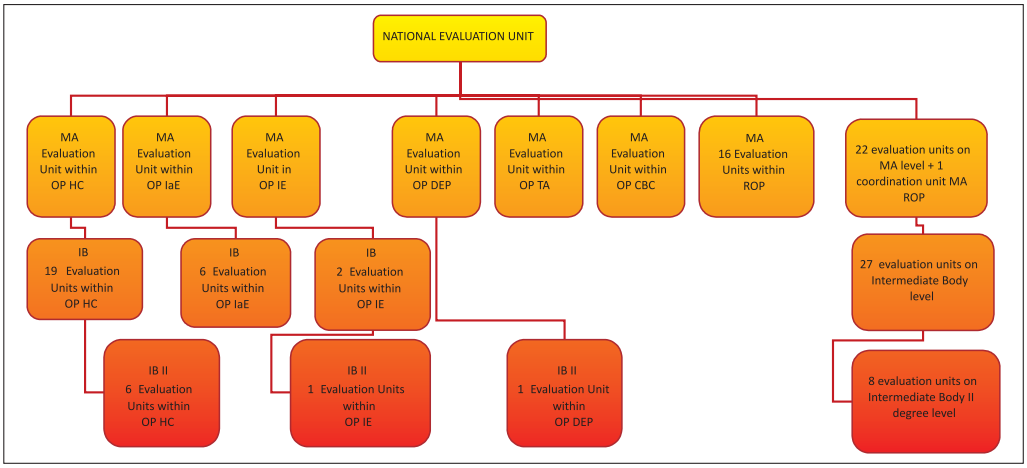
a large group of stakeholders (from implementation system) – allows for the adequate identification of evaluation studies and in consequence use of findings.

Most important documents which define system framework include:

- *Guidelines concerning evaluation of Operational Programmes 2007–2013*
- *Evaluation plan for National Strategic Reference Framework 2007–2013*
- *Evaluation system for the National Strategic Reference Framework 2007–2013 and the National Development Plan 2004–2006*
- *Integrated System for Managing Conclusions and Recommendations from the evaluation researches.*

General structure of the evaluation system is presented on the diagram bellow

Graph: Institutional structure of the evaluation system in POLAND



Source: Own analysis

The most important actors in Cohesion Policy evaluation system include: National Evaluation Unit, evaluation units located in managing authorities and implementing authorities and Steering Committee for the evaluation process of NDP and NSRF. Their efforts are fostered by the involvement of stakeholders of the evaluation process through the work of steering groups and in many cases with a significant role of monitoring committees. Responsibilities of the key actors of evaluation process are presented bellow.

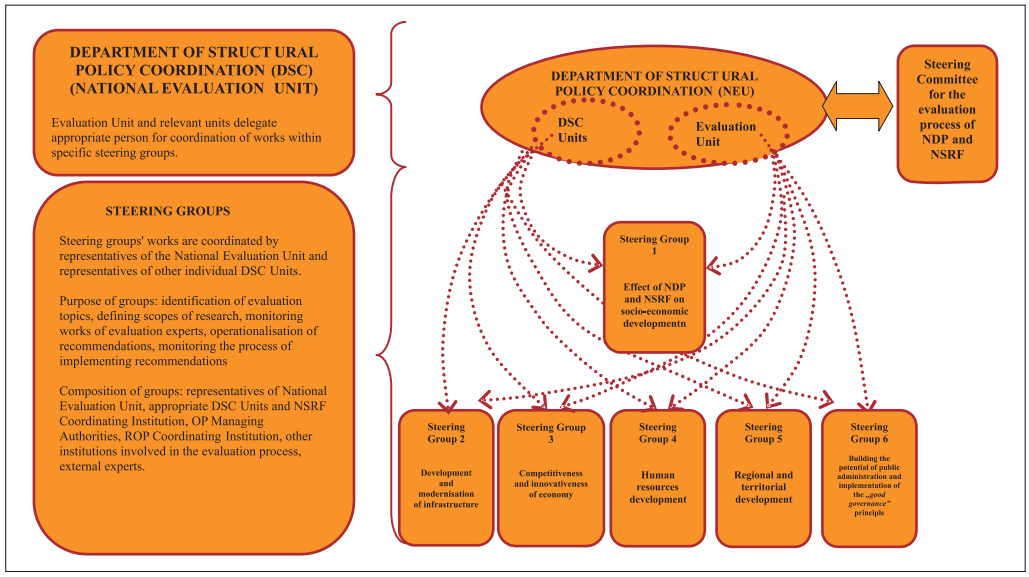
National Evaluation Unit

National Evaluation Unit (NEU) located within the Department of Structural Policy Coordination in the Ministry of Regional Development is responsible for the coordination of the whole system (work of other evaluation units). Furthermore NEU conducts horizontal and thematic evaluation researches on the level of National

Development Plan (NDP)/National Strategic Reference Framework (NSRF). Third responsibility covers coordination of the development of evaluation culture within the Polish administration.

Evaluation system on the level of National Development Plan and National Strategic Reference Framework is based in the institutional sphere in six thematic steering groups (picture below). Groups are comprised of representatives of the National Evaluation Unit, representatives of evaluation units from relevant operational programmes, representatives of relevant units from the Department of Structural Policy Coordination, representatives of Managing Authorities and other institutions involved in management and implementation (especially from monitoring and programming) of the relevant area of intervention.[9)]

Each steering group is dedicated to one sub-objective of National Strategic Reference Framework. At the beginning steering groups met regularly every 2–3 months, at present they are working more at ongoing basis but mostly through internet.



Source: Evaluation system for the National Development Plan 2004–2006 and the National Strategic Reference Framework 2007–2013.

Steering Committee for the evaluation process of NDP and NSRF

Steering Committee for the evaluation process of NDP and NSRF was established for coordinating purposes; it is a body operating at the level of directors of Managing Authorities. Its role is to ensure coherence of the evaluation process of NDP and NSRF as well as specifying general, strategic directions of the evaluation process. The important purpose of functioning of the Steering Committee for evaluation

[9] Evaluation system for the National Strategic Reference Framework 2004–2006 and the National Development Plan 2007–2013. NEU MRD 2007.

process of NDP and NSRF is to engage decision makers in the evaluation as they are the final and most important recipients of evaluation findings.

Evaluation units

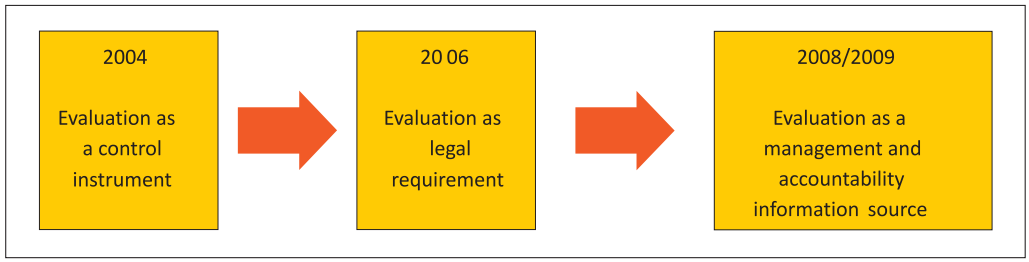
Following decentralisation of the implementation system of Cohesion Policy, also the evaluation process for 2007–2013 was decentralised. Evaluation units (apart from the National Evaluation Unit) are key actors in the evaluation system. Evaluation units are located within each operational programme (national and regional). They are responsible for organisation of the evaluation process within their respective Operational Programme.

Evaluation units were established within Managing Authorities or in some cases – to increase their independence and objectivity – outside Managing Authorities structures. Managing Authorities may delegate competence (and often do so) regarding evaluation to lower implementation levels (to intermediate bodies – IB). Steering groups were also appointed in most operational programmes; their task is to support evaluation units in implementing the evaluation process on appropriate implementation levels.

Evaluation Plans

In 2007 the obligation of preparation of Evaluation Plans has been introduced. According to the *Guidelines concerning evaluation of Operational Programs*, evaluation plans are prepared on the level of NDP/NSRF and on the level of operational programmes (as well as on lower implementation levels in case of delegating competence). There are two types of plans: strategic and operational. Strategic evaluation plans for 2007–2013 – embraces the whole programming period, they determine general areas of evaluation and the organisation of the system on a given implementation level. Operational evaluation plans indicate specific activities planned for launching in a given year: evaluation researches as well as other initiatives dedicated to development of evaluation culture (conferences, trainings, publications…). Evaluation Plans are published on websites, so potential evaluators have more time to prepare for future tender proceedings. The implementation of the above-mentioned Evaluation Plans is subject to monitoring carried out by the NEU. The annual information on evaluation forwarded to the National Evaluation Unit by the Managing Authorities until the end of March of every year, is the instrument for monitoring the evaluation process[10)].

[10] This system was firstly elaborated and launched by MA within Development of Human Resources Operational Programme and after successful implementation embraced whole evaluation system. Due to the cycle of conducting evaluation exercise, full implementation and final assessment on the effectiveness of the system could be made in 2011.



Source: S. Bienias (2009), Evaluation of Structural Instruments in Poland; Evaluation of EU Structural Funds: Reinforcing Quality and Utilisation, Ministry of Economics – Lithuania, 26-27 March, Vilnius

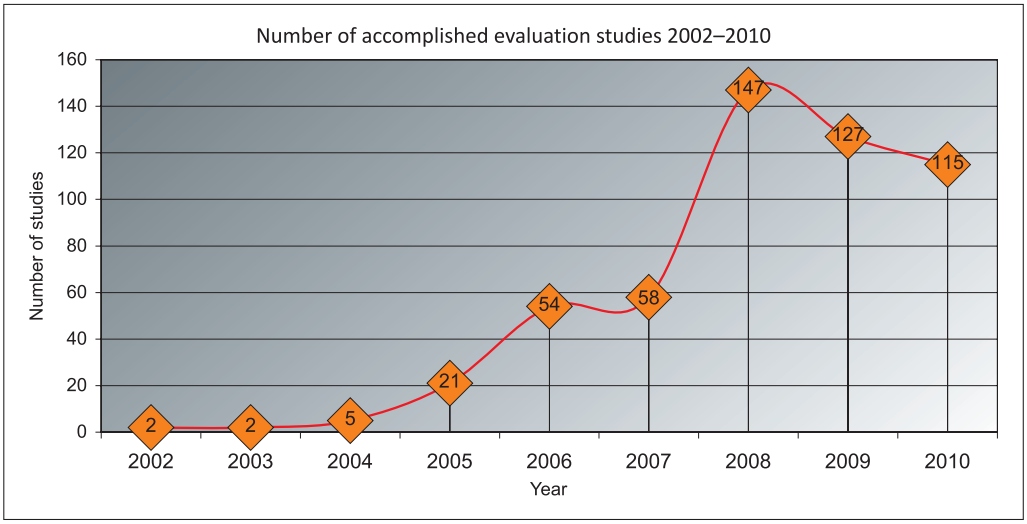
The above graph presents the manner in which the approach of people involved in the implementation of Cohesion Policy towards evaluation in Poland has changed. In 2004, evaluation was above all associated with control and raised serious concern of entities subjected to such controls. In 2006, along with the works on *ex-ante* evaluation (these were very intensive and decision makers participated in the whole process) it started to be perceived as a sort of formal requirement, obligation imposed by Brussels. Overcoming that distrust was a serious progress on the way to fully use this tool. Since 2008 a significant change in Poland can be observed. Evaluation has been used as important, even the most important source of information on actual effects of implementing of Cohesion Policy, and gradually, results of research works begun to be used in the process of operational management of EU programmes[11]. A constant shift towards more strategic oriented use of evaluation findings may now be observed. Significant role in this process has been played by *ex-post* evaluation (internally conducted metaevaluation based on a series of large, external and in-depth studies) which delivered truly strategic and at the same time "manageable" recommendations.

2.3. Structure of evaluation studies in Poland

The number of evaluation studies in 2002-2010[12] has grown significantly and reached 531 researches all together. This number does not include evaluations of single projects and embraces evaluations on different levels of implementation in both 2004–2006 and 2007–2013 Cohesion Policy programming perspectives. The graph below presents dynamics of the number of accomplished evaluation studies of Cohesion Policy conducted in Poland.

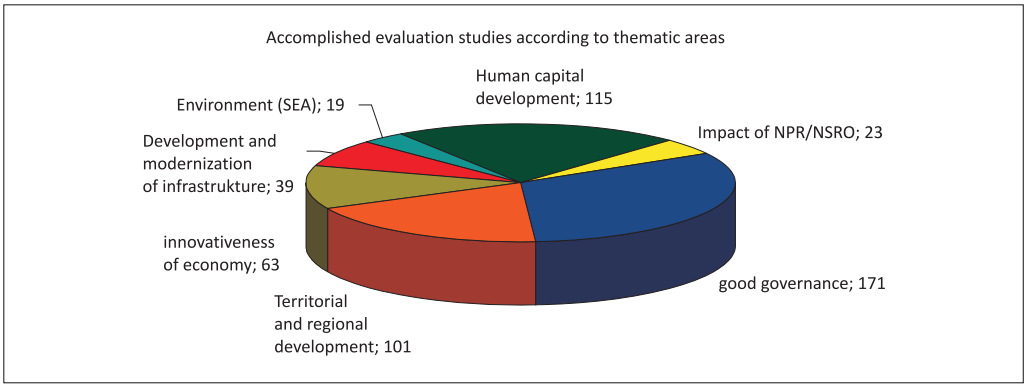
[11] Stanisław Bienias, Iwona Lewandowska (ed.) *Evaluation Systems In the Visegrad Member States, Ministry of Regional Development, Warsaw November 2009*

[12] Researches conducted in the years 2002–2003 were mostly *ex-ante* evaluations of the National Development Plan 2004–2006



Source: Database of evaluation studies as of 1 January 2011.

The change in the number of conducted evaluation researches may be observed especially between 2007 and 2008. This shift in increase of accomplished studies was strictly linked to the change in the system – responsibility for evaluation has been given to newly established managing authorities on regional level. At the same time evaluation units in the Ministry of Regional Development achieved their full capacity which at first resulted in the number of conducted evaluations. After the peak in 2008, the number of evaluation studies has stabilised around 120 researches a year. The challenge for all evaluation units is now quality and utility of evaluation findings so a significant change in future as far as the number of accomplished studies is concerned should not be expected.



Source: Database of evaluation studies as of 1 January 2011.

Taking the structure of conducted studies into consideration, "good governance" evaluations, mostly dedicated to analyse implementation issues of Cohesion Policy, have been the most popular type of evaluation among all those conducted in the reference period. Relatively highest number of researches (almost 1/3 of all conducted studies) in the area of good governance results above all from the initial stage of National Strategic Reference Framework implementation. Especially in 2008

and 2009, due to the lack of measurable effects of the intervention managing authorities concentrated their efforts on evaluation of system implementation issues. A relatively high number of researches concerns areas of territorial and regional development and human capital. Human capital seems also to be the best evaluated (in terms of the effects of intervention) area of Cohesion Policy support

***Ex-post* evaluation of National Development Plan for 2004–2006 in Poland**

National Evaluation Unit decided to conduct full *ex-post* evaluation of the first programming perspective of Cohesion Policy implementation in Poland. Evaluation took form of a series of thematic studies divided into three main components:

- Horizontal (macroeconomic impact assessment, impact of Cohesion Policy implementation in Poland on the economies of EU-15 countries, impact on the level and quality of employment, complementarities of interventions and a system of implementation),
- Sectoral (development of transport infrastructure, quality of environment, competitiveness of enterprises and human capital),
- Territorial (development of cities, development of Eastern Poland, strategic management on regional level and effects of cross-border programmes).

All those studies were contracted to external evaluators and conducted in 2009–2010. On this basis internal metaevaluation has been carried out by the National Evaluation Unit. The draft of the synthesis was discussed during the VI Evaluation Conference in December 2010 and then supplemented by main conclusions and recommendations from the discussion.

Additionally National Evaluation Unit decided to prepare a short document (10 pages) for the use of decision makers, namely Conclusions from the management of Cohesion Policy in the light of conducted evaluations and analysis. This document was a subject of two in-depth discussions with all directors of Managing Authorities. After it was submitted to the Coordination Committee of NSRF and then send to subcommittee of Prime Minister Council for Coordination of Development Policy.

This document is supposed to be one of the most important sources of information for preparing the strategic documents of the next programming perspective (the conclusions have been already used when shaping the Polish position for the future of Cohesion Policy). It is also assumed that conclusions and recommendations will be used in shaping the national strategic documents (major part of evaluation findings has already been used in the National Strategy for Regional Development 2010-2020 and in the Concept of Spatial Development of the Country 2030).

For more information see: http://www.ewaluacja.gov.pl/Ewaluacja_ex_post_NPR/

Key conclusions following from Cohesion Policy evaluation process

The lessons learned during Cohesion Policy evaluation process make it possible to formulate several basic conclusions, which constitute grounds for the further development of Cohesion Policy evaluation system, as well as the evaluation system for other national policies:

- The most efficient method of capacity building for the purposes of evaluation is the “learning by doing” method (mistakes and failures are an inher-

ent element of this process, and the fastest method to learn how to avoid them successfully is to do so in practice)

- The evaluation aims at providing relevant information at the right time to the right person (key issue is how to translate hundreds of pages of analytical materials generated under each study, into several brief, precise and most important conclusions and recommendations accurately targeted at the right time during the decision making process)
- Decentralisation makes it possible for evaluation to provide information concerning the actual problems (it is significant for the entire process to be focused, as close as it is possible, on the actual problems)
- Cooperation and partnership: the most relevant evaluation questions should be asked by the people, who are directly engaged in evaluation.
- Evaluation does not cover single studies – knowledge should be obtained from the entire evaluation process (significant changes at the strategic level may be made in a responsible manner only on the basis of analyses carried out on a greater number of studies).
- Evaluation is not the objective itself – it is a tool aimed at optimum achievement of strategic objectives
- Decision-makers must be involved in the evaluation process at each stage of works, as they are the future, major recipients of the research results

3. National public policies evaluation system in Poland

At the moment, Polish public authorities do not have at their disposal a coherent analytical system, which would provide a methodological support to the decision processes at all management levels. There is a need to construct a system, which would collect, store and analyse the data in order to generate and provide information of diagnostic and evaluation nature.

A starting point to design a public policies evaluation system should be system solutions operating in Poland within the frameworks of implementing Cohesion Policy, which were described in Chapter 2. The system covering the entire public policies in Poland should become an extension of the already applied solutions. A the same time, it must be flexible and adapt the general rules applied in the already operating evaluation system to the conditions and capacity of Polish administration.

Within the national administration, the system should be more centralised (especially at the very beginning) when there is sufficient capacity. Low administration capacity may result in low quality of studies and decision makers (not fostered by the European Commission – as in EU funds) may be reluctant to take the risk. In time, as experience is gained, the move towards more decentralised evaluation system should be taken. This decentralisation should be necessarily followed by delegating competences concerning the formulation of themes and scope of evaluation studies

in order to ensure the possibility of providing flexible responses to the demand for knowledge and experience which develops in the course of implementation.

Especially at the very beginning, public policies evaluation process should be managed and closely coordinated in the strategic dimension, as well as in the dimension of general principles and standards. Therefore, a single centre should be appointed that would have the relevant factual and administrative capacity. The role of the centre would cover the process of evaluation as well as the process of coordination in order to ensure cohesion of the entire system and usability of evaluation results. **Evaluation process coordination centre should be situated in the proximity of the entity responsible for development policy coordination (Ministry of Regional Development) or it should operate directly under the Coordinating Committee for Development Policy**. However, the first solution (when joining the work of this centre with the scope of competences of National Evaluation Unit) will allow to use EU resources and broad evaluation experience to launch the whole process.

The existence of formal and legal regulations on the implementation of evaluation process preconditions launching the public policies evaluation system, especially at the initial stage and under the conditions of insufficiently developed evaluation culture. These regulations do not have to be (and even should not be) detailed – they should only constitute an obligation to establish the necessary elements of the system. The recently adopted National Strategy for Regional Development 2010–2020 and subsequent legal changes will create suitable conditions for such changes.

The establishment of a knowledge-based administration that is targeted at objectives and effects, under which one of the elements is public policies evaluation, requires solutions independent from evaluation, but preconditioning its efficiency and effectiveness. It especially concerns the principles and standards referring to the process of programming and implementing public policies, strategies and programmes. It is impossible to evaluate policies, which do not have clearly defined strategies and quantified strategic and operational objectives (performance indicators system). The first steps to develop principles and standards of strategic programming were made by the Ministry of Regional Development (the Act on the principles of development policy and the project implemented by the MRD from the EU resources, which concerns strategic programming methods and the use of economic and financial indicators). Nonetheless, it is necessary to intensify the works in this regard in the entire administration. The functionality of the evaluation system depends also on the existence of public policies monitoring system and adjustments of the scope and manner of data collection carried out by the Central Statistical Office to satisfy the needs of the administration.

A significant issue in this regard covers completion of works on the activity-based budget. Its operation has the possibility to become a breakthrough in the Polish administration due to reorientation of activities to objectives and results. At that time evaluation will become a necessary and natural element of the public adminis-

tration system. Ensuring a sustainable source of financing, the evaluation process is prerequisite for efficient implementation of evaluation activities. Including the costs of evaluation activities as a fixed item in the budget should be a logical consequence of setting out an activity-based budget.

The need to safeguard independence and objectivity of the public policies evaluation process is one of the major and generally accepted European and global evaluation standards. Public policies evaluation in Poland cannot become a tool to justify the decisions that have already been taken, therefore, one should seek to ensure the greatest possible degree of independence of evaluation from policy implementation processes. Although the evaluation results are often critical of the effects of public authorities' actions, it should be remembered that openness to constructive criticism and discussion shows the strength and efficiency of a State. The best methods to ensure independence of the public policies evaluation process is the **adequate institutional separation of the evaluation system** in the State administration, e.g. through locating the evaluation units at relevant Ministries and through introducing **the external evaluation principle**, which is currently being successfully implemented within the framework of Cohesion Policy in Poland, as well as throughout the entire European Union. External evaluation consists in commissioning evaluation studies to external experts by the evaluation units after a prior identification of information needs and their formulation in the form of the scope of the study. Commissioning studies to external evaluators is associated with the need to comply with the public procurement law, which at the moment is not fully adapted to these types of contracts and has to be amended (for more details see Chapter 4 *Key challenges*).

The introduction of the above-mentioned principle requires from the public administration to be more open, to cooperate with public and private external entities in order to use their knowledge, practice and experience, including in particular the establishment of a forum for cooperation with the world of science, for example, through institutionalisation of such a cooperation within the framework of Steering Groups described below.

Within the framework of the projected public policies evaluation system, there should also be a place for Impact Assessment. The potential of this useful institution, which is successfully applied in many other countries, is not presently used to its full. It seems that **incorporating the Impact Assessment into the public policies evaluation system** would lead to a greater use of assessment, which is now treated only as a formal requirement rather than a real and effective tool in the management of public policies.

A communication system for notifying the effects of public policies implemented by the authorities is a necessary element supplementing the evaluation system. The public **presentations of the results of the evaluation studies** organised with the participation of representatives of the administration, the world of science,

socio-economic partners, as well as media representatives, which are successfully used within the framework of Cohesion Policy evaluation system constitute a useful tool of disseminating and discussing the results of the implemented public policies (apart from electronic and traditional publications) These presentations provide a useful discussion forum on the efficiency and effectiveness of actions undertaken by public authorities.

Institutional solutions

The centre coordinating the entire process and the National Evaluation Unit should become the major and the central institutions under the public policies evaluation system. The role of the NEU, apart from its coordinating tasks, should cover also factual supervision over evaluation, including, *inter alia*, formulating the general guidelines and standards. The coordination should concern in particular the strategic dimension of the evaluation process, *inter alia*, by initiating and monitoring horizontal evaluation studies. The tasks of the NEU should also include providing support to the establishment of the evaluation capacity of the Polish administration.

The **evaluation units**, which are situated in the ministries and entities competent for implementation of individual national policies, should become the core of the public policies evaluation system. Their major task would cover substantive support to the programming and implementation process through providing results of the evaluation studies referring to the issues which are the object of a given policy. The evaluation units should function at different implementation levels depending on the manner of policies institutionalisation. In the case of self-governments, the evaluation system may be based on the presently existing solutions operating within the framework of Cohesion Policy, including the evaluation units situated in the Offices of the Marshal.

The **evaluation Steering Groups** form the second pillar of the evaluation system. Their tasks cover monitoring and factual supervision over the research process. The Steering Groups should be comprised of representatives of institutions involved in the implementation of the policies, which are interested in the results of the evaluation and the experts from the areas that are the object of the study. The Steering Groups will make it possible to ensure high factual quality of the evaluation process and especially high usability level of the formulated conclusions and recommendations. These groups are a link between the evaluation process (managed by evaluation units) and the public policies implementation system (managed by respective ministries or self-governments). The system of Steering Groups forms an element of a modern administration model, under which work does not boil down exclusively to thinking and acting in office or department categories, but it is based on teamwork in interdisciplinary project groups.

Under such public policies evaluation system, the Council of Ministers would become the major and final addressee of the most important and strategic conclusions and recommendations formulated in the course of evaluation process, as it provides the platform for taking decisions prepared on the basis of information generated, *inter alia*, by the public policies evaluation process. As a result, it is necessary to **establish a supporting and consultancy body, which would steer the evaluation process, including, *inter alia*, the analysis of the evaluation of individual policies and submitting to the Council of Ministers the proposals for operationalisation of the formulated conclusions and recommendations**. Such an entity (Sub-committee, Council, Team, etc.) should be situated under the Coordinating Committee for Development Policy or directly under the Council of Ministers.

4. Key challenges

Focusing activities on a few key areas of the system preconditions smooth operation of the public policies evaluation in Poland.

Using the experience of Cohesion Policy and the institutional memory

The public policies evaluation system in Poland should not be constructed "from scratch" and separated from the Cohesion Policy evaluation system. Failure to use the resources of the institutional memory would be a repetition of the mistakes made at the moment of launching the implementation of Structural Funds. The knowledge and experience of the administration obtained during the PHARE programmes implementation was not used to the optimum then. Therefore, the evaluation process and system should be skilfully and systematically "extended" to cover other national policies. Even more so, as the Polish Cohesion Policy evaluation system is commonly recognised (also by the European Commission) as one of the best systems in Europe. Note also that Cohesion Policy is one of the elements of the development policy and over time the EU Structural Funds will be replaced by the national State resources. At the moment, activities are initiated that aim at transferring the experience and solutions within the scope of evaluation to the implementation systems of other national policies. These, above all, cover legislative activities in the form of the Act on the principles of development policy, which covers provisions imposing the obligation to carry out the *ex-ante* evaluation for the programmes and development strategies, as well as attempts to use the results of Cohesion Policy evaluation to form recommendations for other national policies (*see box on Ex-post evaluation of National Development Plan 2004–2006*).

Building evaluation capacity in the administration

The experience gained during the implementation of Cohesion Policy evaluation process point to the need for the employees dealing in evaluation activities to have high qualifications. The implementation of the evaluation process does not consist

entirely in the simple management, but requires knowledge and skills both within the scope of research project management, as well as factual knowledge from the fields covered by the evaluation studies and methodological knowledge. As a result, one of the major challenges related to the establishment of public policies evaluation system is building the evaluation capacity in the Polish administration. It requires undertaking a number of training and information activities carried out in a continuous and systematic manner. Cooperation with the academic circles is also necessary (Academy of Evaluation launched by the Ministry of Regional Development in cooperation with Warsaw University may serve as a good practical example here(Euroreg)).

Adjusting the provisions of the Public Procurement Law to the contracts concerning analytical and evaluation studies

The possibility and ability to use the external experts' knowledge (see Chapter 1) preconditions smooth operation of the modern administration. The manner and procedure of tendering experts' services, including evaluation studies, is governed by the Public Procurement Law. Nonetheless, the experience related to the implementation of the evaluation studies carried out under Cohesion Policy point to the fact that the regulations are inappropriate to these type of innovative services provided for the administration. If we aim at building a modern knowledge-based administration, the provisions of the Public Procurement Law should be modified in order to provide for the possibility of ensuring high quality and usability services of experts, including evaluation studies. Introducing a separate procedure for this type of contracts could be a solution to the problem, introducing the possibility of applying the selection criteria which are to a greater extent based on the quality assessment than on price, as well as the possibility of applying criteria that concern the experience of the research teams.

Closer cooperation with the world of science and development of the methodology of evaluation studies

The use of knowledge and experience of external entities in the administration, including in particular scientific units, requires ordering and institutionalising the cooperation between both these parties. Cooperation with the representatives of the world of science may be useful at all stages of evaluation process, i.e. study problems identification and formulation, preparation of the scope of the study, selection and development of the methodology and the use of the results of evaluation studies. Such cooperation enables knowledge and technology transfer from the world of science to administration and, in line with the feedback principle, it brings science closer to the practical issues, and, thereby, enables verification and development of socio-economic theories.

Building the monitoring system

The existence of the data collection, storage and processing system, which enables constant monitoring of progress in the implementation of the planned objectives, is a prerequisite for successful construction of administration targeted at objectives and effects, under which evaluation becomes one of the management tools. Implementation monitoring is a fixed element of implementing projects and programmes financed under the Structural Funds and the experience in this regard should be used in the national policies implementation process. It is especially important to establish a coherent system of performance indicators (especially in the context of work on activity-based budget). It is also important to encourage greater involvement of the Central Statistical Office in the national policies monitoring process (including adapting the scope and manner of data collection carried out by CSO to the needs of the administration).

There is a need to create legal regulations concerning evaluation

Apart from activities of informational and motivational nature, which are targeted at building evaluation capacity in Polish administration, it is also necessary to create formal requirements concerning the performance of evaluation activities (especially, at the initial stage of evaluation system establishment). It does not relate to detailed regulations (that may lead to dysfunctionality of the system; its main advantage should be flexibility and ability to respond quickly) or to the general legal provisions concerning the obligation to carry out evaluation activities. First steps in this direction were made in the provisions of the Act on the principles of development policy, which regulates the issues associated with the need to implement *ex-ante* evaluation for the development programmes and strategies[13)]. The recently adopted National Strategy for Regional Development 2010-2020 as well as forthcoming Polish Presidency creates a friendly environment to achieve this strategic objective.

Ensuring permanent financing sources for the evaluation process

One of the basic conditions for ensuring stability of the evaluation system is to establish permanent financing sources for the evaluation activities, including especially the evaluation studies. At the moment of setting up the activity-based budget, the evaluation activities and thereby their financing sources will become a natural element of the State budget. Nonetheless, even now it is necessary to create appropriate financing sources of evaluation activities in parallel with the process of building the system.

[13] Act of 6 December 2006 on the principles of the development policy (Dz. U. 06, No. 227, item 1658, as amended)

Setting up the system for communicating the effects of public policies

The process of communicating the objectives of the authorities and effects of their performance should run in a systematic and institutionalised manner. At the moment, the process is *ad hoc* and inconsistent. The publicly available results of evaluation studies should be the basic instrument to communicate the results of public policies. Cooperation with the media is also necessary in this respect.

* * *

Efficient use of the capacity, which is inherent in evaluation, by the Polish administration is very strongly determined by the institutional and legal context. If the philosophy of operation of the administration and the manner of implementing its tasks does not undergo fundamental changes, it will never be possible to optimally use the public resources. Under such circumstances even the best developed and managed evaluation system will not operate fully efficiently.

The three challenges, which seem of key significance in this regard are presented below:

Targeting the administration at implementing objectives and not at carrying out procedures

The Polish administration operates in the "areas of competences" and procedures and it is not targeted at achieving specific objectives. This approach (together with the absence of any bonuses stimulating to undertake risk) causes:

- multiplication of unnecessary procedures (which supposedly should enhance the safety of public resources expenditure – in practice they only dilute responsibility),
- losing the major external objective of operation (the officials often act for their own benefit, and their work only slightly contributes to the actual objectives of the institutions that employ them).

As a result, the entire administration becomes too fixed and "resistant" to the changing external challenges, as well as unwilling to seek any pro-efficiency solutions. Moreover, avoiding irregularities seems to become the key motivation.

Therefore, it is necessary to introduce the principle of formulating quantified, coherent and hierarchical system of objectives (strategic and operational) for individual public administration units together with the programmes for their implementation. It is also necessary to ground the interim evaluation of public administration officials on the criteria of efficiency and effectiveness in obtaining specific objectives (quantified in line with the relevant performance indicators) that were formulated for individual employees and directly contribute to the implementation of the objectives of respective units. Thereby, the principle of personal responsibility for the

objectives and principles of awarding bonuses for their achievement – success fee, was introduced

However, the implementation of the principle requires a fundamental reform in the manner of spending public resources and introduction of an performance-based budget.

Introduction of the performance-based budget

At the moment, public authorities base their operation and the implementation of most activities on annual budgets. The budget implementation consists in allocation of relevant amounts to the established areas and paragraphs, and efficient implementation is identified with full use of resources. However, the actual area and the effects that are brought by the interventions are completely disregarded. This solution strengthens the above-mentioned model of administration targeted at procedures, under which the main motive of action is passivity and, hence, the desire to avoid responsibility.

Therefore, it is necessary to introduce, as soon as possible, the performance-based budget, which will have clearly, precisely and correctly quantified objectives. It is also necessary to resign from the current system solutions. A key element here is the appropriate selection and quantification of performance indicators that will enable the subsequent settlement of the scheduled tasks. It is important to precondition obtaining of public resources with the efficient and effective implementation of previous tasks. The system of indicators should, therefore, be simple (at the level of products and results) and it should be clearly defined leaving no room for interpretation.

However, in order to introduce such an approach, it is necessary to previously define precise strategic and operational objectives for individual institutions and areas of State action.

Appropriate formulation of public policy objectives

Lack of coherent strategic vision, absence of coordination of public interventions and the low quality of programming process are among the most significant weaknesses of the implementation of activities by Polish public authorities. At the moment, the strategies at all levels (from the local level to the national level) are to legitimate the expenditure of public resources allocated to all socially acceptable objectives and they are not created with the view to precisely concentrate the available resources. Most of the up-to-date strategies are not used to prioritise the actions but are rather focused on ensuring that the financial resources are obtained. This approach, given the level of development of the Polish economy, generates enormous inefficiency of the already limited public funds.

Therefore, it is necessary to develop, as fast as possible, mutually and internally coherent strategic visions in respective areas. This vision should clearly and very precisely define and prioritise strategic objectives which will be based on reliable diagnosis[14)].

* * *

The above-mentioned three major challenges which have to be faced by the Polish administration are closely related and determine each other. In order to establish an administration focused on results (instead of procedures) it is necessary to implement instruments which are favourable for the purpose, including, above all, performance-based budget. Creation of a performance-based budget is possible and makes sense only if the public authorities have the ability to precisely define the objectives of its public policies. Correct formulation of the objectives of the public policies, however, is possible only if the administration is targeted at objectives and effects. On this basis a conclusion may be formulated that the process of improving the operation of the Polish administration should be initiated under all key areas mentioned above. Only this approach gives a possibility to leave the above-described "vicious circle" of mutual conditionings and enables to establish a modern, efficient and effective public administration in Poland, which will form the fundament of an good-governed and efficient State.

[14] One should hope that currently prepared nine integrated strategies will change this situation.

Bibliography:

Bienias, S. Żuber P.(2008), *System ewaluacji w Polsce – dotychczasowe doświadczenia i wyzwania na przyszłość [Evaluation system in Poland – current experiences and challenges for the future]*; in: Kozak M., Ledzion B., Olejniczak K., (ed.) *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki [Evaluation of public interventions in theory and practice. Academic handbook]*, Warsaw, Wydawnictwa Akademickie i Profesjonalne.

Bienias S., Sudak S. (ed.) (2008), *Proces ewaluacji polityki spójności. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość [Cohesion Policy evaluation process. Summary of the current experience. Plans and challenges for the future],* Warsaw, Ministry of Regional Development.

Bienias S., Lewandowska I. (ed.) (in publication), *Evaluation of cohesion Policy in Poland, in Evaluation of cohesion Policy in Visegrad Countries*, Warsaw, Ministry of Regional Development.

Ex-post Evaluation of Cohesion Policy Programmes 2000-06 Co-financed by ERDF. Work Package 11 – Mini-case study Poland (2009), Brussels, European Policies Research Centre,

Ferry M., Olejniczak K. (2008), *The use of evaluation in the management of EU programmes in Poland*, Warsaw, Ernst & Young – Program "Sprawne Państwo" ["Efficient State" Programme].

Kierzkowski T. (2002), *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej [Assessment (evaluation) of programmes and projects of socio-economic nature at the background of Polish accession to the European Union]*, Warsaw, Polish Agency for Enterprise Development.

Mathison S. (2005), *Encyclopaedia of evaluation*, Thousand Oaks, Calif., London, SAGE.

Olejniczak K. (2007), *Ewaluacja jako narzędzie zarządzania w sektorze publicznym [Evaluation as a management tool in the public sector]*, in: Gorzelak G., Tucholska A. (ed.) *Rozwój, region, przestrzeń [Development, region, space]*, Warsaw, Ministry of Regional Development.

Pawson R. (2006), *Evidence-based Policy. A Realistic Perspective*, Thousand Oaks, Calif., London, SAGE.

Act of 20 April 2004 on the National Development Plan (Dz. U. No. 116, item 1206, of 005 No. 90, item 759 and No. 267, item 2251 and 2006 No. 149, item 1074).

Act of 6 December 2006 on the principles of the development policy (Dz. U. 06, No. 227, item 1658, as amended).

National Regional Development Strategy 2010–2020, MRD, 2010

THE EVALUATION OF REGIONAL POLICY IN EUROPE: CULTURE, COMMITMENT AND CAPACITY

John Bachtler

1. Introduction

Evaluation has a long history in Europe. As in North America, the growth of public expenditure in the 1950s and 1960s, associated with increasing government intervention through social and economic programmes, prompted the need for research on how best to allocate public funds in support of government objectives (what works?). With greater constraints on government spending during the 1970s and 1980s, evaluation was increasingly used for accountability purposes (how well is it working?) to justify the continuation or rationalisation of programmes (Derlien and Rist, 2002). Latterly, the complexity of policymaking has added further impetus with a proliferation of audit, assessment and evaluation activities intended to improve the design, governance and legitimacy of policies and processes (Nilsson *et al*, 2008).

Much of the evaluation of regional policy in Europe dates from this period, most notably in the United Kingdom – where there was extensive research on the effectiveness and efficiency of regional initiatives – as well as in Germany and the Netherlands (Nicol, 1982; Wadley,1986). An important stimulus to evaluation activity in the field of regional policy in the 1990s came from the European Union through the regulations of the Structural Funds, which required EU Member States and the European Commission to evaluate the outputs, results and impacts of regional development programmes at different stages of the spending cycle. These EU regulatory requirements have had certain spillover effects on the evaluation of domestic regional policy, which has become much more common in Europe over the past 10–15 years, not least in those Member States which acceded to the EU in 1995, 2004 and 2007 (Bachtler and Michie, 1995; Bachtler and Wren, 2006; Bachtler *et al*, 2009).

Notwithstanding this recent increase in evaluation activity, questions remain about the nature of the evaluation culture in Europe. Although there have been advances in the professionalisation of the discipline – reflected in the creation of evaluation societies and new evaluation journals – the degree to which evaluation is institutionalised within governments across Europe varies. There is a significant gap between the debates taking place on evaluation theory and methodology, on the one hand, and the requirements of administrative authorities commissioning

evaluations, on the other. There is also limited evidence on how much influence the increased evaluation activity is actually having on policy decisions.

In this context, the following chapter examines the evaluation of regional policy in Europe based on an extensive review of the literature as well as surveys of Member State domestic regional policy and EU Cohesion policy evaluation practices (Polverari *et al*, 2006; Polverari and Bachtler, 2004; Raines and Taylor, 2002). The chapter begins with a brief discussion of the evaluation culture in Europe, drawing on the comparative work of Furubo *et al* (2002). It then focuses on the evaluation of regional policy in different groups of countries – those where regional policy is long established, the 'second wave' of countries where regional policy evaluation developed in the 1990s, and the emerging evaluation practices in the newer Member States of Central and Eastern Europe. The chapter concludes with some observations on the future of regional policy evaluation.

2. The evaluation culture in Europe

Over the past two decades, the evaluation of public policy has become increasingly important across Europe. This reflects several factors. Most important are the pressures on public expenditure, and the need for policymakers to demonstrate that governments are spending money wisely and well. The emphasis placed on evaluation is also part of the trend in 'new public management' during the 1980s and 1990s which involves a culture that sees the citizen as client or customer, promotes organisational choices in public service provision (involving both public and private sectors) and requires greater accountability. A further factor is the growing complexity of policy challenges, requiring policymakers to have access to relevant, accurate and timely information to assist with the development of policy objectives, the setting of priorities and the design of policy instruments. Regionalisation and thicker institutional landscapes, with more actors involved in the design and implementation of policy, and interacting through different types of relationships, partnerships and networks at national, regional, local and community levels, have also promoted a demand for more evaluation by stakeholders within and outside government.

Some of the emphasis on evaluation is long-standing. In Germany, for example, all public intervention measures have been subject to regular 'success controls' (*Erfolgskontrollen*) since 1970. In the Netherlands and Sweden, formal government and parliamentary reviews of policy have long been an established part of the policymaking culture. Policy evaluation has also been undertaken in the UK for more than 30 years. Recent years have seen a reinforcement of this attention (Polverari and Bachtler, 2004). For example, in the Netherlands, a new Ministry of Finance Regulation on Policy Implementation and Evaluation was introduced in 2002 (*Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid*) to ensure that the evaluation function was being satisfactorily carried out within government and

to make certain that, within the framework of departmental budget estimates and accounts, quality information was being supplied. In the UK, HM Treasury produced a Green Book 'Appraisal and Evaluation in Central Government' in 2003 providing binding guidance to government departments and executive agencies stating that "all new policies, programmes and projects, whether revenue, capital or regulatory, should be subject to comprehensive but proportionate assessment, wherever it is practicable so as best to promote public interest" (HMT, 2003, 1).

Since the 1990s, many other European countries – Austria, Denmark, Finland, Ireland, Italy, Poland, and Spain – have also given a higher profile to evaluation. This 'growing market' (Leuuw, 2004) is evidenced by a certain institutionalisation of evaluation within government departments – at least in some policy domains – through the establishment of evaluation units, the regular commissioning of evaluation studies as part of policy cycles and the publication of evaluation reports. In some countries, such as Denmark, Finland and Ireland, this institutionalisation process has been extremely rapid over the past decade.

The importance accorded to evaluation is part of wider developments in policy design and delivery, which have both 'information' and 'allocation' functions for policymakers (Derlien and Rist, 2002), such as:

- the demand for 'evidence-based policymaking' to make sense of complex policy challenges, seen as part of 'good governance' in responding to economic challenges (OECD, 2007a);
- the use of foresight techniques for strategic thinking and policy planning in conditions of rising uncertainty (e.g. in the field of science and technology, regional development) to analyse future challenges, from a multi-disciplinary perspective, and to identify policy options and strategies (Gavigan and Scapolo, 2002a, 2002b);
- the application of performance budgeting – developing indicators, benchmarks and targets as part of the budget-setting process, to evaluate performance and to demonstrate value for money to the public (OECD, 2007b);
- benchmarking – comparing performance across territories, sectors or policy domains to assess relative performance and to assess the scope for a better use of resources; and
- results-based management – the management of organisational resources and the application of processes to achieve relevant and measurable results (OECD, 2001).

The evaluation discipline has also developed with the creation of evaluation societies – in the United Kingdom (1994), Switzerland (1996), Germany (1997), Italy (1997), Finland (1999), France (1999), Denmark (2000), Poland (2000), Spain (2001), Sweden (2003), Hungary (2006), Slovakia (2006), the Czech Republic (2007), the Netherlands (2007), Romania (2008) – and the launch of new evaluation journals,

internet-based discussion groups and networks, the drafting of evaluation standards ands guidelines, and the organisation of regular evaluation conferences. As the International Organisation for Cooperation on Evaluation has noted, this is part of a global trend: "in the 1980s there were only three national and regional evaluation societies, by the late 1990s there were nine and by the beginning of the 21st century the number had grown to fifty[1]. "

An important stimulus to evaluation in Europe (at least in some countries) has been European integration and the role of the EU, particularly through policies managed under the Community method involving direct expenditure (development aid, research, TENS) and indirect expenditure (CAP, Cohesion policy). The 'shared management' of EU funding through multi-level governance is said to have spurred evaluation in similar ways to the experience of the United States (Stame, 2008). Stame argues that, in the US "'shared responsibilities' are at the origin of the spread of programme evaluation in the 1960s when the federal programmes…had to prove their effectiveness in face of a traditional aversion [by the states] to federal intervention into what had up to then been considered state affairs" (120). She considers that the development of federalism in the EU is linked to "federalist mechanisms as 'subsidiarity' and 'additionality'" (134).

It is certainly true that the use of evaluation within the Commission services, and under certain Community policies, has grown since the 1980s. Williams *et al* (2002) identified "three specific families of evaluation practice" in EU policies: the use of project cycle management and logical framework approach within overseas development projects and other external policies (e.g. PHARE, TACIS); evaluation and monitoring systems with independent expert panels and peer review methodologies in research policy; and a methodological framework for evaluating regional development and employment interventions under EU Cohesion policy. Such initiatives have been strengthened by more recent EU reforms to promote 'sound and efficient management' which have encouraged the use of evaluation as part of effective programme design and implementation. Since 2000, the Commission services have been required to carry our evaluation for activities directed at addressees or beneficiaries outside the European institutions, obligations which were reinforced in 2007 with a DG BUDGET evaluation action plan (European Commission, 2000; 2007). Other external international pressures on using evaluation to improve policy design and performance have come from the OECD, UNDP and World Bank.

The EU influence is acknowledged to have been particularly strong in Ireland and Spain, both countries having received substantial funding from EU Structural and Cohesion Funds (Derlien and Wrist, 2002). Indeed, in Ireland, O'Hara *et al* (2007) consider that, without the external pressure from the EU (and also the OECD), an evaluation culture would not have emerged. In Spain, a review of the development of evaluation by Fernández-Ramirez and Rebolloso (2006, 137) claimed that "the

[1] http://ioce.net/overview/general.shtml

directives of the European Union have been the basic motor" for the rapid increase in evaluation practice since the 1990s, a conclusion shared by Viñas (2009) who highlighted the powerful impetus of Structural Funds evaluation obligations. The same can be said about the development of evaluation in the newer Member States of Central and Eastern Europe.

Notwithstanding these developments in European evaluation culture, some researchers argue that, compared to North America, the discipline is still in its 'infancy', with inadequate training capacity for evaluators at university level, the use of 'quasi-professional' evaluation teams for multinational studies, a prevalence of top-down evaluation processes, and an excessive focus by evaluation studies on resources and administrative practices rather than outcomes and impacts (Leeuw, 2004; Schröter, 2004).

Indeed, putting EU trends in an international context is salutary. Table 1 shows how Furubo *et al* (2002) ranked countries using indicators of an evaluative culture, taking account of the degree to which evaluation takes place in different policy domains, the supply of evaluators, national evaluation discourses, and the professionalisation and institutionalisation of evaluation. In this ranking, the benchmark is set by evaluation culture in North America; despite progress in the 1990s, (Western) European countries were generally seen as being some way behind. Only Sweden, the Netherlands and the United Kingdom came close to the scores of the US and Canada, and several countries – Italy, Ireland and Spain – were given very low scores. Many countries were ranked poorly on the degree of institutionalisation of evaluation (within government or parliament), the limited proportion of outcome evaluations undertaken, and the (under)use of evaluation by national audit offices. This analysis was undertaken at the start of the 2000s, but other rankings undertaken since then also deem much the same group of countries – the United Kingdom, Australia, Canada, the United States, France, Norway, the Netherlands and Sweden – to be leading in terms of the institutionalisation of evaluation and the state of the discipline (Jacob and Varone, 2003; Varone, 2007).

Table 1 also illustrates the significant variation within Western Europe (Central and Eastern European countries were not included in the survey). It has been argued that differences in evaluation culture are related to systems of law and administration (Bœuf, 2001; Casavola and Tagle, 2003). For example, the so-called 'Continental' administrative structure and law systems in France, Italy, Portugal and Spain are deemed as having been less favourable to the emergence of an evaluation culture. By contrast, Anglo-Saxon common law and public management-based administration are considered to have been more conducive to the development of such a culture, with the emphasis that it places on value for money and accountability.

Table 1: Ranking of countries on indicators of an evaluative culture

	Indicators of an evaluative culture									
	I	II	III	IV	V	VI	VII	VIII	IX	Total
United States	2	2	2	2	2	2	2	2	2	**18**
Canada	2	2	2	2	2	1	2	2	2	**17**
Sweden	2	2	2	1	2	1	2	2	2	**16**
Netherlands	2	2	2	1	2	1	2	2	1	**15**
United Kingdom	2	2	2	2	1	1	2	1	2	**15**
Germany	2	2	1	2	1	1	2	1	1	**13**
Denmark	2	2	2	1	1	0	2	1	1	**12**
France	2	1	1	2	2	1	1	1	0	**11**
Finland	2	1	1	1	1	1	1	1	1	**10**
Switzerland	1	1	2	2	0	0	2	0	0	**8**
Ireland	1	1	1	0	1	0	1	1	1	**7**
Italy	1	1	1	2	0	0	1	1	0	**7**
Spain	1	0	1	2	1	0	0	0	0	**5**

I) Evaluation takes place in many policy domains. II) Supply of evaluators in different disciplines. III) ational discourse on evaluation. IV) Professional organisations. V) Degree of institutionalisation – government. VI) Degree of institutionalisation – parliament. VII) Pluralism of institutions or evaluators performing evaluations in each policy domain. VIII) Evaluation within the Supreme Audit Institution. IX) Proportion of outcome evaluations in relation to output and process evaluations.

Source: Extract from Table 1.1 in Furubo et al (2002).

Other factors explaining the differing positions of evaluation include the responsiveness of government authorities to empirically based research, the degree to which outcome orientation and programming is a natural part of administration, the nature of public discourse, the accessibility of data and other information, and the supply of evaluation training and studies (Beywl, 2006). Schmidt (2007) considers that policy evaluation would only flourish in "societies that provide liberal access to data, accept the merits of randomized assignment and guard the independence of research institutions" (379).

Constitutional, legal and administrative traditions and cultural values can be major obstacles, as the cases of Belgium, Switzerland, Italy and Spain illustrate. The historically low profile of evaluation in Belgium, which is not ranked in Table 1, has been attributed to the phenomenon of partitocracy (the undesirability of challenges to political consensus), the weakness of parliament and the federalisation of the country (with lack of coordination between levels) (Varone *et al*, 2005). Switzerland was also given a low score (although the institutionalisation of evaluation has been increasing since the late 1990s) and again the degree of power-sharing in the political system limits the scope of evaluation. As Bussmann (2008, 501) graphically explained:

> *"In this political system, so heavily influenced by various political actors, evaluation cannot play a decisive role. It must serve as a resource for all partners (or rather opponents). Evaluation, at its best, can help to illuminate the stakes at hand and to improve the quality of argumentation within the legislative process. Often it is used as ammunition in the political process by different interests involved. At its worst, it is distorted or completely ignored".*

In Italy, Stames (1998, 92) explained the reasons for the absence of an evaluation tradition in blunt terms: "nobody bothers to demonstrate the effectiveness of a program, and even its manifest failure is tolerated if it can be attributed to a latent consensus". While political awareness on the potential of evaluation grew in Italy during the 1990s, the major constraint was adequate knowledge. Spanish interest in evaluation has also been growing since the 1980s (Pazos and Zapico-Gõni, 2002), with the creation of evaluation institutions, publications, training courses and programmes; an institutionalisation of evaluation within government is said to be underway, albeit it "is not part of a consistent policy evaluation system" (Viñas, 2009, 467) and lacks a systematic approach. As Fernández-Ramirez and Rebolloso (2006, 144) noted:

> *"it is a growing field in the process of becoming consolidated and formalized, with a multitude of professionals and scholars, but lacking in identity and with limited networking. The effort of a few to reconcile evaluation added to political support contrasts with the lack of definition of what evaluation means, what the technical competencies are, or the goals and procedures."*

Ireland is much further down the road of institutionalising evaluation, especially through recent public service reforms, but here one of the potential difficulties is the lack of systematic initiatives to develop evaluation skills (Boyle, 2002). In the Irish context, as noted above, it has also been argued that evaluation has developed largely in response to external pressure, with little "appetite for evidence driven policy among senior political or public sector leaders" (O'Hara *et al*, 2007). However, evaluation practice is now spreading beyond EU-funded programmes and being adopted in domestic policy as part of a broader approach to the quality assurance of government action (McNamara *et al*, 2009).

Even in those countries which score relatively high scores in Table 1, there are question marks over the evaluation culture, particularly related to the standards of evaluation. In Germany, for example, it has been argued that despite evaluation activity since the 1960s, an evaluation 'profession' as such did not develop until the mid 1990s (Struhkamp, 2005). In Denmark, where evaluation has grown rapidly since the 1990s to become "almost routine in many areas of the public sector", the quality of evaluations is still deemed to be very inconsistent (Albæk and Rieper, 2002). In Sweden, despite the long tradition of evaluation, there are still big differences in evaluation activity between sectors (and the institutionalisation of evaluation

among government departments), with limited coordination and exchange of best practice (Forss, 2008).

Compared to Western Europe, the Central and Eastern European countries do not have a tradition of policy evaluation, and (at least in the past) the term has tended to be conflated with monitoring, audit and control. In preparing for EU accession, considerable efforts were made to develop an evaluation culture, encouraged by PHARE pre-accession programmes. Particularly during the 1990s, however, the scope for utilising evaluation within government administrations was limited by institutional instability, incomplete administrative reforms, the lack of qualified staff, and inexperience with tools and systems for programme development (Salamin and Peti, 2002; Toulemonde and Bjornkilde, 2003; Blazek and Vozab, 2006).

These problems have persisted until recently. A 2006 survey of evaluation capacity in the new Member States found that despite improvements (e.g. increased availability of evaluators), an evaluation culture was not well-established compared with the EU15, most notably because of the lack of procedures to ensure that evaluations contributed to policymaking (CSES, 2006). As in some EU15 Member States, the use of evaluation in several Central and Eastern European countries (e.g. Romania) is being externally driven, and it is questionable whether there is always a genuine public sector interest in developing national knowledge on evaluation, compounded by the scarcity of time, dedicated professionals, information and resources (Mihalache, 2006; 2009)[2]. Conversely, in Lithuania, a recent assessment of evaluation arrangements indicated that growing evaluation demand on the part of government authorities was being constrained by limited supply (Jarmalavičiūtė, 2008); discussions in the DG REGIO Evaluation Network suggest that the same may be true of some other small countries[3]. The accelerating interest in a professionalisation of the discipline is indicated by the fact that the longer-standing Polish Evaluation Society (formed in 1999) has recently been joined by the Hungarian Evaluation Network, launched in September 2006; the Slovakia Evaluation Society, created in 2006; the Czech Evaluation Society formed in 2007; and the Romanian Evaluation Informal Network (currently seeking legal status).

3. The evaluation of regional policy in Europe

3.1. Drivers of regional policy evaluation

The evaluation of regional policy is a prime example of the trends discussed above. More so than ever before, regional policymakers in European countries are using

[2] This applies still more to countries neighbouring the EU, such as Moldova, Russia, Turkey and Ukraine, where evaluation activity is dependent on international evaluation networking and the evaluation requirements of foreign donor organisations (Karimov et al, 2007; UNDP, 2007).

[3] See meeting minutes of the DG REGIO Evaluation Network (and discussion of 2007-13 evaluation arrangements) at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/tech_en.htm

evaluation to assess the effectiveness and efficiency of policy, to guide the design of new policies and to support the implementation of policy programmes and instruments. This is strongly associated with the profound changes experienced by regional policy over the past 15 years, sometimes termed a shift in the paradigm of regional policy (Bachtler, 2001; Pezzini, 2003; Bachtler and Yuill, 2007). Compared to traditional policy approaches, contemporary regional policies have several distinctive features. They target both equity and efficiency at the same time, shifting the policy focus from redistribution to competitiveness. They favour supply-side instruments and bottom-up local economic development initiatives, embodying a stronger spatial but also thematic/sectoral targeting of resources. And they are implemented and delivered by different (broader) actors and mechanisms, according a greater role to regional and local public and private actors, and planned within a multi-annual strategic framework.

Crucially, regional policies also now accord greater importance accorded to evaluation. As noted above, this reflects broader trends for public bodies to be more accountable for government expenditure and of governments requiring evidence to improve the management and delivery of policies. Integral to the new regional policy paradigm is the need for better understanding of how regional economies function and the effectiveness of different combinations of policy inputs, particularly in the context of more regionalised management of regional development policies as a result of devolution, deconcentration or other decentralisation trends.

Apart from the changes to the design and content of regional policy, an important driver of policy change in regional development, and particularly the use of evaluation, has been EU Cohesion policy (Bachtler and Wren, 2006; Bachtler and Michie, 1995). Since the reform of the Structural Funds in 1988, the evaluation of Cohesion policy has been given a much greater profile in EU regulations, requiring the Commission services and managing authorities in the Member States to monitor and evaluate Structural Funds programmes at different stages in the programme management cycle. From the start in 1988, the preparation of regional development programmes had to include an *ex ante* appraisal of the expected impacts of programme interventions with quantified targets, an evaluation which was increasingly undertaken by external evaluators in the 1994–1999 and 2000–2006 programme periods. This was complemented by interim evaluations of programme progress undertaken at the mid-point of programme periods, carried out by some programmes in 1989–1993 and 1994–1999 and systematically by all programmes in 2000–2006.

The early experiences of this approach were mixed. Initial assessments found a lack of Member State compliance with EU evaluation requirements, inconsistent evaluation methodologies and inadequate capacities and skills within the Commission services (Levy, 1996). However, over time, the Commission has placed considerable pressure on Member States to take the regulatory requirements on evaluation seriously (Bachtler and Mendez, 2007; Raines and Taylor, 2002), and it has undertaken

ex post evaluations of all interventions at the end of each programme period. The DG REGIO Evaluation Unit has also commissioned an extensive range of strategic and thematic evaluation studies and supported the development of evaluation capacity with guidelines, investment in evaluation methodologies and resources[4], as well as the organisation of a network of evaluation administrators and international evaluation conferences.

The effects have been threefold (Bachtler and Wren, 2006; Polverari *et al*, 2006; Polverari and Bachtler, 2004; Raines and Taylor, 2002). First, national and regional authorities responsible for managing Structural Funds across Europe have been required to comply with the EU evaluation obligations, investing time and resources in planning and commissioning evaluation studies – in many cases for the first time. The requirements have also stimulated the supply of evaluation services from consultants, universities and research centres. Second, although many authorities may have a 'compliance approach' towards evaluation, fulfilling their obligations for accountability purposes, there is evidence that the value of evaluation for the effective design and management of programmes has been increasingly recognised. Third, the outcome of these two trends, has been a 'spillover effect' on the evaluation of domestic public policy, with the adoption of EU evaluation procedures in non-EU funded areas of government intervention, and more generally an impact on national evaluation cultures.

3.2. National case studies in the evaluation of regional policy

To illustrate the above trends, the following section provides brief summaries of the development of evaluation in several EU Member States. Drawing on research reported in Polverari and Bachtler (2004), it begins with countries where evaluation of regional policy is long-standing – Germany, the United Kingdom, the Netherlands, Sweden – followed by Member States where regional policy evaluation emerged in the 1990s – Austria, Finland, Ireland, and Italy – driven by policy change and Structural Funds requirements. Lastly, an example of the more recent development of regional evaluation in the new Member States is summarised, with the case of Poland.

In ***Germany***, regional policy (like other areas of policy support) has been subject to regular 'success controls' (*Erfolgskontrolle*) for more than 30 years. The approach taken to evaluation has involved a mix of annual project-level monitoring, to assess the investment and employment associated with assisted projects, and national or regional-level studies assessing the economic impacts of policy of interventions on regional disparities. However, attempts to undertake meso-level evaluations, analysing the performance of individual businesses in receipt of regional aid were inhibited during the 1980s by data protection legislation. Since the mid-1990s,

[4] The MEANS programme in the 1990s and see EVALSED in the 2000s (see: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evaluation_en.htm)

a new evaluation concept has been developed by a specialist institute (IAB), on behalf of the Federal Economics Ministry, involving business surveys and labour market region studies (Deutscher Bundestag, 2007). This has allowed the matching of time series data of assisted firms with a reference panel of ca. seven million firms in terms of employment, investment, wages, firm stability and contributions to regional income and employment, providing new insights into the effectiveness of regional policy instruments (IAB, 2004).

Notwithstanding this tradition of regional policy evaluation with respect to regional aid, it should be noted that, since 1989 the Structural Funds have widened the use of evaluation in other areas of regional development in Germany. Indeed, it has been said that the value placed on evaluation in Germany changed dramatically based on the experience of evaluating Structural Funds in the 1994–1999 period, albeit with some concerns about the way in which the EU evaluation obligations were being applied (Scheffel, 2001). At *Land* level, in North-Rhine Westphalia, Jakoby (2006) commented that "systematic regional policy evaluations were not undertaken in NRW before the fundamental reform of the Structural Funds in 1988/1989"; since then, evaluation has been increasingly recognised for its learning potential and scope for catalysing changes to policies and programmes.

The ***United Kingdom*** has a long-established evaluation culture, in particular in the field of regional policy where evaluation has been undertaken for over 40 years. The effectiveness and efficiency of regional policy as a whole, and individual policy instruments, have been periodically evaluated throughout that period (see for example Arup, 2000), mainly through surveys of assisted businesses and modelling, with a strong focus on analysing employment creation and cost-per-job measures. Over the past decade, evaluation has been given new impetus by the UK government's commitment to 'evidence-based policymaking' requiring that new policies must have evaluation of their effectiveness built into them from the start (Cabinet Office, 2000). This has been part of a wider agenda of policy reform and public service modernisation, in which evaluation as been used as a 'change agent' by central government, for example to encourage local authorities to improve performance in service provision (Martin and Sanderson, 1999).

Regional policy evaluation has also been driven by the UK government's changing assessment of the 'regional problem', viewing regional differentials largely the result of long-term differences in performance against several 'drivers' of productivity, and demanding a more sophisticated approach to evaluation, accommodating long time horizons with high levels of complexity (Ling, 2003). Further stimulus has come from devolution (to Scotland, Wales and Northern Ireland) and regionalisation within England, leading to new regional development bodies and strategies and associated demands for analysis, monitoring and evaluation at different spatial scales (Martin and Thurston, 2001). Evaluation throughout the UK is governed by the Treasury 'Green Book' providing binding guidance for Departments and executive agencies on appraisal and Evaluation. The Department for Business Innovation & Skills (formerly

the Department of Trade & Industry), which is responsible for regional policy, has its own Performance and Evaluation Team which coordinates a rolling programme of evaluation. Similar units operate in the devolved administrations.

In ***the Netherlands***, policy evaluation has been part of the regional policy regime for many years. Regional policy operates on a 4–5 year cycle, and it is an established practice that policy instruments are evaluated towards the end of each cycle, with a view to improving and re-targeting future policy. The outcome is a spatial policy memorandum setting out policy objectives and priorities (see for example, Ministry of Economic Affairs, 2000; 2004). This was strengthened as a result of a government-wide system of management control introduced by the Ministry of Finance in the late 1990s, creating a budgetary framework within which central government departments were required to link together their objectives, activities and allocation of resources with the aim of ensuring "*the accountability of the minister for the efficiency and suitability of policy and its* periodical examination" (Ministry of Finance, 1999; 2002). Subsequently, the Ministry of Economic Affairs (responsible for regional policy) produced its own Handbook on Policy Evaluation and Implementation (*Handboek Beleidsevaluatie & Prestatiegegevens*) setting out evaluation guidelines, which have governed recent evaluations of all the main regional policy measures as well as the policy as a whole (IBO, 2004).

In ***Sweden***, evaluation is embedded in the policy-making process in the form of National Public Reviews (SOUs), i.e. evaluations carried out by Parliamentary Committees to review policy or its delivery. The most recent such review on regional policy was carried out in 2000 and formed the basis for a fundamental change in regional policy the following year (Regeringskansliet, 2000). Independent national authorities also undertake periodic independent evaluations on specific policy instruments or programmes. Accession to the EU and the implementation of Structural Fund programmes have increased the perception of the relevance of evaluation amongst policy-makers and have also impacted on the way evaluations are managed. For example, it has increased the use of calls for tenders with exact specification of the tasks to be performed; and it has promoted recourse to the '*ex-ante/mid-term/ex-post*' timetable also for domestic policy (e.g. the domestic regional policy Regional Growth Programmes have been evaluated on this cycle).

The heightened concern with evaluation also flows from the 2001 regional policy legislation which sought to improve the analytical base for policy development. A government-commissioned report from the Swedish Institute of Growth Policy Studies (ITPS) advocated that central and regional actors should possess systematic instruments in the form of holistic analytical foundations and overarching objectives. It proposed a common set of indicators to be used for formulating the Government's regional development policy and also as a basis for identifying specific regional developmental problems or unused development potential, as well as a basis for follow-up and evaluation (Kolmodin and Melkersson, 2004).

By contrast with these longer standing traditions of regional policy evaluation, are European countries where evaluation activity has emerged more recently, particularly since the 1990s and often attributable to the evaluation requirements of EU Cohesion policy.

Austria is representative of this second group of countries. The evaluation of regional policy – and policy evaluation more generally – has developed mainly in response to EU requirements (Holzinger, 2001). Apart from the activities of the Austrian Court of Auditors, which focuses primarily on financial and legal aspects, there was traditionally only a limited domestic evaluation of regional policy. As Huber (2006) has noted, prior to EU accession, evaluation was neither required by law nor part of the political culture; while some regional policy instruments were evaluated, this was voluntary and unsystematic. There was also considerable scepticism about the evaluation of regional policy (Holzinger, 2001). One explanatory factor is that Austria's domestic regional policy is relatively limited in terms of resources and instruments, not least because regional disparities are not very acute, and there are effective equalisation mechanisms for redistributing funds between regions. Another is that long-standing cooperation relationships between Austrian governmental institutions and other actors fostered a tradition of voluntary learning through informal networks rather than formal evaluation practices (Huber, 2006).

Since EU accession, evaluation has become more integrated to policy management and implementation for both Structural Funds and domestic programmes. Huber (2006, 279) states that "Austria decided to take the opportunity to start a joint-learning process rather than follow the easy route and limit their evaluation efforts to the minimum required for a purely formal exercise". The guidelines for domestic regional policy initiatives (*Richtlinien*) now include detailed provisions on both the implementation of interventions and their evaluation, and initiatives are systematically evaluated. A forum was also established – initially 'Checkpoint EVA' up to 1999, 'KAP EVA' from 2000 to 2006 – which facilitated cooperation on evaluation between federal and *Land* levels (the *Land* being responsible for Structural Fund programmes and many regional development initiatives), raised evaluation capacity amongst administrators and evaluators and more generally contributed to the development of an evaluation culture in Austria (Strohmeier and Holzinger, 2004; 2006).

In ***Finland***, the evaluation of regional policy, as in other areas of public policy, has been strongly influenced by the EU. Although evaluation studies can be traced back to the 1970s, Ahonen *et al* (2002) argue that the first comprehensive evaluations by government took place in the early 1990s as part of assessments by the Ministry of Finance of administrative reforms, and that "the advent of true program evaluation in Finnish government coincided with the Finland's establishment as an EU member state" (p.50). The use of evaluation for regional policy-making evolved rapidly after Finland joined the EU in 1995, in particular due to the interrelationship that exists between domestic policy and EU programmes (the Structural Funds are one of the funding sources that contribute to the strategic regional programmes), and the

creation of Regional Councils to manage Structural Funds and domestic regional development programmes.

A related factor was the evolution of regional policy in Finland. In the early 2000s, an extensive review of regional development policy was undertaken which reshaped the Finnish approach to regional policy. As part of the subsequent Regional Development Act passed in 2002, a series of requirements were introduced regarding the evaluation of any programme under the Act. Specifically, three distinct components of regional policy had to be evaluated by external evaluators at least once during or at the end of their lifetime: the strategic four-year regional programmes developed and delivered by each Regional Council; the regional development plans and targets agreed and implemented by certain sectoral ministries; and the special programmes contributing to national regional development goals (such as the Centre of Expertise Programme, the Regional Centre Development Programme, the Rural Policy Programme, and the Island Development Programme). The continuous assessment of policy effectiveness, notably through evaluation also figured prominently in the implementation of the Aid to Business Act (2000) which governed the award of regional aid (and business aid more generally). Lastly, in 2004, a Government Decision stated that "monitoring, effectiveness, evaluation and anticipation concerning government allocations for regional development in various administrative sectors will be enhanced", part of which involves the annual evaluation of the regional policy impacts of the proposals made with respect to the State budget each year (Ministry of the Interior, 2004). The cumulative effect of these steps has been an extensive programme of regular evaluation of both Structural Funds programmes as well as domestic initiatives, which has provided important insights into the effectiveness of regional policy programmes (OECD, 2005).

Ireland is one of the most striking illustrations of the impact that EU Cohesion policy rules have had on evaluation use and culture. As in Austria and Finland, during the 1970s and 1980s policy and programme evaluation was limited and peripheral to decision-making, although there was a relatively well-developed system of project evaluation (Tutty, 1994). The demand for evaluation of EU-funded programmes and a programme of public management reforms have contributed to a significant shift in the role and weight of evaluation in Ireland, involving not just regional policy but the whole of public policy in the broad sense.

Legislation was introduced in the 1990s to ensure the periodic review of public expenditure and to promote the development of evaluation procedures in government departments. The 1993 Comptroller and Auditor General (Amendment) Act and the 1997 Public Service Management Act together provided a legal mandate and requirement for the development of evaluation in all government departments. As a result, in 1997 an 'Expenditure Review Initiative' (ERI) was launched to provide a systematic analysis of what was being achieved by expenditure in each programme and to provide a foundation on which to base informed decisions for the prioritisation of expenditure within and between programmes. As part of the ERI, all areas of

expenditure became subject to review at least once every three years. In addition, triennial reviews of industrial policy became mandatory, to analyse implemented policy, to outline context changes and to identify the necessary changes in the direction of future policy.

The Structural Funds have had a major impact on the development of the Irish approach to evaluation (Hegarty, 2003). Independent evaluation units were established in the early 1990s in the government departments managing operational programmes for human resources and industrial development, and in the course of the 1990s an extensive programme of evaluation activity was undertaken related to the Structural Funds. This has had spillover effects on national evaluation practices. Most notably, under the National Development Plan for 2000–2006, an evaluation approach was adopted based on the 1999 Structural Funds regulations, even though the amount of NDP resources provided for by European funds was relatively minor. As under EU requirements, this approach was based on a cycle of *ex ante*, interim and *ex post* studies, supplemented by a programme of on-going thematic evaluations.

Lastly, under this group of countries, regional policy evaluation in ***Italy*** can be linked to the procedures and practices introduced by the Structural Funds (Polverari and Bachtler, 2004). Historically, Italy has not had a strong evaluation culture, at least as regards regional and economic development (other policy spheres, such as the health sector or labour market policies have a stronger evaluation tradition). Prior to the late 1980s, evaluation was mainly project-related and *ex ante* in the sphere of regional and economic development. No *ex post* evaluations or programme-wide evaluations were conducted. It is only with the requirements of Structural Fund regulations, mainly from 1988, that evaluation became programme-wide.

Like Ireland, Italy is distinctive in its use of EU obligations as part of a proactive (although not always coherent) strategy for evaluation capacity building. As Casavala and Tagle (2003, 2) note: "Italy started building its capacity from a particularly low point. European requirements played a substantial role in shaping the way evaluations are conducted". Specifically, the evaluation framework of the Structural Funds were used by regional policymakers in the Department for Development and Cohesion Policies as a lever to develop the use of evaluation in all of the policies for the problem regions ('underutilised areas'). As part of a wider 'strategic turn-around' in Italian regional policy which sought to improve the effectiveness and efficiency of intervention, improvements in evaluation practice were seen as a way of increasing the verifiability of policy and improving policy design (Barca, 2006). This was done particularly by building up evaluation demand and by generating greater understanding of the utility of evaluation among those involved in programming and implementation, mainly through capacity building and networking activities. Annual reporting to parliament by government departments providing regional support (Ministry of Economy and Finance, Ministry of Productive Activities) increasingly took on an evaluation character. Initial assessments of the new strategy were judged

to be 'remarkable' (Barca 2006), encouraging more transparent decision-making in the light of expected and actual results and greater accountability. The longer term impact is less clear, particularly whether a commitment to evaluation is truly embedded within government authorities responsible for regional development at national and regional levels.

Looking beyond the EU15 Member States, as noted above, evaluation in the Central and Eastern European countries is still in its infancy. Nevertheless, here too, the evaluation obligations of EU Cohesion policy are playing a part in encouraging greater use of evaluation in the regional development field. The pace of change in evaluation was significant in the first Structural Funds programme period (2004–06), with considerable investment in capacity and methods, albeit with varying levels of commitment and scope. Notable also is the evidence of spillovers from EU evaluation experience into domestic practice within a relatively short time frame (Bachtler *et al*, 2009).

Among the Central and Eastern European countries, regional policy evaluation has evolved quicker and more substantively in Poland than elsewhere among the new Member States. The evaluation of Polish regional policy was driven first by PHARE and then by Cohesion policy regulations. Initially, however, while government officials were ready to implement monitoring and evaluation procedures, the evaluation capacity was lacking. As Olejniczak (2002, 8) noted: "there were no actors able to do it and no knowledge about the theory and methodology of evaluation of public policies. The concept of evaluating public intervention or the support was still alien not only to the administrators but also the academics". Politicisation of the system was another difficulty.

Nevertheless, as part of the regionalisation of public administration and introduction of strategic planning in the pre-accession period, evaluation requirements were also introduced as part of domestic regional development legislation, monitoring and evaluation units were created within government departments, a monitoring system was developed, and the first evaluations commissioned (Olejniczak, 2002). Since accession, evaluation activity has accelerated; the National Evaluation Unit calculated that 135 evaluation studies related to Cohesion policy were carried out between 2004 and 2007, although most of them were commissioned by evaluation units in just two managing authorities (regional development and human resources) (Ministry of Regional Development, 2008). The supply of evaluation services and the quality of evaluations were also said by the Unit to have increased, supported by investment in training and regular evaluation conferences promoting exchange of experience nationally and internationally.

4. Conclusions

This chapter has provided an assessment of the development of the evaluation culture in Europe, focusing on the evaluation of regional policy. It has highlighted the different stages in the development of evaluation in various countries, contrasting those where regional policy evaluation is long-standing with those where it is more recent, and it has identified the key drivers of change. This section concludes with some observations on the future development of evaluation.

An important vector in the growth of evaluation in Europe has been the EU. The ongoing debates at EU level on financial management and budget discipline suggest that evaluation, along with performance measurement, cost-benefit analysis and impact assessment, will continue to be important in EU policymaking in the future[5]. With respect to regional policy, in many EU Member States, Structural Funds regulations have played an important role in promoting an evaluation culture with increased awareness of, and commitment to, evaluation. However, some commentators have questioned the embeddedness of this commitment and whether it will survive the reduction or withdrawal of Structural Funds intervention (McNamara *et al*, 2009).

The limitations of current evaluation practice should also be noted. This chapter has focused largely on the evaluation culture in Europe, with little mention of the methods for evaluating regional policy. The paradigm shift in regional policy discussed above present major challenges for evaluation, with a requirement for new or different methodological approaches that have yet to be fully appreciated in evaluation practice. Diez (2002), Stern (2003) and McVittie and Swales (2006) cite several characteristics of regional policy which complicate regional policy evaluation and require new methodological thinking. These include: the fact that policy innovations are based on new theories that are more difficult to test, especially in the absence of relevant data; the presence of intangible objectives, such as the creation of knowledge, learning and capacity building; the integration of transversal themes into regional policy, such as gender equality or environmental sustainability; the complexity of cause-effect relationships in processes involving large numbers of agents and organisations, and multiple funding sources; the systemic nature of regional policy intervention, targeting different stages in innovation or entrepreneurship processes; the difficulties of taking account of the cultural and political context and socio-economic conditions, in which new regional policies are rooted; the dynamism and flexibility of development processes; the role of the regions as 'animateur' involving the greater participation of regional actors; and the extended timescales over which new regional policy interventions are expected to operate.

A further issue concerns capacity development. This applies not just to the institutionalisation of evaluation within government administrations, at different levels,

[5] See, for example, Member State contributions to the current budget reform debate at http://ec.europa.eu/budget/reform/issues/read_en.htm.

but also the professionalisation of the discipline. With the growth of evaluation societies in many European countries over the past 5–10 years, there is scope for more systematic training – especially through university courses – and the development of EU-wide evaluation standards that can contribute to raising the quality of public policy evaluation.

Finally, there is the question of what this increased evaluation activity is achieving. As Leuuw (2004) points out, European evaluations have tended to focus on processes and outcomes rather than impacts, with implications for the degree to which evaluation can contribute to policy design. The experience of the Netherlands is typical of other countries also: five years after the introduction of a new Dutch approach to policy accountability, an official assessment concluded that while "Policy Budgets and Policy Accountability had stimulated people to give more thought to the intended results of policy……the relationship between objectives, instruments and resources is still not clear enough" (IOFEZ, 2004, 42). Further, the state of central government evaluation was judged to be disappointing: "a great deal of *ex post* evaluation takes place, but this does not review the effects of policy, and hardly any research is *ex ante*" (*ibid*). Notwithstanding the scope for much more impact assessment, it is important to recognise the limitations of what is possible, especially in the light of the complexity, overlap and interrelationships of current policy development, which suggest that expectations of what evaluation can deliver, may be too high. There are also potential unintended or undesired effects associated with an (over)emphasis on evaluation where effective organisational or policy management may be inhibited by focusing on performance measures and evaluation methods that do not facilitate dialogue and learning (Leuuw, 2000; Van Der Knaap, 2006).

Most intriguing is the lack of evidence of evaluation influencing policy decisions. The nature of the policymaking process is of course unlikely to show any direct cause-and-effect relationship between evaluation evidence and policy design amid the multiple, complex and competing influences on policymakers. Indeed, Nilsson *et al* (2008) argue that evaluations are primarily conducted to confirm – rather than inform or even challenge – policy decisions. This suggests a need for better understanding among evaluators of how evidence is treated in different institutional systems, with a concomitant adaptation of methods and tools to suit the practices and procedures of different jurisdictions.

References

Ahonen P., Virtanen P. and Uuskiylä P. (2002) Evaluation in Finland, in J.-E. Furubo *et al* (Eds) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, 47–62.

Albæk E. and Rieper O. (2002) Evaluation in Danish Governance, in J.-E. Furubo *et al* (Eds) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, 27–46.

Arup (2000) *Evaluation of Regional Selective Assistance, 1991–1995*, Report to the DTI, National Assembly for Wales, Scottish Executive, Arup Economics & Planning, London.

Bachtler J. and Michie R. (1995) A new era in EU regional policy evaluation? The appraisal of the Structural Funds, Regional Studies, 29, 745–752.

Bachtler J., Polverari L., Oraže, Clement, Tödtling-Schönhofer H., Gross F., McMaster I. and Naylon I. (2009) *ex post evaluation of Cohesion policy: the management and implementation of the ERDF programmes, 2000–06*, Final report to the European Commission (DG Regio), European Policies Research Centre, University of Strathclyde, Glasgow.

Bachtler J. and Wren C. (2006) Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges, *Regional Studies*, 40(2), 143–154.

Barca F. (2006) European Union Evaluation between Myth and Reality: Reflections on the Italian Experience, Regional Studies, 40(2), 273–276.

Beywl W. (2006) The role of evaluation in democracy: Can it be strengthened by evaluation standards: a European perspective, *Journal of MultiDisciplinary Evaluation*, 6(3), 10–29.

Blazek J. and Vozab J. (2006) *Ex-ante* Evaluation in the New Member States: The Case of the Czech Republic, *Regional Studies*, 40(2), 237–248.

Bœuf J.-L. (2001) *L'evaluation des politiques publiques, problèmes politiques et sociaux*, No.853, 23 February 2001.

Boyle R. (2002) A Two-Tiered Approach: Evaluation Practice in the Republic of Ireland, in J-E Furubo *et al* (Eds) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, 261–272.

Bussman, W. (2008) The Emergence of Evaluation in Switzerland, *Evaluation*, 14(4) 499–506.

Casavola P. and Tagle L. (2003), *Building Capacity for Evaluation: Lessons from Italy*, Paper presented at the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26–27 June 2003.

CSES (2006) *Evaluation Capacity in the New EU Member States*, Paper to the European Evaluation Society and UK Evaluation Society, Joint Conference, London, 4–6 October 2006.

Derlien H.-U. and Rist R. C. (2002) Policy Evaluation in International Comparison, in J-E Furubo *et al* (Eds) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, 439–455.

Deutscher Bundestag (2003) *Sechsunddreißigster Rahmenplan der "Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur" für den Zeitraum 2007 bis 2010*, Deutscher Bundestag, Drucksache 16/5215, 27 April 2007.

Diez M.-A. (2002) Evaluating New Regional Policies: Reviewing the Theory and Practice, *Evaluation*, 8(3), 285–305.

European Commission (2000) *Focus on Results: Strengthening Evaluation of Commission Activities, Communication to the Commission*, SEC(2000)1051, Brussels, 26 July 2000.

European Commission (2007) *Responding to Strategic Needs: Reinforcing the use of evaluation*, Communication to the Commission, SEC(2007)213, Brussels, 21 February 2007.

Forss K. (2008) *Evaluation in Sweden – the persistence of traditions*, Paper to the French Evaluation Society meeting, Strasbourg.

Furubo J.-E., Rist R. C. and Sandahl R. (Eds) (2002) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London.

Gavigan J. P. and Scapolo F. (2002a) *Regional Foresight – Future-Proofing and Validating Development Strategies*, ITPS Futures Group.

Gavigan J. P. and Scapolo F. (2002b) *Foresight and the Long-Term View for Regional Development*, IPTS Futures Group.

Hegarty D. (2003) *Framework for the Evaluation of the Structural Funds in Ireland*, Paper to the Fifth European Conference on the Evaluation of the Structural Funds, Budapest, 26–27 June 2003.

Holzinger E. (2001) *Sechs Jahre Programmevaluation – eine Bilanz*, Österreichisches Institut für Raumplanung (ÖIR), Vienna.

Huber W. (2006) Evaluation of European Union Cohesion Policy: Window-dressing, Formal Exercise or Coordinated Learning Process? *Regional Studies*, 40(2), 277–281.

HMT (2003) *The Green Book. Appraisal and Evaluation in Central Government*, Treasury Guidance, Her Majesty's Treasury, London, The Stationery Office.

IAB (2004) *Wie erfolgreich sind Subventionen? Investitionsförderung auf dem Prüfstand (How effective are subsidies? investment assistance under examination)* Gutachten im Rahmen der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur", IAB-Gutachten Nr 1/2004, Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit, Nürnberg.

IBO (2004) *De effecteviteit van regionaal beleid (The effectiveness of regional policy)* Interdepartementaal Beleidsonderzoek (IBO) Regionaal Economisch Beleid, The Hague, June 2004.

IOFEZ (2004) *Policy budgets and policy accountability: evaluation – Lessons from practice*, Interministerial Consultations for Financial and Economic Affairs (IOFEZ), The Hague, December 2004.

Jacob S. and Varone F. (2003) *Évaluer l'action publique: Etat des lieux et perspective en Belgique,* Gent, Academia Press.

Jakoby H. (2006) *Evaluation as Part of the Regional Policy Life Cycle: The Example of North Rhine-Westphalia*, Germany, Regional Studies, 40(2), 281–285.

Jarmalavičiūtė N. (2008) *Evaluation arrangements in Lithuania 2007–2013*, Presentation to the DG REGIO Evaluation Network meeting, Brussels, 26 June 2008.

Karimov A., Borovykh A., Kuzmin A., Abdykadyrova A., Efendiev D., Greshnova E., Konovalova E., Frants I., Palivoda L., Usifili S. and Balakirev V. (2007) Program evaluation development in the Newly Independent States, *Journal of MultiDisciplinary Evaluation*, 4(7), 84–91.

Kolmodin A. and Melkersson M. (2004) *Den nye regionala utvecklingspolitiken – Hur följa upp och effektutvärdera?* Swedish Institute for Growth Policy Studies, ITPS, May 2004, Stockholm.

Leuuw F. L. (2002) Evaluation in Europe 2000: Challenges to a Growth Industry, *Evaluation* 8(1), 5–12.

Leuuw F. L. (2004) Evaluation in Europe, in R. Stockmann R. (Ed.) *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*, 2nd edition, Opladen, Leske & Budrich.

Levy R. (1996) Managing Value-for-Money Audit in the European Union: The Challenge of Diversity, *Journal of Common Market Studies*, 34(4), 510–529.

Ling T. (2003) Ex Ante Evaluation and the Changing Public Audit Function: the Scenario Planning Approach' *Evaluation* 9(4): 437–452.

Martin S. and Sanderson I. (1999) Evaluating Public Policy Experiments: Measuring Outcomes, Monitoring Processes or Managing Pilots? *Evaluation* 5(3), 245–258.

Martin S. and Thurston R. (2001) *Evaluation Post Devolution: Challenges and Opportunities*, A Discursive Symposium, UK Evaluation Society Annual Conference.

McNamara G., O'Hara J., Boyle R. and Sullivan C. (2009) Developing a Culture of Evaluation in the Irish Public Sector: The Case of Education, *Evaluation*, 15(1), 101–112.

Mihalache R. (2009) Evaluation Networking in Romania, *Evaluation*, 15(4) 473–482.

Ministry of Economic Affairs (2000) *White Paper on Spatial Economic Policy in the Netherlands: Dynamics in networks*, Ministerie van Economische Zaken, The Hague

Ministry of Economic Affairs (2004) *Peaks in the Delta: Territorial Economic Prospects – Making room for enterprise*, Ministerie van Economische Zaken, The Hague

Ministry of Finance (1999) *From Policy Budgets to Policy Accountability (Van beleidsbegroting tot beleidsverantwoording)* Ministerie van Financien, The Hague, May 1999.

Ministry of Finance (2002) *Management Control Framework: The frame of reference statement on operational management*, Ministerie van Financien, The Hague, April 2002.

Ministry of the Interior (2004) *Government Decision on Development Targets under the Regional Development Act*, Sisäasiainministeriö, Helsinki, 15 January 2004.

Ministry of Regional Development (2008) *Evaluation of the cohesion policy in Poland: Summary of previous experience. Plans and challenges for the future*, National Evaluation Unit, Department of Structural Policy Coordination, Ministry of Regional Development, Warsaw.

Mihalache R. (2006) *Evaluation and knowledge management in the regional and national context*, Paper to the UNDP Evaluation Policy Regional Workshop, Istanbul, 6–8 December 2006.

McVittie E. and Swales J. K. (2006) Regional Policy Evaluation: Ignorance, Evidence and Influence, *Strathclyde Discussion Papers in Economics*, No.03-06, Department of Economics, University of Strathclyde, Glasgow.

Nicol W. R. (1982) Estimating the effects of regional policy: A critique of the European experience, *Regional Studies*, 16(3), 199–210.

Nilsson M., Jordan A., Turnpenny J., Hertin J., Nykvist B. and Russel D. (2008) The use and non-use of policy appraisal tools in public policy making: an analysis of three European countries and the European Union, *Policy Sciences*, 41(4), 335–355.

O'Hara J., McNamara G., Boyle R. and Sullivan C. (2007) Contexts and constraints: An analysis of the evolution of evaluation in Ireland with particular reference to the education system, *Journal of MultiDisciplinary Evaluation*, 4(7), 75–83.

OECD (2001) *Results-based management in development cooperation agencies: a review of experience*, DAC Working Party on Aid Evaluation, Organisation for Economic Cooperation and Development, Paris.

OECD (2005) OECD *Territorial Development Reviews: Finland*, Organisation for Economic Cooperation and Development, Paris.

OECD (2007a) *Globalisation and Regional Economies: Can OECD Regions Compete in Global Industries?* Organisation for Economic Cooperation and Development, Paris.

OECD (2007b) *Performance Budgeting in OECD Countries*, Organisation for Economic Cooperation and Development, Paris.

Pazos M. and Zapico-Gõni R. (2002) Program evaluation in Spain: Taking off at the edge of the Twenty-First Century? in J-E Furubo *et al* (Eds) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, 291–306.

Polverari L. and Bachtler J. (2004) *Assessing the Evidence: The Evaluation of Regional Policy in Europe*, EoRPA Paper 04/5, European Regional Policy Research Consortium, European Policies Research Centre, University of Strathclyde, Glasgow.

Polverari L., Mendez C., Gross F. and Bachtler J. (2004) Making Sense Of European Cohesion Policy 2007–13: On-Going Evaluation And Monitoring Arrangements, *IQ-Net*

Thematic Paper, 21(2), European Policies Research Centre, University of Strathclyde, Glasgow.

Raines P. and Taylor S. (2002) Mid-term Evaluation of the 2000-06 Structural Funds Programmes, *IQ-Net Thematic Paper*, 11(2), European Policies Research Centre, University of Strathclyde, Glasgow.

Regeringskansliet (2000) *Regionalpoltisika utredningens slutbetänkande*, Statens offentliga utredningar, SOU 2000:87, Näringsdepartementet, Stockholm.

Salamin G.and Peti M. (2002) *First steps towards systematic regional programme evaluation: experiences of an accession country, Hungary*, Paper to the RSA Conference 'Evaluation and EU Regional Policy: New Questions and New Challenges, Aix-en-Provence, 31 May – 1 June 2002.

Scheffel K. (2001) Evaluierung – Chance oder Last? Eine Betrachtung aus Bundessicht, *Informationen zur Raumentwicklung*, Heft 6/7, 315–319.

Schmidt C. (2007) Policy evaluation and economic policy advice, *Advances in Statistical Analysis*, 91(4), 379–389.

Schröter D. (2004) Evaluation in Europe: An Overview, *Journal of MultiDisciplinary Evaluation*, 1, 58–62.

Stame N. (1998) Evaluation in Italy: Experience and Prospects, *Evaluation*, 4(1), 91–103.

Stame N. (2008)The European Project, Federalism and Evaluation, *Evaluation*, 14((2) 117–140.

Stern E. (2003) *The Rationale & Challenges for Evaluation of Public Policies*, Keynote Speech to the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26–27 June 2003.

Strohmeier G. and Holzinger E. (2004) *KAP-EVA Koordinations- und Arbeitsplatform Evaluierung: Bilanzbericht*, Österreiches Raumordnungskonferenz, Vienna.

Strohmeier G. and Holzinger E. (2006) *Die Koordinations- und Arbeitsplatform KAP-EVA: Ein Lernprozess in der Evaluierung der Strukturfonds in Österreich*, Österreiches Raumordnungskonferenz, Vienna.

Struhkamp G. (2005) Evaluation in Germany: An Overview, *Journal of MultiDisciplinary Evaluation*, 2(3), 180–194.

Toulemonde K. and Bjornkilde T. (2003) *Building evaluation capacity: experience and lessons in Member States and Acceding Countries*, Paper to the Fifth European Conference on Evaluation of the Structural Funds, Budapest, 26-27 June 2003.

Tutty M. (1994) *Improving Evaluation of Policies and Programmes in the Irish Public Service*, University of Dublin, Trinity College, unpublished Msc thesis (quoted in Boyle R. (1997) *Evaluating Public Expenditure Programmes: Determining a Role For Programme Review,* Institute of Public Administration, Dublin).

UNDP (2007) *Workshop Report*, UNDO Regional Workshop on the UNDP Evaluation Policy: European and Commonwealth of Independent States, Istanbul, 6–8 December 2006.

Van Der Knaap P. (2006) Responsive Evaluation and Performance Management, *Evaluation* 12(3), 278–293.

Varone F. (2007) Développer les capacitiés d'évaluation, in S Jacob, F Varone and J-L Henard (Eds.) *L'évaluation des politiques au niveau regional*, Brussels, Lang, 157–77.

Varone F., Jacob S. and De Winter L. (2005) Polity, Politics and Policy Evaluation in Belgium, *Evaluation*, 11(3) 253–273.

Viñas V. (2009) The European Union's Drive towards Public Policy Evaluation: The Case of Spain, *Evaluation*, 15(4), 459–472.

Wadley D. (1986) *Restructuring the Regions: Analysis, Policy Model and Prognosis*, Organisation for Economic Cooperation and Development, Paris.

Williams K., de Laat B. and Stern E. (2002) *The Use of Evaluation in the Commission Services*, Final Report to the European Commission, Technopolis France, Paris.

The Methodological Shortcomings of Current ERDF Evaluation Practice

Karol Olejniczak

1. Introduction

Evaluation has been a formal requirement in the implementation of Structural Funds since their very beginning (European Council, 1988 Article 6). The first experiences in this regard date back to the first round of the Cohesion Policy (in the late 1980s), and since then the number, complexity and scale of the exercise has been consistently and significantly growing.[1)]Nevertheless,methodological discussions, especially those devoted to the assessment of the interventions' effects, date back only to the mid-1990s. They are relatively new when compared to the American debate (which started in the mid-1970s) or the developments that took place in national administrations of some Member States (e.g., UK, France and Germany).

The milestones in this European Union debate include the MEANS initiative (Evaluation Methods for Actions of a Structural Nature) and its follow-up publications,[2)] in addition to a series of *Working Papers* (e.g. European Commission, 2000; 2002b; 2002a; 2006) which deal with basic issues of SF programming, monitoring and evaluation. However,these publications are intended merely as guidelines or manuals. They do not inform about the "current state of the art" of specific methodological developments in the practice of evaluation of regional interventions. They also offer only limited insight into real-life applications of the individual methods or approaches in relation to Structural Funds programmes. So far, the only holistic review of methodologies used for the evaluation of the effectiveness of Structural Funds has been presented by the EPRC &Allander Institute (2000). This study was based on a case study method, providing valuable insights but telling us very little about the scale of specific practices. Additionally, this particular analysis was completed nearly 10 years ago, and thus took into account only the experiences of the old Member States.

Even though programmes co-financed from the EU have for many years been significant constituents of Member States' regional development policies, issues related to structural funds research are rarely encountered in the mainstream of

---

1 Olejniczak (2008, p. 86) estimated that over 2,000 EU regional policy evaluation studies were conducted in the EU-15 between 1994 and 2006. If we were to include such studies from the new Member States and 2007–2013 *ex ante* evaluations, this number would be doubled.

2 The first set of publications prepared following a series of seminars waspublished in the mid-1990s (European Commission, 1995a, 1995b, 1995c, 1995d, 1995e, 1995f and 1995g). We also refer to the main MEANS edition from 1999 (European Commission, 1999) as well as its two web-based versions (Tavistock Institute *et al*., 2003; GHK Consultants, 2006).

evaluation debates held in scientific periodicals or papers given at European Evaluation Society.[3)]

The issues discussed above show that although the practice of SF evaluation is dynamically growing, the know-how on current methodological developments is neither easy to trace nor access, assearching for the potential inspirations would require an in-depth and EU-wide effort.

In the case of Poland, the methodological experiences in the evaluation of the effects of public intervention have been both short in time and limited in scale. The reasons for this are twofold. Firstly, result-based management culture did not take root in Poland until the mid-1990s (and evaluation is inherently connected with such a management paradigm). Such an approach was first introduced in the early 2000s, owing to EU requirements and procedures (Olejniczak, 2007, p. 393; Kierzkowski, 2002). Secondly, Polish programmes have thus far not been covered by *ex post* evaluation studies, since the evaluation of assistance funds was directly commissioned by the European Commission, and as a rule the orders were placed with foreign evaluators. In the case of Structural Funds studies, most attention was focused on current management issues. Also, the first programming period has only recentlyended, and delays in projects do not allow for any comprehensive analysis of the programmes' results earlier than in the coming years.

The first empirical studies on the use of evaluation in Poland also showed that both the contracting bodies and contractors have had problems with assessing the intervention effects and with developing strategic knowledge based on *ex post* evaluations (Ferry, Olejniczak, 2008). Moreover, according to some national experts (Szlachta, 2006), evaluation methodologies used so far rarely went beyond the Commission's guidelines, and as such were predominantly technical and derivative in nature.

This lack of a straightforward point of reference for methodological discussions at the European level, combined with the Polish limited experience and pressure for the need to evaluate the results of wide-ranging EU investments provided the impulse for the research project entitled: "Evaluating the effects of regional development programmes – a comparative study of international practices".

---

[3] Structural Funds are absent from mainstream opinion-making scientific journals (such as *The American Journal of Evaluation, Evaluation and Program Planning andNew Directions in Evaluation*), which is mainly due to the American focus of the journals. That aside, in the European quarterly Evaluation, only as few as three papers on Structural Funds have been published over the past 10 years. Although EU structural funds are often a subject of articles in the quarterly *Regional Studies,* only one issue thus far has been strictly devoted to the question of Structural Funds evaluation (2006/40). At the most recent conference of the largest professional association, the European Evaluation Society (Lisbon, October 2008), Structural Funds issues were present only in few of the conference papers.

This project was carried out in 2008 by an international research team,[4] as part of a research grant scheme organised by the Polish Ministry of Regional Development and co-financed by the EU Technical Assistance programme.

The project was exploratory in nature, and its first part aimed to address the following issue: what are the evaluation practices concerning (EU co-financed) regional development programmes in Poland and selected European countries?[5] This chapter shows the findings from that particular part of the study, and is divided into three sections. The first section summarises the adopted definitions of the underlying notions and outlines the research methodology applied. The second section discusses the study's findings: the structure of the analysed population of reports, the methods used in developing research concepts and research methodologies, the methods used to compile and analyse data, and the actual orientation of results and recommendations from the reviewed evaluation studies. The third section offers conclusions as well as suggestions for further research.

2. Research methodology

Definitions

The research question for the project was: **What are the methodologies used in summative evaluation studies, devoted to the assessment of the effects of regional development programmes co-financed by the European Union?** Before addressing thisquestion, four elements need to be clarified.

Firstly, the term "**methodology**" refers to both the adopted overall research concept – the so-called evaluation approach[6] – and to its constituent, individual tools and research methods of data compilation and analysis used in a given evaluation.

The term "summative evaluations" refers both to the evaluation of completed or advanced programmes. Thus the focus is on both *ex post* evaluations and on-going evaluations (including mid-term evaluationsand their updates) of programmes which are more than halfway through.

---

4 The research team was composed of Polish and foreign staff and EUROREG contributors. On the Polish side, the team included: mgr Adam Płoszaj, mgr AlicjaWeremiuk, mgr KatarzynaWojnar, dr DominikaWojtowicz, dr Michał Wolański and Monika Kur. The contributors from other countries were: dr Kai Böhme of SWECO EURORFUTURES, Stockholm; dr Richard Boyle of the Institute of Public Administration (IPA), Dublin; dr Martin Ferry of the European Policies Research Centre (EPRC), University of Strathclyde, Glasgow, and Lukas Wortmann of Metis GmbH, Vienna. The team was headed by dr Karol Olejniczak of EUROREG, Warsaw University.

5 In the second part of the report, we looked for methodological inspirations from beyond the EU and the Structural Funds.

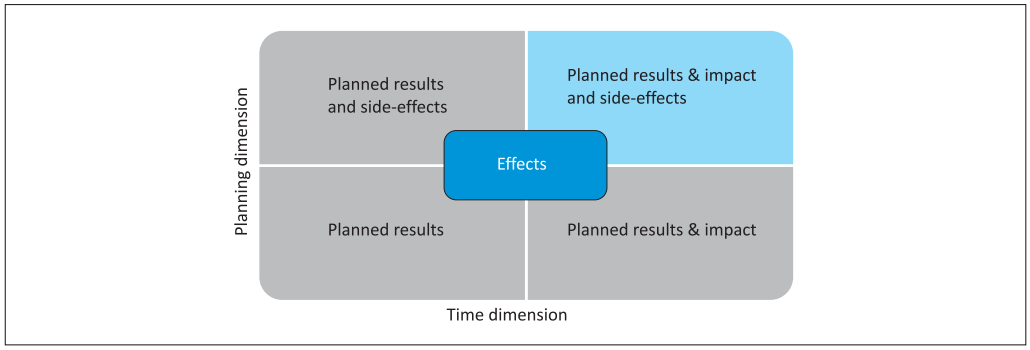
6 Mark *et al*. (2000, p.75) refer to it as "inquiry modes", arguing that this is a link connecting overall objectives of the study with very technical methods used in the analysis of a given intervention. On the other hand,Stufflebeam& Shinkfield (2007) adopt a broader interpretation, and use the term "approach" interchangeably with "evaluation models", while identifying 26 types of such models(according to the research objectives on which they focus their attention and the set of research methods). In this analysis, we tend toconcur with the interpretation proposed by Mark and contributors.

Thirdly, the term "**effects**" refers to the degree and scale of changes which a given programme is aimed and which it has actually achieved. The definition of this term should be viewed in two dimensions:

- The time dimension, which covers both results (the immediate effect of the intervention for its beneficiaries) and impact (long-term structural change);[7)]
- The planning dimension, which addresses both the effects envisaged by those who designed and implemented a given intervention and the so-called side-effects (whether positive or negative), which are visible despite being unanticipated in the original intervention blueprint.

It should be pointed out that these two dimensions can be interpreted in four different ways, some of them being highly reductive. The possible options are showed in Figure 1.

Figure 1. Two dimensions of intervention effects



Source: prepared by the author

In our analysis, we adopted a comprehensive interpretation of the definition of "effects", understood as planned effects and side-effects as well as impacts of a given intervention (the top right box in Fig. 1).

Last but not least, we should define the term "**regional development**". For the purpose of this analysis, it was assumed that regional development programmes refer to all programmes spatial in character, co-financed from the European Regional Development Fund. In the case of pre-accession programmes, the definition's latter component was not taken into account.

To sum up the discussion on terminological issues, we should also clarify the expectations concerning the evaluation of effects. According to publications in this area (Olejniczak, 2008, pp. 177–178; Scriven, 1991; Stern, 2005, p. 25) as well as EU legislative guidelines, summative evaluations are intended to perform two basic functions – an accountability function and a learning function. Both of these are strategic in nature,the practical implications of which are shown in the table below.

7 For a detailed definition of these terms, see: European Commission, 1999 vol. 6.

Table 1: Scope of the evaluation of effects

Field of exploration	Orientation of conclusions
I. Examine the extent to which the planned objectives of a given intervention have been achieved.	Conclusions which take intoaccount how the managers fulfilled the plan.
II. Identify real effects of the intervention and analyze the degree of the desirable social change which a given programme has generated.	Conclusions explaining the causal relations and showing the programme's utility for the society.
III. Analze the success/failure mechanism for a given intervention, including: • The role played by context (external) factors • The role played by the way in which a given intervention was delivered (implementation) • The role of programme assumptions concerning the mechanisms of desirable change (theory and earlier practices on which the assumptions concerning causal relations were formulated)	Conclusions explaining the mechanism of an effective social change in the areas addressed by a given programme, and showing the directions of future, effective strategies.

Source: prepared by the author.

According to Mark *et al*. (2000, p.78), it should be emphasised that the second and third row in column Orientation of conclusion reach beyond the simple description of visible effects, as it refers to causality. In order to assign changes observable in a given spatial system to a specific, completed programme, first of all any possible alternative explanations of such changes must be ruled out, and proofs must be found to confirm the cause-and-effect relationship between a given change and the programme's results. Such an approach poses a real challenge when we analyse multi-sectored programmes, with a large number of measures and multidimensional structure of goals.

The adopted research methodology

The overall research plan adopted for this particular study was cross-sectional.The research strategy consisted of four stages, which are discussed below.In search of inspiration, we decided to focus our attention on evaluation reports from countries with a well-developed evaluation culture.

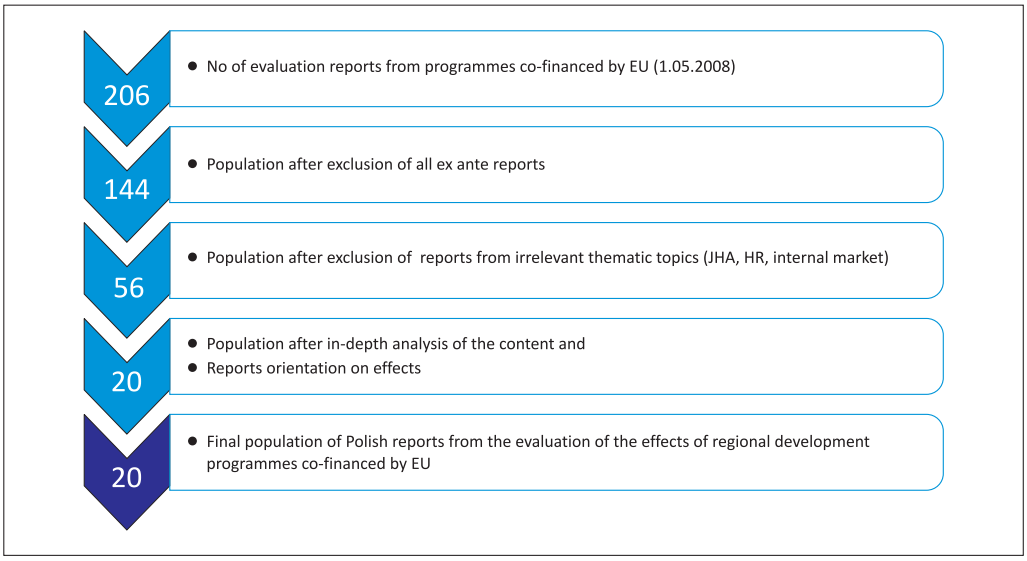
In the first stage the subject of the research was selected – these were 100 evaluation reports. Based on the findings from previous analyses, four EU countries with extensive experience evaluating public programmes were chosen (Bachtler *et al*., 2000; Furubo *et al*., 2002; Toulemonde, 2000). These were: Ireland, Sweden, Germany and the United Kingdom. Poland was added to these four countries to provide a point of reference. Then national experts chose 20 reports for each of the countries based on the following four criteria:

1. It was a regional policy programme– that is, it had to be financed from the ERDF or be implemented regionally or be clearly oriented to spatial aspects;

2. The evaluation had to be an *ex post* (i.e. for completed interventions) or – in the case of on-going or mid-term studies – for a programme whose implementation was considerably advanced;
3. Analysis and assessment of the effects of a given programme were listed among the formal objectives of a given evaluation;
4. The report had a separate section on methodology.

As a result of this procedure, 100 reports have been chosen for in-depth analysis.[8] A sample selection procedure is shown using Poland as the model (see the figure below).

Figure 2. Report selection procedure – the case of Poland

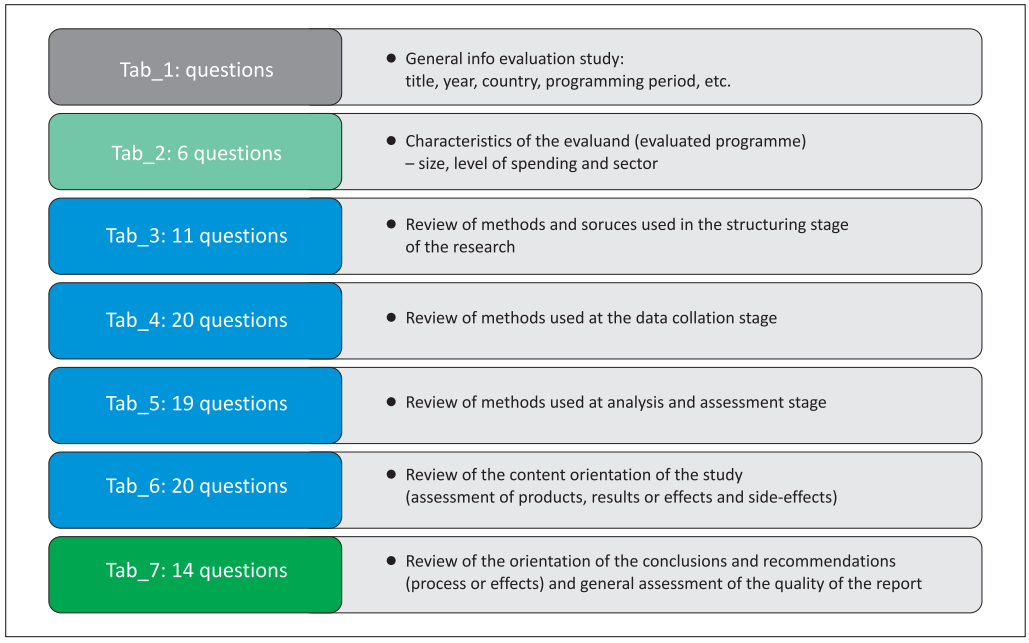


Source: prepared by the author.

At the second stage of our research, we formulated an analytical framework in the form of a database to ensure a systemic review of the methodologies adopted in each of 100 evaluation studies. The database's topical categories were developed on the basis of literature review (European Commission, 1995f, 1997, 1999, 2007; Howell, Yemane, 2006; Mathison, 2005; Shadish *et al*., 2001; Tavistock Institute *et al*., 2003). In effect, seven review tables (see Figure 3) were created with detailed analytical questions (100 analytical items in total).

[8] Initially, it was assumed that 20 reports from each of the countries with the best-developed chapters on methodology (i.e. where the probability of interesting inspiration was the highest) would be selected for the final analysis from among such a population of reports. In practice, however, it turned out that once the first three aforementioned criteria had been applied, a sample of 20 reports from every country for all practical purposes exhausted the entire population of studies on the effects of regional development programmes. We can say, therefore, that the group of 100 reports from five countries represents all of the evaluation research dealing with EU regional policy effects conducted at the level of national administrations.

Figure 3. Analytical sets of the database



Source: the project's database.

At the third stage of our research, the content of all 100 reports was analysed by a team of experts using the detailed analytical questions and standardised definitions formulated during stage two. The answers to these questions were input into the database.

The final stage of the project involved an analysis of the database records, which covered simple statistical listings compiled for the entire population of reports covered by the study. Methods used in these 100 studies were compared for each of the standard stages of the evaluation research: structuring data collection, analysis and assessment. Additionally, the real orientation of and recommendations from the studies were analysed.

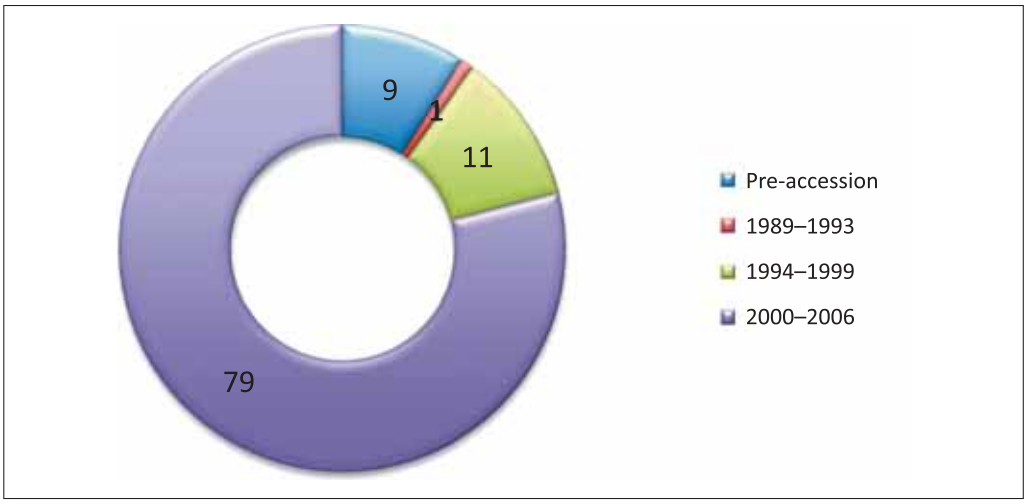
3. Results of the analysis

3.1. Basic characteristics of the analysed population of reports

As indicated above, the analysed population consisted of 100 reports from five countries: Ireland, Germany, Sweden, United Kingdom and Poland. Initially, we looked for evaluations from all related Structural Funds budgetary periods (i.e., years: 1989–93, 1994–99 and 2000–2006). In practice, however, it turned out that most reports dealt with programmes from the years 2000–2006 (see Figure 4). To some extent, this is due to the gradual improvements in evaluation practices in subsequent programming periods. But the primary reason for this was the require-

ments placed in the Regulations from 2000–06, which obliged old Member States to conduct mid-term evaluations and mid-term updates for all Structural Funds programmes (Council of the European Union, 1999 Article 42).

Figure 4. Number of studies by Structural Funds programming periods



Source: prepared by the author on the basis of the project's database.

This legal requirement in effectled to a predominance of mid-term evaluations and mid-term updates, which account for 60% of the entire surveyed population of reports, with only 19% of them being classic *ex post* evaluations (carried out only after the entire intervention had been completed), while the remaining 17% were on-going evaluations.[9)]

Naturally, all *ex post* evaluations faced the problem of transposing the conclusions from their analyses to new rounds of programming. This problem arose due to mismatches between the final evaluations and the planning of subsequent rounds of the same type of intervention. In this context, reports from the evaluation of PEACE (Northern Ireland) and PHARE CBC Poland-Germany programmes are particularly interesting as they strive to combine *ex post* evaluation with evaluation of subsequent rounds of the programme (see the box below).

[9] The difference in the number of mid-term evaluations and their updates arises from the exclusion of several reports in the selection process. These studies did not investigate effects but merely on-going processes in programmes which began very late (with expenditure in 2003 at a level of 0%) – and therefore no effects were available to be analysed. Such analyses were not possible until the mid-term update in 2005.

Table 2: Examples of combining various types of evaluations

The aim of PEACE was to support the peace process in Northern Ireland through regional development fostered by local project partnerships. In its attempt to ensure continuity of the overriding concepts between individual programme rounds, the Managing Authority combined in one contract *ex post* evaluation of PEACE I and mid-term evaluation of PEACE II. The Implementing Authority for PHARE CBC 1999–2001 Poland-Germany tried to transfer *ex post* experiences gained at Poland's western borders to the eastern border, where the Neighbourhood Programme Poland-Belarus-Ukraine was being prepared. In the Terms of Reference for the *ex post* evaluation, a provision was made that conclusions from the PHARE-CBC evaluation should also present good practices and suggestions which could be useful for future beneficiaries of the eastern Programme.

Source: prepared by the author on the basis of the project's database and evaluation reports

As regards the characteristics of evaluands[10)] (i.e., programmes being the subject of the reports) in addition to the prevalent group of multi-sectoral regional programmes (nearly half of all reports), a number of topical subgroups emerged: SMEs & innovation (20), cross-border cooperation (12), R&D (10) and transport (10). In terms of the financial scale of the intervention, the average value of programmes covered by the reports was EUR 300 million; there were 28 examples in the group of large programmes (i.e. with budgets over EUR 1 billion), of which as many as 11 interventions had budgets in excess of EUR 4 billion.

3.2. Methods used for planning and structuring evaluations

We started our analysis by looking at sources of inspiration of the evaluators (such as literature, methodologies and earlier studies) and tools they employed to assist them in the key stage of investigation – the structuring phase.

The sources of inspiration are presented in Figure 5. They were quite disappointing. Nearly half of the analysed studies lacked any scientific underpinnings of their research approach, while 15% of the population just copied earlier similar evaluations.

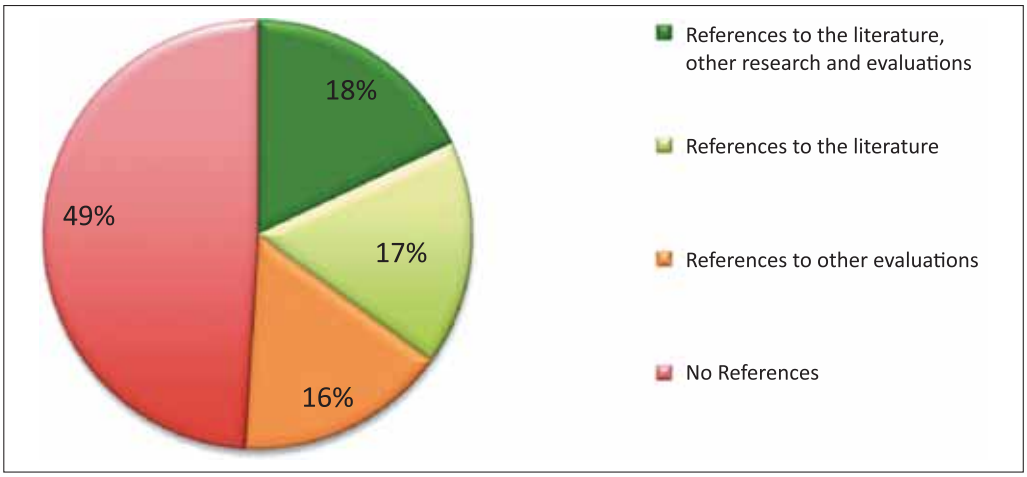
Only a small share of the studies (18%) were founded on broader methodological and conceptual groundwork, whether theoretical (textbooks and manuals) or practical (other research studies). The bulk of such casesare from Sweden (7) and – quite encouragingly – from Poland (also 7).

Looking closer at the set of studies that referenced methodological literature (35% of the entire report population), we can say that nearly half of them limited themselves to such sources as MEANS, Evalsed and the Commission's *Working Papers*. The *Working Papers*, however, cannot be regarded as a "methodological source" in its own right, since these are standardadministrative guidelines specifying the

[10] Evaluand is a term introduced by Michael Scriven to describe any object of an evaluation study. Thus, „evaluand" may be a group of persons, a programme, project, process or an idea (Mathison, 2005, p.139).

minimum requirements concerning questions and methods for a given type of evaluation study. To sum up, we can therefore conclude that the methodological inspirations for the surveyed set of reports were of an exceptionally low quality.

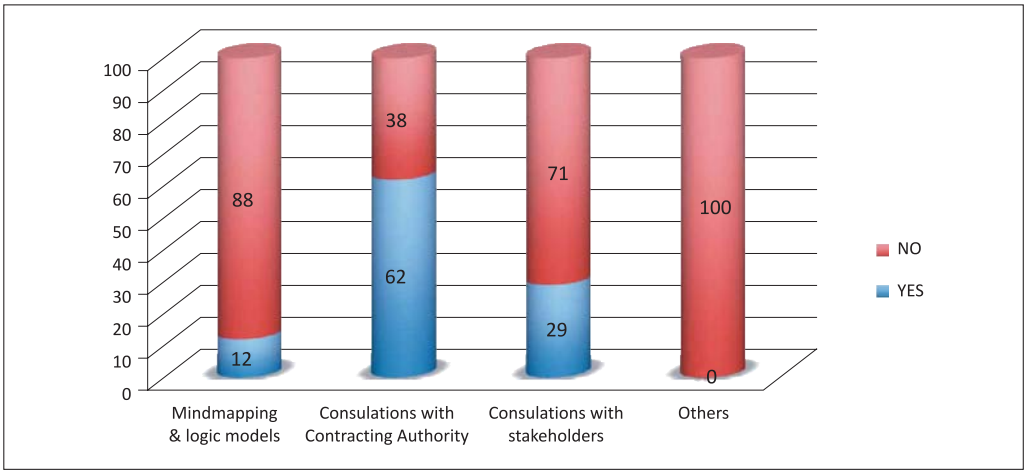
Figure 5. Sources used in the development of methodology



Source: prepared by the author on the basis of the project's database.

Tools used during the structuring stage are shown in Figure 7. The overwhelming majority of the evaluations merely fulfilled the standard of "talking to the client" and in some cases to other stakeholders,such as the Evaluation Steering Group, key institutions of the programme or Commission-level stakeholders. A striking lack of innovativeness can be observed.None of the analysed reports provided a single example of the structuring method other than one from the aforementioned three options (or rather two, because the second and the third option differ only in who the interlocutor is).

Figure 6. Tools used at the structuring stage



Source: prepared by the author on the basis of the project's database.

Nevertheless, special attention should be given to the small group of cases where logical models and maps were employed. As a rule, evaluators began by using the standard Commission's model,[11]then adapted it to meet the specific needs of the programme (as in the UK example summarised below), or in some cases, made far-reaching modifications. Normally, these models helped clarify the hierarchy of the programme's objectives and effects, and highlighted the distinctions between outputs and effects by relating them to indicators (which was particularly useful as their interpretation was in many cases distorted by beneficiaries during the implementation phase). Finally, these tools also allowed linking the analytical approach with the evaluation criteria. In two examples from Sweden, the adopted model additionally showed the cause-and-effect sequence of the intervention and was used as a "backbone" for structuring the entire study (all of its fields of analysis) and then as a template for assessment (this example is discussed in the next section).

Table 3. Example of the model's application in a research concept

In the Mid-Term Report for the East Wales Programme, evaluators used the standard Commission's model (inputs – activities – outputs), albeit with a modification, as they distinguished between top-down and bottom-up dimensions. The former included impacts, which were evaluated using the criterion of the sustainability of effects, measured by changes in regional trends in the context of the programme.The latter was ascribed results which were assessed on the basis of their utility for the final beneficiaries, and measured by the degree to which both expectations and needs have been satisfied. Both these dimensions made up the Community Added Value.

Source: (CSES, 2003, pp. 13–14)

3.3. Methods used for obtaining data

In this section, we analysed the sources and methods of data collection used in 100 evaluations of effects. The list is shown in Figure 7.

The inventory indicates that secondary data were used most frequently.This includedthe use of databases from the monitoring of programmes and general national and regional statistics (category: other secondary data), as well as a review of the programme records and documents, applications for funding, beneficiaries' applications, annual reports, as well as other administrative data.In this context, the quality of databases from the monitoring of programmes seems to be of primary importance.

[11] That is, the model involving the following components: needs – objectives – inputs –activities – outputs – results – impacts] and the related evaluation criteria.

Table 4. Example of the usefulness of monitoring databases

In the Irish-English PEACE I & II programmes, the Managing Authority entrusted the development and maintenance of the monitoring database to a private company. The database was designed using the five basic segments of the intervention logic model. The first segment (inputs) provided information about the programme's funds and additional sources of financing. The second segment included process indicators – number of partnerships, selected project orientation, information about equal opportunities (in the context of Irish-English proportions). The third module covered activities and products as well as project data. Module four summarised data on results – local increase in wages, jobs, revitalisation as well as social inclusion – integration of target groups, development of local communities. The fifth and final module on impacts provided data on reinforcing the cooperation effect (for example, in the form of joint projects) as well as proofs of promoting the idea of reconciliation. Such a structured database allowed evaluators to make a comprehensive analysis of the extent to which the programme objectives had been met, and to pose questions on the expediency or usefulness of some of the activities.

Source: prepared by the author on the basis of the project's database and evaluation reports

As regards primary data, most of them came from interviews and questionnaire surveys. The former were mostly conducted with key persons in the implementation system of a given programme, selected project managers or beneficiaries (usually as a component of a case study), while surveys typically covered direct project beneficiaries –that is, institutions receiving funding, those applying but not accepted in the selection process, and in some cases potential beneficiaries who did not take part in the calls for proposals. Surveys from final beneficiaries (individuals, communities and companies thatexperienced the impact of the projects) were extremely rare.

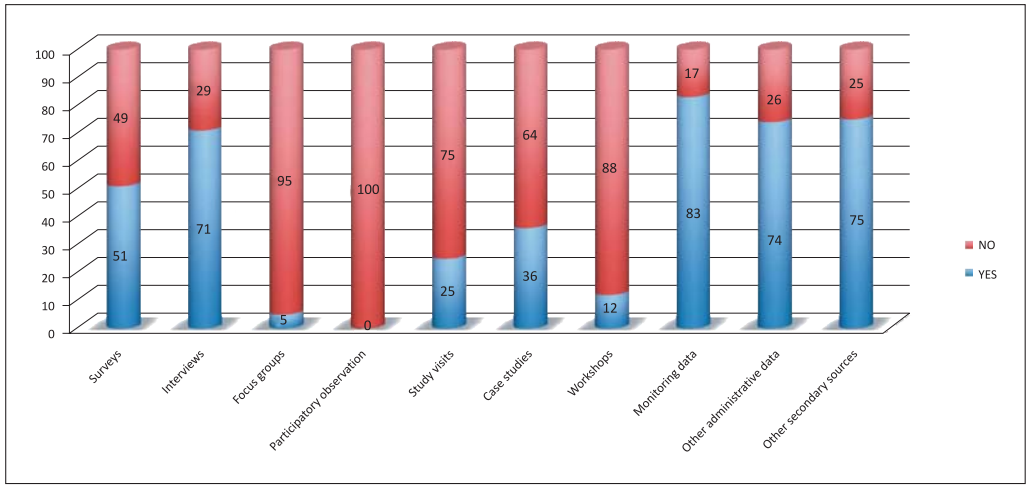
In nearly 40 situations, case studies were used to collect and then analyze the data. This was usually connected with on-site visits.[12)]As a rulethe subjects of case studies were projects; although there were a few situations when the beneficiaries' partnerships were surveyed (e.g. institutions or networks of institutions which won several projects), and in one case not a project but a selected spatial system (municipality) covered by the programme was subject to analysis. The average number of case studies ranged from 5 to 10, even though 6 evaluations covered over 20 projects. The Polish evaluation of the RAPID programme was an absolute exception, as it discussed 55 case studies; however, on the whole these had the form of descriptive information about the projects rather than in-depth studies. This example shows a correlation whereby a larger number of cases meant more limited and superficial investigation methods; e.g. the evaluators restricted themselves to telephone interviews and a brief review of documents.

[12] In two situations, on-site inspections were employed to identify projects for case studies.

Focus group interviews were relatively rare; in two out of five cases, they were in the form of discussions within the working groups organised on the basis of the programme's priorities.

Novel practices related to data compilation are synthetically presented in Table 5 below.

Figure 7. Methods and sources used at the data collection stage.



Source: prepared by the author on the basis of the project's database.

Table 5. Examples of innovative data collection methods

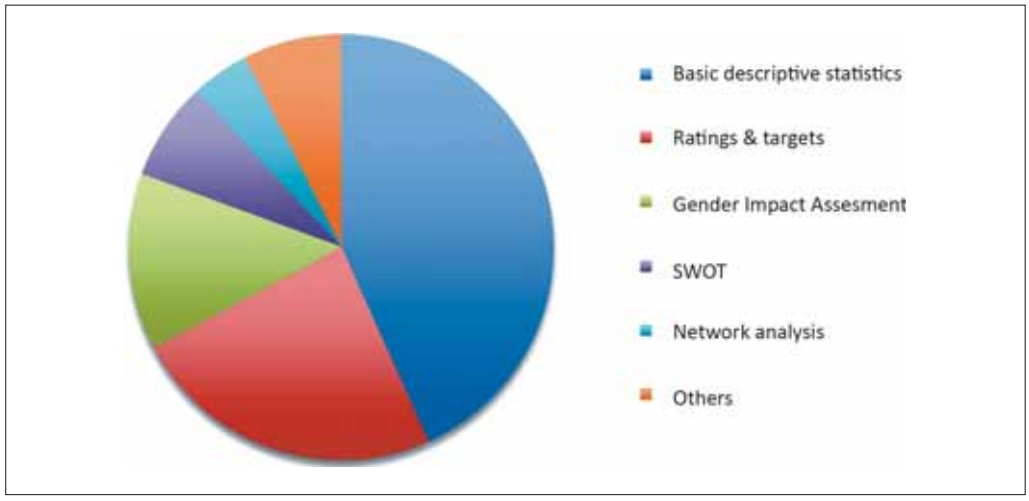
Programme	Method/Tool	Description
Poland: *ex post* evaluation of the PHARE-CBC 1994–99 Poland-Germany Programme *ex post* evaluation of Poland-Germany Programme	School questionnaire Review of local press	In order to assess the impact of the PHARE-CBC programme on the awareness of the residents of the border region, a school questionnaire was conducted in the so-called "divided cities" (that is, cities divided by a border,such as Zgorzelec and Goerlitz). The teachers from selected primary schools circulated the questionnaires among pupils, who gave them to their parents. The return rate was very high, and the results were very much in line with the results of a mirror survey of a representative group of Poles living in the border region. In addition to school surveys, a review of Polish and German local presswas conducted. This second method investigated when (if they were made) the context in which references to the PHARE programme were made.
UK: South Yorkshire Objective 1 Programme Mid-Term Evaluation	Social surveys Workshops	For evaluation purposes, survey questionnaires involving 160 programme stakeholders in the subregion were conducted with NGOs, educational institutions, local authorities and business associations. They were asked questions concerning identification and evaluation of visible programme effects. Workshops were organised in three waves. The first included workshops with three strategic and thematic programme groups, with the aim of identifying key programme areas for corrections or adjustments, based on tentative conclusions. The second involved meetings with each of the Strategic Local Partnerships which had implemented the projects. Here, the programme's implications for specific local areas were discussed. The third wave of workshops dealt with horizontal issues, with discussions from experts supported by representatives of specific regions.
Poland: PHARE SSG 2003 – The regional component	Interviews	As part of the exercise, interviews were conducted with a multiple groups of final programme beneficiaries, in order to understand the mechanisms underlying their success as well as the motives which prompted them to take part in the programme.
Sweden: On-going evaluation "EU Structural Funds – tools for regional development"	Case studies	This is an example of an interesting, unconventional choice of case studies, which were selected on the basis of an analysis of statistical data. Two municipalities with similar development conditions but different effects of Structural Funds programmes were selected for the exercise. In all, 17 in-depth interviews were conducted.

Source: prepared by the author on the basis of the project's database and evaluation reports.

3.4. Methods used for analysis and assessment

In the subsequent part of the study, we looked at the methods used at analysis and evaluative judgement. The most popular methods are detailed in Figure 9 below. The only group that made a positive difference were the Swedish reports, which employed multi-criteria and input-output analyses. It should be noted, however, that some of the reports provided only a simple calculation of project costs and their outputs as the input-output analysis, which is far from the original procedure. Also, and quite surprisingly, tools for spatial analysis were extremely rare even though the assessed programmes clearly dealt with spatial phenomena. The lack of GIS may be explained by high costs, technological barriers or the time required to input the data, but the lack of ordinary maps is absolutely astonishing! The few maps that were included were merely supposed to provide the background – an illustration serving as an introduction ofthe main analysis. The items listed under the (quite numerous) heading "other" call for some kind of clarification. Its structure is shown in Figure 8.

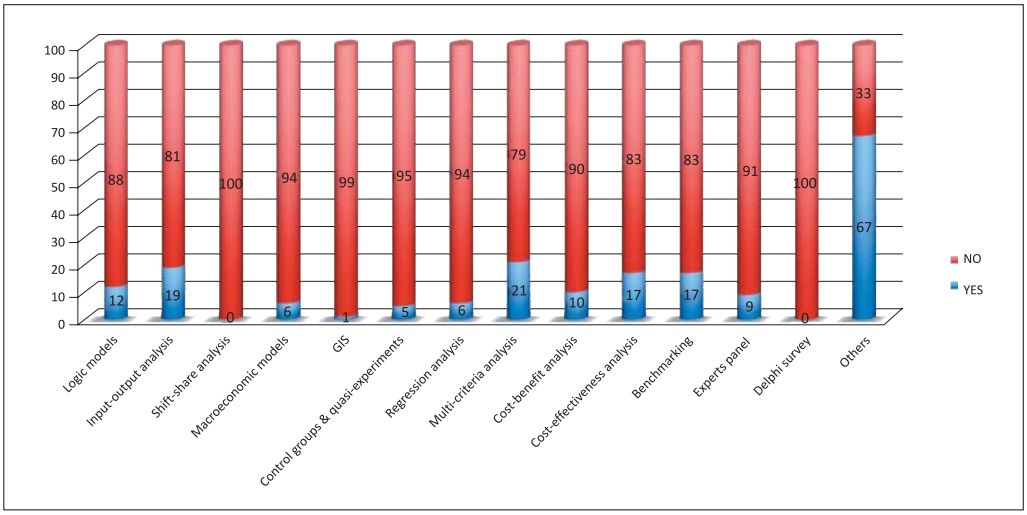
Figure 8.Analysis and judgement methods labelled as "others"



Source: prepared by the author on the basis of the project's database.

As we can see, a wide majority of evaluation studies were based on ordinary, descriptive statistics showing socio-economic trends in a given field (the sector of the programme's operation), but without linking them to the effects of the programme, specification of expenditure and outputs. What is exceptionally disappointing is that very few attempts were made to look into the causal relationships or the lack of unconventional methodological solutions. Their overview is provided in Table 6.

Figure 9. Methods used in data analysis and evaluative judgement



Source: prepared by the author on the basis of the project's database.

Table 6. Examples of innovative analysis and evaluation methods

Programme	Method/Tool	Description
Ireland: The Programme for Research in Third-Level Institutions (PRTLI) – Impact Assessment (International Assessment Committee, 2004)	Analysis of quotations and peer review evaluation	Bibliometric analyses were conducted, measuring the frequency of quotations regarding publications by staff affiliated with relevant institutions (beneficiaries of the programme). Additionally, international committee members evaluated 10 publications in terms of the recognition (awareness) of a given research centre across the world.
Sweden: Structural Funds, Entrepreneurship and Situation of Enterprises. Lessons from Structural Funds Programmes 2000–2006	Typologies	A typology of enterprises was developed for companies participating in the programme, based on their lifetime stages (survival, dynamic growth and maturity). In addition, a typology of ideal public assistance programmes for each of those stages was prepared (respectively: consulting, financing, networking, etc.). Later, individual companies were categorised into groups according to the statistics, In this way, the correspondence of the programme vis-ŕ-vis real needs was tested.
Sweden: EU Structural Funds – Tools for Regional Development	Comparisons	The "difference-in-difference" method was applied: municipalities with SF projects were compared with municipalities where no such projects were delivered; time series of statistical data were investigated, and discussion panels were organised.
Ireland: Evaluation of Agency Supports for R&D Performed in the Business Sector (Evaltec, 2004)	Specifications	A list of 300 enterprises classified as "top R&D" (150 from Ireland and 150 foreign-owned) was verified in terms of their participation in programmes promoting R&D.

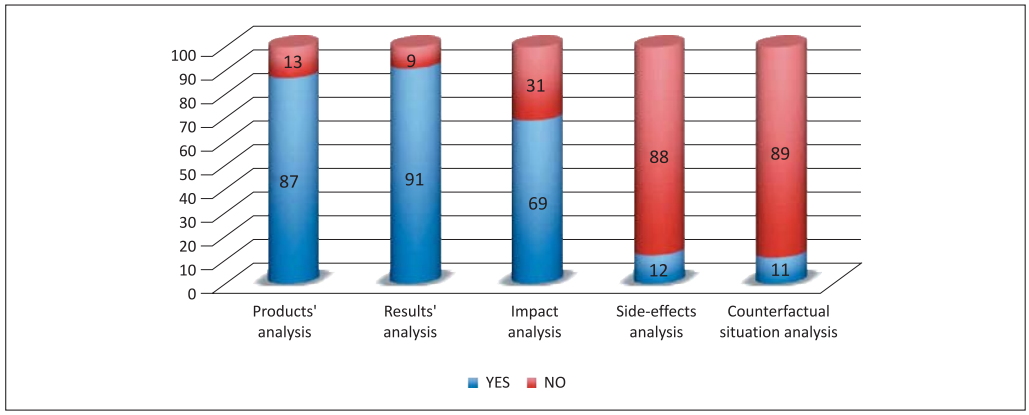
Source: prepared by the author on the basis of the project's database and evaluation reports

3.5. Orientation of findings and recommendations from the evaluation

The final component of our analysis of evaluation reports was an overall view on the content of their conclusions and recommendations.

Below you see a list of the objects of analyses based on the nomenclature used in individual reports.

Figure 10. Objects of study based on nomenclature used in reports

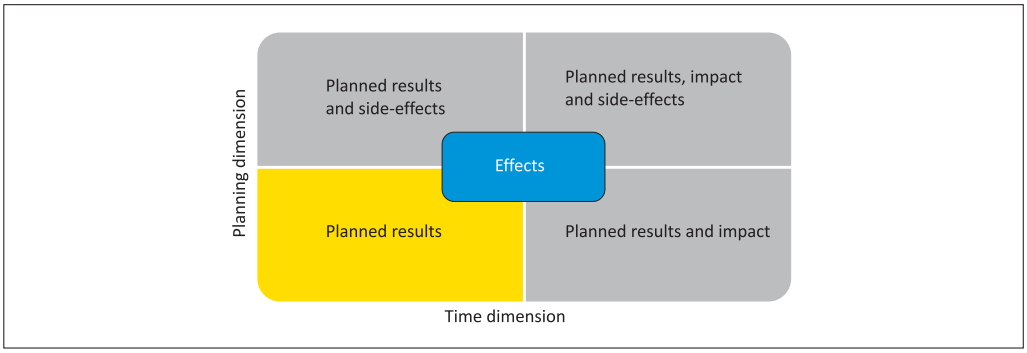


Source: prepared by the author on the basis of the project's database.

Interestingly, studies of results prevail over studies of outputs. This is partly due to the fact that the analysis in several cases began at the level of results (and the programme's outputs were merely listed). However, this predominance of results (large number of impact analyses) is superficial. In reality, what the evaluators referred to as results or impacts, after carefully reading and comparing the relevant definitions, turned out to be outputs, or at best short-term results. The reason for this was that the number of projects was defined as an output. Consequently, an analysis of outputs was an ordinary list of the statistics of completed projects, whereas their outputs (such as the number of persons trained or the infrastructure completed) were interpreted as results. With such an approach, the majority of evaluation studies focused on descriptive statistics, thereby avoiding the issue of actual effects. This is confirmed by rare cases when a situation without intervention, and in some cases potential side effects, were subject to analysis (an exercise which was usually based on a straightforward extrapolation of trends).

When we relate this analysis of reports orientation to the two dimensions of effects discussed in the first part of the chapter, we can conclude that the evaluation studies under analysis belong to the bottom left corner of the diagram below. However, if we were to use a stricter definition of results and impacts, these studies would not fit into the diagram at all; that is, they could not be regarded as studies of effects.

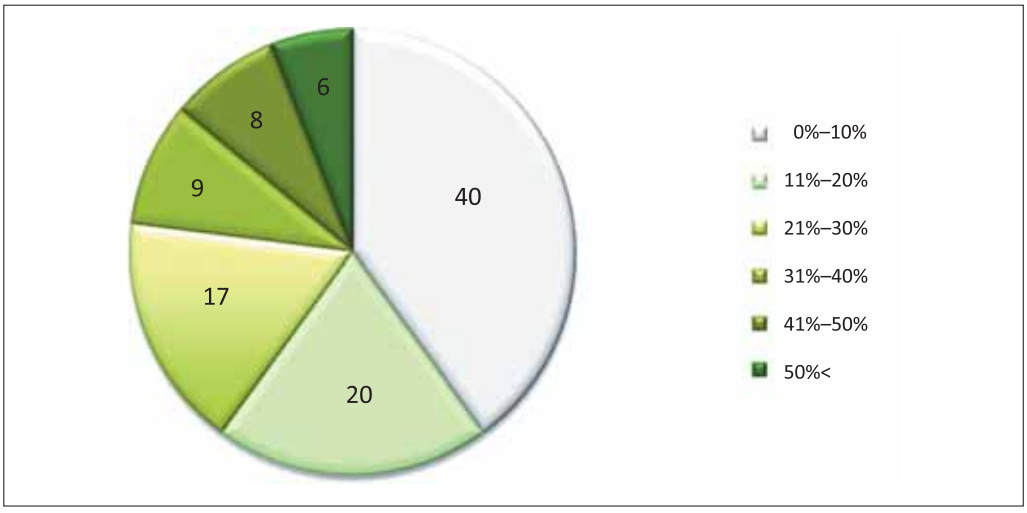
Figure 11. Two dimensions of intervention effects



Source: prepared by the author

Another source of the verification of this phenomenon is the analysis of the orientation of the conclusions and recommendations. Figure 12 shows the number of reports by the percentage of their conclusions dealing with the effects of a given intervention. Only six reports devoted more than 50% of their conclusions and recommendations to the issue of effects, whereas an overwhelming majority discussed effects only in one of ten of their conclusions. The diagram shows very clearly the predominance of process-related and on-going issues over strategic issues. Therefore, it can be concluded that the bulk of the studies were process-oriented evaluations.

Figure 12. Number of reports by the percentage of conclusions dedicated to effects



Source: prepared by the author.

Last but not least, we observed that the evaluators referred to theory and broader development concepts only in four cases, which confirms the extremely (and narrowly) technical nature of the reviewed evaluations.

4. Conclusions and Recommendations

In this final section of the chapter we first present the main conclusions of the research, then discuss the possible explanations of the observed phenomena and finally put forward a set of suggestions for future improvements.

There are three main conclusions we can draw from the executed review of 100 evaluation studies.

Firstly, when we relate the studies in question to the definition of methodology outlined in the first part of the chapter, we see that vast majority of them were lacking methodology understood in a holistic way (methodology = the research concept + a set of tools and methods)! These evaluations were characterised by an absence of a coherent research approach that is grounded in the literature of the subject. Moreover, with the exception of few cases (usually model-based evaluations), the reports did not contain any broader references or thorough, convincing justification for the methodological choices.

Secondly, the proposed set of methods and tools in fact was limited to a few basic components: survey questionnaires of the beneficiaries (but not the final beneficiaries), interviews with programme managers, general statistics, desk research and monitoring data. Analytical methods were mostly limited to descriptive statistical data.

The methodological shortcomings, and failure to provide any sensible justification and explanation of the choices made by the authors, are particularly unsettling as they undermine the credibility of these studies. Both theory and empirical studies clearly indicate that evaluation methodology (its transparency, coherence and quality) is the critical factor that determines the quality and reliability of the findings and recommendations, and thereby their use in the public sector (Cousins, Leithwood, 1986; Ferry, Olejniczak, 2008; Weiss, Bucuvalas, 1980).

These first two observations are reflected in the third conclusion – the limited nature of conclusions coming from the analysed evaluation studies. One could pose a rhetorical question: what can be said about the effects of complex public interventions without a coherent research strategy and clearly justified methods? Not much, it seems. This is confirmed by the striking focus of reviewed research on on-going issues, processes and progress in spending funds. If some information about outputs or results appears, theyare purely descriptive in nature. The assessments are limited to the discussion of programme objectives, while the degree of desirable social change has not been pondered.

To sum up, we can conclude that the analysed studies do not offer a reliable foundation for developing more extensive knowledge about the effects of Structural Funds, and thereby the mechanisms for effecting desirable socio-economic change. Instead, they proved to be a simple, bureaucratic exercise in describing ways of expending funds. As we can see, the sources of inspiration are rather uninspiring.

How can we explain the surprising methodological weakness of the reviewed evaluation studies of Structural Funds? A possible explanation of this situation seems to be related to five issues.

The first issue is a broader policy context and formal division of accountability. According to provisions laid down in the Structural Funds Regulations, *ex post* evaluation (i.e., comprehensive reflection on the effects of SF programmes) rests with the European Commission. In practice, the national administrations are rarely interested in real effects of EU programmes, as, firstly, they are not formally obliged to do so, and, secondly, the main, if not sole interest of both national politicians and the wider public is on fast absorption of EU funds. Thus, the measure of success by which the administration is assessed is the level of spending, not the quality of structural effects. As a result, the transfer of national evaluation practices to the area of funds is limited.The commissioned studies focus on what the national officials need the most – advice on improving the procedures and the flow of spending funds.

The second reason for the diagnosed situation has been a structural error regarding the scope of mid-term evaluations and their updates (which were the majority of the population of the analysed reports). The scope of tasks required by the Commission covered both an evaluation of on-going progress and assistance in enhancing the implementation processes as well as an evaluation of the effects. However, as empirical studies show (Eser, Nussmueller, 2006; Olejniczak, 2008), these two functions – operational and strategic – cannot be combined into a single study. When evaluators were forced to make this type of choice, they focused on offering advice on operational issues. This was seen as the more important issue from the point of view of the contracting authority. What is more, this was also a much simpler task and less politically sensitive than an evaluation of the effects.

The third reason seems to be the application of a narrowed definition of the evaluation of effects – one oriented exclusively on programme goals (the so-called objective-oriented approach). Virtually all of the analysed studies (except a very small number of those which looked at side effects) applied this narrow interpretation; that is,theychecked only to what extent the programme assumptions had been fulfilled (whether a given objective was achieved), without going into a demanding analysis of the relationships between the desirable social change and the actual effects of the activities.

The weakness of the methodology (understood holistically as both concept and methods) and a striking lack of reference to broader literature on the subject can be explained by the fourth reason – the dominance of the “consultancy” approach over a more scientific approach. Evaluators, who, for the majority, were employed by consulting companies (while academic or research institutions were involved in merely 20% of the analysed cases) normally assumed the role of consultants, as it was easier for them and did not require any specific knowledge or expertise in a given field. However, for the evaluation of effectsthe evaluators have to play the

role of teachers – mentors. This must be based on an in-depth knowledge, both theoretical and practical, of a given area of intervention (Ferry, Olejniczak, 2008, pp.74–75).

The final explanation for the deficiencies of the research methodologies can be the lack of literature on the methodologies of Structural Funds. Astoundingly, there are practically no publications of this kind in the European market – not even from renowned publishing houses. As previously indicated in the introduction, papers on the evaluation of EU programmes (and their methodologies) are very seldom published in professional periodicals. Naturally, this can be the result of feedback – since there are no academically-oriented studies, there are no publications having a research orientation; and lack of publications reinforces a simplified, consultancy-style approach to methodology.

The above findings call for some recommendations on how the present situation can be improved. We put forward few suggestions.

Referring to the first of the identified problems, we should emphasise the great value of *ex post* evaluations initiated by the administrations of Member States. Studies commissioned by the European Commission are conducted at the European level for the entire EU space and consequently have a high degree of generality. However, in order to take account of changes, to capture the causal relationships and explain the underlying mechanisms of the success or failure of a given programme, detailed empirical studies are needed, and they must be conducted on specific, clearly targeted populations or spatial systems. Such a level of analysis and detail could only be achieved by studies commissioned at the national or regional level. These do not have to be purely *ex post* evaluations. In our opinion, we should also support on-going evaluations focused on the effects assessment of completed groups of projects as part of programmes underway.

The second under lying problem – the structural error of combining strategic and operational functions in single mid-term evaluations – has already been resolved. The European Commission renounced its requirements concerning mid-term evaluation. Nevertheless, national administrations should be careful not to make the same mistake at the national level while commissioning individual studies for on-going evaluation.

Assistance in addressing the third problem – anarrowed definition of effects – can be found in the classic literature on evaluation. Back in the 1970s Michel Scriven introduced the concept of a “goal-free evaluation”. This is research in which evaluators examine the value of the programme by investigating what it has done and changed in the population or area, rather than tracing what objectives it tried to achieve (Scriven, 1976). Thanks to this, evaluation can avoid narrowing its scope to only the planned changes and is able to explore the real impacts as well as identify unexpected outcomes of the interventions. In practice, it means reversing the order of the investigation. First, evaluators trace the real impact of the funds and explore

the effects in the field. Later, they compare these findings with the objectives set in the programme documents. Finally, they investigate views of the programme staff and implementation procedures. This sequence puts the effects in the centre of the investigation, while the implementation process is only used as one of the explanatory factors (the others are characteristics of the beneficiaries and conditions of the environment). This solution is further reinforced by our next suggestion.

Our final set of suggestions addresses the last two issues identified – the weakness of methodologies and lack of literature devoted to the evaluation of Structural Funds. In our opinion, studies on Structural Funds should make reference to regional development theories. Financial interventions (and EU-funded programmes are certainly examples of such interventions) are components of development policies, thus they are based on certain (often not clearly expressed) assumptions and theoretical choices crucial to the development, strategies for ensuring economic growth or social progress. Taking such an approach could provide three advantages. Firstly, evaluation studies would be provided with a conceptual framework for their analyses, and a clear narration structure. Actual observed changes would be set against the programme's plans and interpreted in light of the regional development paradigms to the implementation of which they have contributed, whether knowingly or not. This would allow for going beyond narrow technical digressions and using the evaluation findings for informed, factual discussions on the direction and usefulness of public policies. Secondly, evaluation would certainly benefit by using, or relating to, extensive literature devoted to socio-economic development issues as well as analytical tools offered by this kind of literature. Likewise, the introduction of socio-economic theories, their terminological and research apparatus will also impose certain discipline and enhance analytical standards. This is clearly visible when we look at examples of evaluations from the United States, where standard concepts of validity and reliability (Bryman, 2004, pp.70–75) have migrated from social studies and are now firmly grounded in evaluation (Bamberger *et al*., 2006; GAO, 1998; Mathison, 2005, s.439–442, 370; Mohr, 1995). Thirdly, combining theory and practice could attract wider involvement from academics, and the results of Structural Funds evaluations will make an impact and be discussed in specialised literature.

As both practice and theory show, in the end it is the topic and quality of the evaluation research that determine its utility forpublic sector decision-making. The extensive energy and resources that go into evaluation exercises should have, and do have, the potential to provide an important contribution to evidence-based policy debates.

Bibliography

Bachtler, J., Poverari, L. & Taylor, S. (2000) "Culture, Commitment and Capacity. A Comparative Review of Structural Funds Evaluation in the European Union", *Edinburgh Conference on Evaluation*, Edinburgh, United Kingdom.

Bamberger, M., Rugh, J. & Mabry, L. (2006) *RealWorld evaluation: working under budget, time, data, and political constraints*. Thousand Oaks, Calif.; London: Sage.

Bryman, A. (2004) *Social research methods*. Oxford: Oxford University Press.

Council of the European Union (1999) Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds. *Official Journal of the European Communities*, L161, 26.6.1999, 1–45.

Cousins, B.J. &Leithwood, K.A. (1986) Current empirical research in evaluation utilization. *Review of Educational Research*, 56(3), 331–364.

CSES (2003) *Mid-term evaluation of the East Wales 2000–2006 Objective 2 & Transitional Programme – Final report*, Sevenoaks: Centre for Strategy and Evaluation Services.

EPRC & Fraser of Allander Institute (2000) *Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment. Final Report to Scotish Executive*, Glasgow: European Policies Research Centre, Fraser of Allander Institute.

Eser, T.W. &Nussmueller, E. (2006) Mid-term Evaluations of Community Initiatives under European Union Structural Funds: A process between accounting and common learning. *Regional Studies*, 40(2), pp. 249–258.

European Commission. (1995a) *MEANS Handbook 1 – Organising intermediate evaluation in the context of partnership*. Brussels: Commission of the European Communities.

European Commission. (1995b) *MEANS Handbook 2 – Identifying the structuring effects of community interventions*. Brussels: Commission of the European Communities.

European Commission. (1995c) *MEANS Handbook 3 – Measuring the employment effects of Community Structural Interventions*. Brussels: Commission of the European Communities.

European Commission. (1995d) *MEANS Handbook 4 – Applying the multi-criteria method to the evaluation of Structural Programmes*. Brussels: Commission of the European Communities.

European Commission. (1995e) MEANS *Handbook 5 – Quality assessment of evaluation reports: A framework*. Brussels: Commission of the European Communities.

European Commission. (1995f) *MEANS Handbook 6 – The ex post evaluation of Structural Funds Programmes: a methodological framework*. Commission of the European Communities, Brussels.

European Commission. (1995g) *MEANS Handbook 7 – The ex ante evaluation of Structural Funds Programmes: a methodological framework*. Brussels: Commission of the European Communities.

European Commission. (1997) *Evaluating EU expenditure programmes: A guide for ex post and intermediate evaluation.* Brussels: Office for Official Publications of the European Communities.

European Commission. (1999) *MEANS collection: evaluating socio-economic programmes*. Luxembourg: DG-Regio, Office for Official Publications of the European Communities.

European Commission (2000) *Methodological Working Paper 3 – Indicators for Monitoring and Evaluation: An indicative methodology*, The New Programming Period 2000–2006 Methodological Working Papers; Brussels: European Commission – DG-Regio.

European Commission (2002a) *Methodological Working Paper 8 – The Mid Term Evaluation of Structural Fund Interventions*, The New Programming Period 2000–2006 Methodological Working Papers; Brussels: European Commission – DG-Regio.

European Commission (2002b) *Methodological Working Paper 8a – The Mid Term Evaluation of Structural Fund Interventions – INTERREG III*, The New Programming Period 2000–2006 Methodological Working Papers; Brussels: European Commission – DG-Regio.

European Commission (2006) *Working Document no 2 – Indicative Guidelines on Evaluation Methods: Monitoring and Evaluation Indicators*, The New Programming Period 2007–2013 Working Documents; Brussels: European Commission – DG-Regio.

European Commission (2007) Evalsed– A resource for the evaluation of socio-economic development. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm, accessed: 10.10.2008.

European Council (1988) Council Regulation (EEC) No 2052/88 of 24 June 1988 on the tasks of Structural Funds and their effectiveness, and on the coordination of activities between themselves and with the operations of the European Investment Bank and other existing financial instruments. *Official Journal of the European Communities*, L185, 15.7.88, pp.0009–0020.

Evaltec (2004) *Evaluation of Agency Supports for R&D Performed in the Business Sector*, Dublin: Enterprise Ireland.

Ferry, M. & Olejniczak, K. (2008) *The use of evaluation in the management of EU programmes in Poland*, Warsaw: Ernst & Young – "Sprawne Państwo" Program.

Furubo, J.-E., Rist, R.C. & Sandahl, R. (2002) *International atlas of evaluation*. New Brunswick, U.S.A.: Transaction Publishers.

GAO (1998) *Approach and Methodology Overview – Instructor Manual*, Washington D.C.: Government Accountability Office – Training Institute.

GHK Consultants. (2006) *The Evaluation of Socio-Economic Development. The Evalsed Guide (2nd edition)*. London: GHK Consultants.

Howell, E. &Yemane, A. (2006) An assessment of evaluation designs. Case studies of 12 large federal evaluations.*American Journal of Evaluation*, 27(2), 219-236.

International Assessment Committee (2004) *The programme for Research in Third Level Institutions (PRTLI) – Impact assessment*, Dublin: Higher Education Authority.

Kierzkowski, T. (2002) Evaluation of socio-economic programmes in Poland – the current situation and future challenges. *Annual Conference of Regional Studies Association Evaluation and EU Regional Policy.*

Mark, M.M., Henry, G.T. &Julnes, G. (2000) *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.

Mathison, S. (2005) (red.) *Encyclopaedia of evaluation Sage*, Thousand Oaks, Calif.; London.

Mohr, L.B. (1995) *Impact analysis for program evaluation*. Thousand Oaks, Calif.; London: Sage.

Olejniczak, K. (2007) "Ewaluacja jako narzędzie zarządzania w sektorze publicznym"; w: Gorzelak, G. & Tucholska, A. (red.) *Rozwój, region, przestrzeń*, s.361-396. Warszawa: Ministerstwo Rozwoju Regionalnego.

Olejniczak, K. (2008) *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*. Warszawa: Wydawnictwo Naukowe Scholar: www.euroreg.uw.edu.pl.

PENTOR (2008) *Ewaluacja ex post regionalnegokomponentu MSP PHARE SSG 2003*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Scriven, M. (1976) "Pros and cons of goal-free evaluation"; w: Glass, G.V. (red.) *Evaluation studies review annual vol 1*, Bevery Hills: CA: Sage.

Scriven, M. (1991) *Evaluation thesaurus*. Newbury Park; London: Sage.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2001) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Stern, E. (2005) "What do we know about the utilization of evaluation?", *I konferencja ewaluacyjna – Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych*, Ministerstwo Gospodarki i Pracy & Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 16 September 2005.

Stufflebeam, D.L. &Shinkfield, A.J. (2007) *Evaluation theory, models, and applications*. San Francisco, Calif.: Jossey-Bass; Chichester: John Wiley [distributor].

Szlachta, J. (2006) "Polskie doświadczenia w zakresie ewaluacji", *II Konferencja Ewaluacyjna – Ewaluacja ex-ante jako narzędzie doskonalenia zarządzania strategicznego w administracji publicznej*, Ministerstwo Rozwoju Regionalnego & Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 22 November 2006.

Szlachta, J. (2007) "Wprowadzenie"; in: Haber, A. (ed.) *Ewaluacja ex post. Teoria i praktyka badawcza*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Tavistock Institute, GHK Consultants & IRS. (2003) *The Evaluation of Socio-Economic Development. The Evalsed Guide (1st edition)*. London: Tavistock Institute.

Toulemonde, J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *The Danish Evaluation Society Annual Conference*.

Weiss, C.H. &Bucuvalas, M.J. (1980) *Social science research and decision-making*. New York; Guildford: Columbia University Press.

Evaluating the effects of regional policy – the Experience of the European Commission from the Recent *ex post* Evaluation

Veronica Gaffey

Background

Under the rules laid down by the European Council for Cohesion Policy, the European Commission has responsibility for the *ex post* evaluation of each programming period in collaboration with the Member States, while Member States and regions are responsible for *ex ante* and mid term or ongoing evaluation. Member States and regions are responsible for designing and implementing programmes, but the Commission is responsible to the European Parliament, the European Court of Auditors and, ultimately, the public at large for the expenditure of EU taxpayers' money under the policy.

The Directorate General (DG) for Regional Policy has completed its third *ex post* evaluation with a multi-annual focus covering a large number of countries and programmes. This chapter briefly outlines the history of the previous two evaluation exercises, presents the approach to *ex post* evaluation for 2000–2006 and reflects on the use of its results.

In the background of this evaluation was the review of the EU budget for the post 2013 period. Under this review, all areas of expenditure are being scrutinized. Cohesion Policy is subject to particular examination, given that it absorbs 35.7% of the entire EU budget and 0.38% of the total GDP of the EU. In this context, it is no longer sufficient to have an efficient policy which ensures that resources are spent – which has often been the focus in the past. We need to demonstrate that it has had an impact and one that is greater than what would have happened in its absence. If there are elements of the policy which cannot demonstrate positive effects, these are likely to be challenged and may be dropped for the future.

This chapter is written from the perspective of the Evaluation Unit in the Directorate General (DG) for Regional Policy of the European Commission. The unit manages evaluation processes for the DG, which includes designing evaluations, managing calls for tenders, selecting contractors, steering evaluations, disseminating results and facilitating policy debate. The unit also provides guidance on evaluation to Member States and regions and facilitates networking and exchanges of experience among those in the Member States responsible for evaluation. This chapter focuses

on the experience of the *ex post* evaluation of the European Regional Development Fund (ERDF), 2000–2006. Other *ex post* evaluations have been launched by DGs Employment and Agriculture and also by DG Regional Policy for the Community Initiatives, Interreg and Urban, the Cohesion Fund and the Instrument for Pre-Accession (ISPA).

What were the particular challenges?

If we take the ERDF alone, how can we evaluate the effects of 230 programmes, implemented in 25 Member States, each of these programmes covering a wide variety of activities and ranging in size from €6.4 billion ERDF (Andalucía) to €6.4 million (Hamburg)? The total amount of resources (EU and national) invested through these programmes was €258 billion.

Think of the further complexity, when we include evaluations of 150 programmes under the Urban and Interreg Community Initiatives, involving a further €6 billion ERDF, not to mention the Cohesion Fund/ISPA providing support of €33 billion to over 1,000 transport and environmental projects in 17 countries.

The traditional approach to Structural Funds evaluation has been the logical framework, which structures programmes on the basis of inputs, outputs, results and impacts. However, experience has demonstrated that this framework cannot be applied simply at the EU level. While we can to a certain extent aggregate inputs and outputs to an EU level, this is not possible for results and impacts, since each programme is rooted in its regional or national context and aims to achieve effects in that context. Therefore, we are trying to evaluate the entire policy, recognising that we cannot aggregate all effects. At the end of this evaluation process we need to reflect how well this objective has been achieved and if there are better ways to evaluate the policy[1)].

So, what were the particular challenges confronting the evaluation of the ERDF under Objectives 1 and 2 in the 2000–2006 programming period? Many were the same challenges which faced previous evaluations, but – as will be explained below – we gained through experience a greater understanding of the nature of those challenges and have tried new approaches to tackle them:

- **Scale and Diversity** clearly presented challenges. There was an enormous diversity in programmes, in terms of size, as outlined above; content, with interventions ranging from the building of transport, environmental, education and health infrastructure, to supporting SMEs and larger enterprises, to promoting local, urban or rural development; and context, regions lagging behind, regions facing structural change, EU15, new Mem-

[1] This reflection has started. See DG Regional Policy's Ideas and Concepts Paper for monitoring and evaluation 2013+ at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/tech_en.htm and a methodological paper on identifying outcome indicators by Fabrizio Barca and Philip McCann assisted by a group of experts at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/performance_en.htm

ber States, national and sectoral programmes. A particular challenge was that the 2000–2006 programming period included the full 7 years for the EU15 (10 years including the extended period for expenditure) and only the shorter 2004–2006 period for the EU10 after their accession (5 years in total including the expenditure extension). For the EU10 this period was characterized by learning how to implement the policy including establishing systems for project selection and monitoring, auditing and reporting.

- **Multi-level governance** is at the heart of Cohesion Policy. This means that regions, national ministries and the partnership within Member States all have an influence in designing and implementing programmes. In order to understand programme effects, it is necessary to understand the dynamics within the programmes which involve a wide range of actors and factors.
- **Data Availability** continues to be problematic and is related to the previous point. The regulations governing the ERDF do not impose a common set of indicators. The literature suggests that this is the correct approach and that monitoring indicators are most appropriately defined at the programme level with those who will deliver outputs and results having a say in the indicators and their targets[2)]. This assumes, however, a capacity to define appropriate indicators and targets, which is still not fully developed across the EU. This inevitably creates a challenge for evaluation at the EU level, as even when similar indicators are used across countries, the definitions may be different and the target setting may be on a different basis. There was no centralized database of indicators and reports on achievements – up to 2007–2013, this was a paper based system.
- **Methods** are not straightforward. There is no one method which can successfully evaluate all types of intervention across all contexts. Therefore the mix of methods used in evaluation remains crucial.

DG Regional Policy started planning this evaluation early in 2007, with an intensive series of debates within the DG to define the evaluation questions and the terms of reference. The first contracts with the evaluators were signed in October 2007 and all reports were finally published in early 2010. The evaluations of the Urban and Interreg Community Initiatives were published in summer 2010 with the Cohesion Fund evaluations to be completed later in 2011.

Past experience

For the 1989 to 1993 period, *ex post* evaluations were undertaken of Objectives 1 and 2. The Objective 2 evaluation focused on jobs created, but the quality of data was poor. Extrapolations were made from one context to another and the transition from gross to net jobs was based on a small survey of enterprises. For Objective 1,

[2] DG Regional Policy will propose common output indicators to be used where relevant in the post 2013 funding period, building on the experience of reporting on core indicators in 2007–2013.

there were even greater difficulties with data. In general, the approach tended to be descriptive and also to focus on relatively small samples.

The 1994 to 1999 *ex post* evaluations were more systematic, with evaluations for Objectives 1, 2, Community Initiatives and the Cohesion Fund. The Objectives 1 and 2 evaluations involved two evaluation contracts, one for each Objective, with each contract producing national reports and a synthesis.

The European Court of Auditors carried out a performance audit of the evaluations of Objectives 1 and 2 and was critical of data, the macro-economic model and what they found to be insufficient oversight by the Commission. It acknowledged, however, that useful insights were obtained.

The Commission's own view of this evaluation process was that, while some of the criticisms were justified, the results of the exercise were stronger than those of the past. The main weaknesses were data availability (including on actual expenditure to improve the modelling, as well as data on outputs and results), evaluation design (too many evaluation questions, while the approach of evaluating nationally and then attempting to synthesize at EU level did not work), absence of rigorous methods (linked to both the above) and, therefore, over-reliance on descriptive approaches.

2000–2006: A thematic approach

In 2007, as DG Regional Policy began to design the *ex post* evaluation of Objectives 1 and 2, 2000–2006, three decisions were taken:

1) to evaluate thematically and across Objectives 1 and 2;
2) to phase and interlink the evaluation work packages, ensuring that different work packages started when data would be available and early results would feed into subsequent work packages; and
3) to accept that not everything could be evaluated, but to be more focused in approach and ensure that the evaluation questions could be answered on the basis of available data and the resources provided for original research.

The evaluation was planned to take place over a 27 month period – from October 2007 to December 2009 (compared to one year for the previous *ex post* evaluations). The budget was also increased, from €1.7 million for 1994-1999 to €4.7 million. This increase should be considered in the context that the evaluation covered 25 Member States instead of 15 and an allocation of €123 billion ERDF instead of €70 billion. All work packages have been published[3].

Below is a brief description of the different evaluation work packages, highlighting the key evaluation questions, the methods used and some of the findings.

[3] http://ec.europa.eu/regional_policy/sources/docgener/evaluation/rado2_en.htm

Work Package 1: Co-Ordination and Synthesis

The first contract helped in the overall co-ordination of the exercise, including commenting on all the deliverables of the other work packages, with a view to arriving at an overall synthesis of the entire evaluation in 100 pages. In addition, this contract included the following tasks:

- To review and describe the macro economic situation and regional development trends 2000–2006 and review the coherence of Cohesion Policy with international policy recommendations from the OECD, IMF and selected relevant thematic and empirical academic literature;
- To assemble financial information on allocations and payments, broken down by area of intervention for further analysis in other work packages;
- To present summaries for each of the 25 Member States outlining the stage of development of regional policy and outcomes of Cohesion Policy programmes;
- To draw up a taxonomy of Objective 2 programmes.

The methods used included statistical analysis of Eurostat data, assembly and detailed analysis of administrative data from DG Regional Policy, and desk research within each Member State and of EU and other international documentation.

The final synthesis report was published in April 2010; it draws conclusions on the effectiveness of the policy and raises a number of important policy implications:

- How to deal with the multiplicity of cohesion policy objectives and how to be clearer on the relative priority attached to these objectives and related indicators;
- The need for concentration of expenditure;
- The need for more clarity on the spatial dimension of policy focus;
- The role of indicators for determining and assessing policy, especially indicators which currently do not exist to capture the social and territorial dimensions of the policy;
- The case of Objective 2; and
- A number of implications for different thematic areas of the policy and for its management and implementation.

As we will see at the end of this chapter, many of these implications have been taken up in the review of the EU Budget and the Conclusions of the 5th Cohesion Report.

Work Package 2: Data Feasibility

All physical indicators from 2006 annual reports for all programmes were entered into a database. This information was available to other work packages for analysis and interpretation, and to the wider public, as it is published on the DG Regional Policy website. Through this work package, the Commission significantly improved data availability compared to previous evaluations. At least this time, the consultants on other work packages knew what was available and did not have to spend time searching for information which did not exist.

Work Package 3: Macro Economic Modelling in Objective 1 regions

This work package was carried out in DGs Regional Policy and Economic and Financial Affairs. The DG for Regional Policy owns the HERMIN macro economic model and is investing in its development. The impact of actual expenditure up to the end of 2008 on physical infrastructure, human capital and productive investment on the overall economy (GDP, employment and labour productivity) was estimated. Results are available for Objective 1 regions in Italy and Germany, Greece, Portugal, Ireland and Spain as well as for the 10 new Member States. In addition, the DG for Economic and Financial Affairs ran its model QUEST for the evaluation. The results of both models were incorporated into the synthesis report.

This work package overcame the most serious criticism of the modelling in 1994-1999, which was unable to be run on the basis of actual expenditure, given the lack of available data on matching funds expenditure[4)].

Work Package 4: Structural Change and Globalisation

The main evaluation question under this work package was to ask if Objective 2 programmes supported regional structural change and enabled adaptation to globalisation. 12 programmes were studied in depth and the terms of reference required both a quantitative and qualitative assessment of their added value in relation to structural change and globalisation. A final element of the evaluation was to explore the extent to which the results of the 12 case studies could apply for all Objective 2 regions.

Work Package 5a: Transport

The transport evaluation focused on the contribution of the ERDF to the development of the EU transport system. 18 Member States were studied, with information gathered on the main financial and physical outcomes of the transport interventions supported by the ERDF compared these to targets. 15 regional case studies looked

[4] This improvement was already in place for the simulations run for the 4th Report on Economic and Social Cohesion.

in more detail at the transport strategies, the types of interventions supported, their rationale and intermodality. Two further case studies looked at road safety and two examined ERDF support to regional airports – particularly the demand analysis and the contribution to regional development.

€34 billion ERDF was invested in transport measures and findings suggested that the bulk of these resources were targeted at measures with a focus on the development of the regions rather than the trans-European transport networks, which were more the focus of the Cohesion Fund.

Work Package 5b: Environment and Climate Change

The fundamental issue of the work package dealing with the environment was to understand the contribution of the ERDF to the implementation of environmental strategies in the EU and at the same time to regional development. The first task was to establish an appropriate conceptual basis, to understand the theoretical foundations and limits for the contribution of environmental infrastructure investment to the development of regions.

The scope of the evaluation was 14 Member States. In addition to the usual collection, presentation and analysis of physical and financial indicators, 10 regional case studies analyzed differences of effects between regions and assessed if and how environmental investments contributed to the wider socio-economic development of the region. A further three case studies focused on waste prevention and the management of waste.

Three case studies examined how the ERDF can help to reduce the emission of green house gases and adaptation to climate change. While climate change was not a priority in 2000 when programmes were adopted, it is clearly now an important issue and the *ex post* evaluation provided an opportunity to examine the extent to which programmes have been able to adapt to meet new challenges.

Total investment in the environment in the EU in the 2000–2006 period was an estimated €104 billion, with the EU (ERDF and Cohesion Fund) contributing almost €30 billion. Little evidence was found to support claims that environment infrastructure investment leads to economic development, although it can contribute to improved attractiveness of regions. Questions arose in the evaluation on the attention paid to value for money considerations in this area of investment.

Work Package 6: Enterprise Environment and Innovation

This was the third largest work package of the evaluation (after the co-ordination and management work packages) reflecting the significant proportion of ERDF investment (36%) which goes to support enterprise and innovation. Unfortunately, the contract was terminated in summer 2009 because DG Regional Policy was not

satisfied with the quality of what was being produced and the time it was taking (see comments below on quality assurance). As a result, two further work packages were launched to gather results for the synthesis evaluation, but this caused a delay and limited the evidence available. The two new work packages examined:

- The results and impacts of the 30 programmes which spent the most on enterprise support (which comprises 60% of total enterprise support under Cohesion Policy); and
- The performance of supported and non-supported enterprises in Eastern Germany, using counterfactual methods. This introduced more rigour to the examination of deadweight, which is usually analysed using *ex post* surveys of enterprises, asking if investment would have gone ahead without support. The answer to this question is subject to bias. The results of this evaluation were able to demonstrate no deadweight in investment grants to enterprises.

Work Package 7: Gender and Demographic Change

While gender was a horizontal priority for Structural Funds programmes, demographic change was not high on the policy agenda in 2000. This changed, particularly with enlargement. The work package examined the impact of the ERDF in supporting gender equality and enabling adaptation to demographic change. Again, the approach was literature review to establish concepts and criteria for the selection of case studies, followed by 12 case studies of regions.

The evaluation found a potential for a virtuous link between demographic change and gender, but one that is not sufficiently exploited by the ERDF. Examples of good practice were found, e.g., in the ERDF co-financing childcare facilities. In general, however, gender issues were perceived as being "for the ESF". The evaluation concluded that there is a need for locally designed strategies to tackle the challenges of gender inequality and demographic change and that the ERDF, as well as the ESF, could play an important role in this regard.

Work Package 8: Transport Modelling

The Institute for Prospective Technological Studies at the Joint Research Centre in Seville has developed a sectoral model to simulate the impacts of transport projects. DG Regional Policy explored with the Institute the possibility of using this TRANS-tools model to assess the effects on regional GDP and other economic aggregates of over 100 transport projects which received support from the ERDF. The exploration found that the model is currently not suited to provide such estimates, as it can only work with entire transport corridors. Further development of the model will be explored when we can map all ERDF and Cohesion Fund transport projects.

Work Package 9: Rural Development

The 9th work package asked what is the nature and importance of the contribution of the ERDF to rural development. Based on a literature analysis, a typology of rural areas was developed. Information on the role of the ERDF in rural development in 5 Member States was gathered and analysed (France, Spain, Germany, Sweden and Poland) and case studies of one region in each Member State deepened the analysis.

The conclusion was that the ERDF invests considerably in rural areas, more per capita than in urban or intermediate areas, but that the driving force of the ERDF is the support of weaker regions. The evaluation found that the perception of people in rural areas was often that they did not receive support, often due to the location of decision making. In line with the findings of the evaluation on gender and demography, it argued for more local "ownership" of Structural Funds programmes and greater co-ordination between the different funds.

Work Package 10: Efficiency of Major Projects

Major projects are large infrastructures and productive investments (business support) which are subject to *ex ante* appraisal by the Commission services before EU funds are granted. This study established the actual unit costs of 96 projects and compared these to the plan and established benchmarks. Completion times were also analysed. Statistical analysis was anticipated in this work package, but the data available was not sufficiently strong or extensive to support such analysis. However, the conclusion of the evaluation was that establishing a database of unit costs is feasible and would represent a useful resource for those concerned with major infrastructure projects. This work is being built upon in the evaluations of the Cohesion Fund.

Work Package 11: Management and Implementation

The final work package focused on management and implementation systems for ERDF co-financed programmes. The evaluation concluded that:

- In the EU10 the systems to draw down the EU resources were put in place; in itself a very significant achievement. However, making them work well was difficult, especially in the beginning, due to weak administrative cultures and incomplete administrative reforms, unstable institutions, high fluctuations of staff, etc.. Systems tended to be too rigid and "compliance-oriented". Decisions were often driven by the fear to fail, in particular concerning financial control standards, and not by performance considerations.
- In the EU15, the evaluation found improvements in the strategic management of the policy, with evidence of spill-over effects on domestic govern-

ance systems. These effects were strongest when driven by committed decision-makers or when the status and weight of Cohesion Policy was significant. They were perceived less positively – as an administrative burden – in Member States with well functioning administrative systems.

- The main recommendation was to strengthen the performance focus of Cohesion Policy. Other recommendations concerned the fostering of leadership among Cohesion Policy managers, embedding a learning reflex in the policy, and supporting institutional capacity.

Methods Used in the Evaluation

Having presented the components of the evaluation and some of the findings and policy implications, it is evident that each work package had a particular focus which required a mix of methods. Our terms of reference provided guidance to the evaluators on what we sought. They move beyond the traditional approach of evaluating relevance, efficiency, effectiveness and utility across entire programmes to define more specific evaluation questions. Each work package required a specific mix of methods – designed to respond to the evaluation questions, the data available and the data required to answer the questions. Most of the thematic evaluations had a number of steps:

1. Establishing the conceptual basis for the evaluation: Too often evaluations jump straight into the collection of data, without reflecting on what we want to know. It is important to reflect on the theoretical foundations for support in a particular area, including market failure and theories of change. This should influence the choice of hypotheses to test in case studies.
2. Review of administrative data – financial and physical: We invested significantly in gathering available data and making this available to the consultants. Therefore, this task was usually comprehensive across all programmes and Member States. Aggregate information on inputs and outputs was analysed. This is an essential element of the evaluations, responding to our accountability obligations.
3. Case Studies: To go more in-depth on particular issues, we cannot examine all programmes; hence the case study approach. The rationale for the selection of case studies is important here in order to achieve some representativity, even if it is difficult to extrapolate from a few examples to the EU as a whole. The methods used within the case studies varied from contract to contract, but include more detailed document research and analysis, stakeholder interviews, focus groups, surveys, project visits, etc..
4. Conclusions and Recommendations: The final challenge for the evaluations is to pull together all the analysis into conclusions which answer the questions posed by each set of terms of reference.

The thematic work was complemented by the modelling and the economic analysis in work packages 1, 3 and 8. We also tested the use of counterfactual impact evaluation in enterprise support. There was a balance between the micro and macro approach, the bottom-up and top-down, the qualitative and the quantitative, all aiming at triangulation of evidence.

Steering and Quality Assurance

One of the objectives of the thematic approach was to appeal to the strengths of different evaluation consultancies in the market. Rather than expecting one consultancy company to master all themes and all methods, by focusing thematically we hoped to broaden the pool of evaluators involved. This also meant that we had higher expectations concerning the expertise evaluators brought to our contracts.

The quality assurance of the evaluation was made up of a number of elements:

1. First was the ongoing day to **day management** of the evaluation process, with each work package being the responsibility of a member of staff of the Evaluation Unit in the DG. The Evaluation Unit has a Quality Management Strategy, which defines the procedures to be followed at the various stages of an evaluation from preparation (the decision to evaluate and the preparation and publication of call for tenders, as well as the selection process), through the project milestones (kick off meetings, inception, interim and final reports) and general management issues.
2. Second was the **Steering Group** made up of colleagues from within DG Regional Policy and other interested DGs (Employment, Agriculture, Economic and Financial Affairs, Environment, Transport, Enterprise, etc.) as well as European Agencies, such as the European Environmental Agency and the European Foundation for Living and Working Conditions. One steering group for all the work packages facilitated exchanges of information and knowledge between contractors. Following each Steering Group meeting, the Evaluation Unit wrote formally to the contractor outlining the additional work required before the deliverable concerned could be approved (usually linked to a payment).
3. A third element was the **expert committees** for each work package, which met to discuss methods and interim and final deliverables. These experts were usually academic or other experts unconnected with the contractor. Their views helped to ensure that the knowledge of the academic community was taken on board. Altogether, 27 experts were involved in 9 scientific committees.
4. A final element was the **publication of interim deliverables and dissemination** events. Evaluators were challenged in public debates during the evaluation process – contributing to ensure quality. For final reports, the Commission also published its quality assessment of the product.

Of course, all the steps outlined required hard work and consistent follow-through, both from the Commission and the contractors. We were prepared to take hard decisions and on two of the work packages, contracts were terminated because we were not satisfied with the progress being made. Overall, however, we are satisfied with the quality of what has been produced.

What are the Strengths of the Evaluation?

What did we learn from this evaluation process? Some aspects of the process proved to have been particular strengths:

- More precise terms of reference were positive and helped the evaluators to focus their work and the European Commission to manage the process. In particular, targeting the evaluations at particular evaluation questions, rather than listing the usual exhaustive evaluation criteria of relevance, efficiency, effectiveness and utility, has proved effective. Our conclusion is that it is not possible to conclude on all these criteria across such scale and diversity as is represented in Cohesion Policy. So, it is better to recognise this and focus on what evaluations can realistically deliver.
- The variety of methods used provided extremely useful insights into how to advance evaluation practice in Cohesion Policy. The experience confirmed that there is no single or simple method which will deliver answers on effectiveness. Rather we try to build up a body of evidence using different methods and different evaluators.
- The involvement of academic experts also proved positive. We asked the experts to act as “critical friends”, to comment on and challenge methods, analysis and conclusions. Their contribution improved the quality of the evaluations.
- Finally, the enthusiasm, commitment and professionalism of some of our evaluators were definitely strengths of the evaluation process.

Where do we need to Improve?

Undoubtedly, we have also learned about what should be improved. We should continue to refine terms of reference. The *ex post* evaluations are a start, but we are not convinced that one evaluation will ever deliver all the answers. We must continue to build up a body of evidence and continue the process of debate on results.

A fundamental challenge for evaluation of Cohesion Policy continues to be the weakness of data. In theory, evaluators should be able to base much of their work on statistical and administrative data. In reality, only limited statistical data at NUTS3 level is available and usually with a time-lag. There are further complications with the comparability of NUTS3 regions across Member States. And a further challenge

is that statistical data reflects the effects of the policy and other factors. Although great effort has been invested by Member States and the Commission in developing monitoring systems, the sets of indicators, the targets and the gathering of data on performance – are still often weak. Evidence from the evaluations suggests that programme authorities still find it challenging to quantify objectives and indicators are added on to the programmes at the end, rather than being an integral part of programme design. Fewer and more relevant indicators of results are needed, based more clearly on a theory of change, with an idea of how and when to evaluate the effects of the policy, taking account of other factors. In addition, more use of standard output indicators is needed for accountability purposes.

As regards the quality of the evaluations produced, we were surprised at how challenging it is to turn data gathered into a narrative that links findings from the literature, statistics, data from monitoring systems and interviews with key actors on the ground together with the evaluator's judgements. To have an effect, evaluations need to tell a story and convince the audience with the evidence, analysis and conclusions. Too often too much energy is taken up with chasing data and not enough time is left for the crucial process of turning this into an evaluation. The last stage is an iterative process which needs time to hone and refine analysis and tease out conclusions and possible recommendations.

During the evaluation process, we became convinced of the added value of some more rigorous methods in some areas; for example, counterfactual impact evaluation of enterprise support or in the more recent past, *ex post* cost benefit analysis of major infrastructure projects. However, we are convinced also that we need more rigour in the use of qualitative methods[5)] and are exploring the provision of guidance and encouragement of more good practice in this regard.

Using the Results of the Evaluation

The *ex post* evaluations of 2000–2006 have been used intensively in the design of the future policy after 2013, while we also encourage reflection on the findings in order to improve the implementation of current programmes. The process is through internal discussions within the DG for Regional Policy complemented by presentations and discussions in the Member States.

After 2013, the policy directions adopted by the Commission so far indicate that the messages of the *ex post* evaluation have been heard and have complemented a greater awareness in the policy discourse that the results achieved are more important than being able to demonstrate only that the resources have been spent

[5] During 2011 we will publish a new section of the EVALSED guide on Theory Based Evaluation at http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/theory-based_impact_evaluation/introduction_to_tbie/index_en.htm and we are also exploring an Australian method – „performance story reporting".

correctly. In October 2010, the Commission's Communication on the EU Budget Review announced the following policy directions:

- Concentration on a menu of thematic priorities linked to EU2020; and
- Spending programmes with a real impact which can be measured – rather than focus on inputs.

While the Conclusions of the 5th Cohesion Report published the following month announced:

- The need to reinforce Strategic Programming;
- Conditionalities and Incentives to encourage a performance focus;
- Fixing outcome indicators *ex-ante*;
- *Ex-ante* evaluation to improve programme design; and
- Evaluation plans & more use of impact evaluations.

What is evident is a much stronger orientation towards achieving results. Currently, discussions are ongoing with Member States on how to make these concepts operational and within the Commission work is underway on drafting legal texts which will be adopted later in 2011 and will then be negotiated with Member States and the European Parliament.

Conclusions

The Commission should not on its own be responsible for demonstrating the value of Cohesion Policy. Member States and regions also have a responsibility to evaluate – in the interests of improving the quality of their programmes, but also for accountability. DG Regional Policy's objective – through the *ex post* evaluations – was to be accountable for the Cohesion Policy resources invested during the 2000–2006 period and to initiate a greater methodological rigour. This means being open about the limitations of the evaluation in that it could not cover all issues in depth. By focusing on key issues and testing more rigorous methodologies (particularly quantitative methods), we could change the nature of the debate towards a discussion based on evidence. Over time, and working in partnership with Member States undertaking more such evaluations which increasingly focus on the 2007–2013 period, we can learn more about what works in different contexts across the different fields of intervention.

What conclusions can we draw, based on our recent experience of evaluation at the Commission? The first is that it is not possible to evaluate everything from all angles at one point in time. Even at programme level, we found in the mid term evaluations of 2003 that a weakness was the requirement to evaluate at once effectiveness, efficiency, impact, community added value, management and implementation, etc., across large-scale programmes. The breadth militated against depth and the use of methods which would answer more specific questions related to particular

areas of intervention. Even more so at a European level, it is impossible to evaluate all themes for all programmes. Therefore, a selective approach must be taken. By trying out new methods in samples of programmes or regions, we hope to obtain interesting results and promote these methods for greater use by Member States and regions. The approach to ongoing evaluation of the 2007–2013 period provides greater scope than in the past for Member States and regions to include such methods in their evaluation plans.

A second conclusion is that the evaluation design needs to take account of data availability and be realistic. So the development of evaluation questions should reflect realism about what data are available and how they can be collected. This time the Commission provided to the consultants all programming documents, evaluations and recent annual reports. Two of the first work packages assembled physical and financial information in a readily accessible manner. This meant that evaluators focused their efforts in gathering additional relevant data and could start to evaluate sooner, rather than spending too much time in assembling basic information. That said, however, our experience was that sometimes the basic information was not actually available. We need to reflect on what data is absolutely necessary – reducing the overall volume – and agree between Member States and the Commission that this will be gathered.

A further conclusion is that as important as the methodology is the process of evaluation and a constant focus on high quality analysis and clear presentation of results. We set high standards both for our evaluators and ourselves. We read every deliverable closely ensuring that findings were based on evidence. The critical but constructive inputs of our colleagues within the Commission and the expert scientific committees also contributed.

However, fundamental to a shift towards more evidence based policy making, with evaluation playing an important role to provide such evidence, is a change in the mindset of those responsible for Cohesion Policy programmes. If those responsible for programmes regard “spending the money” as the success of the policy, rather than a precondition for success, we will not have evidence based policy making. Better programme design, including reflection on the choice of indicators, availability of baselines and setting appropriate targets, along with a concern to obtain evidence on programme performance, are essential for evaluation to fulfil its potential role as a support to policy makers to design and continuously improve Cohesion Policy programmes in a transparent and accountable manner.

The review of the Community budget requires us to ask fundamental questions about the performance of Cohesion Policy. The imperative to spend the money is no longer sufficient. What are the market failures which Cohesion Policy addresses in different regional contexts and across different intervention areas? What are the theories of change which influence the design of programmes and can be tested in evaluation? This new focus is challenging and may provide evidence which is

uncomfortable for some areas of the policy. But it is important that we continue to build up the evidence base about what works. If the evidence suggests that some elements of the policy are not effective, we – the Commission and the Member States – need to face up to that. In this context, evaluation has the potential to play a much stronger role than in the past.

Basics of Design for Evaluation of Cohesion Policy Interventions

Kathryn E. Newcomer

Introduction

There are many policies being implemented around the world involving multiple levels of governments that are intended to reduce disparities in the educational and employment opportunities across regions within nations. One of the most promising ways to build knowledge of the effectiveness of these policies and interventions is to undertake rigorous impact evaluations in different geographical regions, and

then share knowledge on what elements of interventions are more like to be effective under what sorts of conditions.

In the program evaluation field there are many different types of evaluation studies. Some focus on program processes and implementation, typically called formative evaluations, and others look at the effects or impacts of programs, typically called summative evaluations. While this chapter addresses the later type of evaluation, it is worth noting that there should be some assessment of program implementation in any high quality impact evaluation, as well.

This chapter addresses basic issues regarding design of impact evaluations of cohesion policy interventions. Advice is offered regarding measurement and design strategies, and on how to address challenges to drawing credible conclusions from evaluations. Specific issues addressed here are provided in four sections: (1) the current context for evaluating policy interventions; (2) designing evaluations to build a strong methodological base; (3) strengthening the credibility and supportability of inferences; and (4) crafting findings and recommendations that are credible.

(1) The Current Context for Evaluating Policy Interventions

There are a variety of policies promoted at the national and local levels of government intended to improve educational and employment opportunities, and the quality of economic development in depressed regions (Blom-Hansen and Jens, 2005;

Basle, 2006; Mairate, 2006; and Bradley, 2008). A variety of different strategies are employed, with differential levels of adherence to policy implementation. In fact, without rigorous evaluation and careful assessment of implementation, it is difficult to know which interventions may work if implemented properly. There are many contextual factors that affect success of cohesion interventions, for example, national and local government social policies, land use, zoning, and business incen-

tives, access to education, training and income support in communities, that great care must be given to designing adequately sensitive evaluations.

The recent policy evaluation environment has been dominated by proponents of evidence-based policy making. The Campbell Collaboration, a group of distinguished American and European social scientists, for instance, proposes that more experimental research be undertaken to identify consequences of social policy interventions. Relatedly, in the U.S. during the Bush administration, the Office of Management and Budget (OMB) provided guidance that randomized control trials (RCTs) were to be utilized to identify the effectiveness of federal programs, and the Bush administration significantly opened public dialogue about the need for systematic program assessment that is likely to continue in the Obama administration.

The George W. Bush administration introduced a new tool to "hold agencies accountable for accomplishing results" that focused on assessing program effectiveness (see U.S. Office of Management and Budget, 2004). The Program Assessment Rating Tool (PART) was a questionnaire consisting of approximately 30 questions (the number varied slightly depending on the type of program being evaluated) that federal managers were required to answer about program purpose and design, strategic planning, program management, and program results. The PART process pushed managers to draw conclusions about the effectiveness of their programs and substantiate them with evidence. It underscored the need for managers to report on how they assess the rigor of evaluation studies and apply them to inform program planning and corroborate program results.

OMB used the PART tool to assess almost about 98% of the federal government's programs between 2003 and 2008. Once a program was assessed, it was also re-assessed, albeit with less intensive scrutiny, in subsequent years, as well.

A notable aspect of the PART tool for program managers was its explicit focus on program results. A set of questions addressing program results were to be answered with "Yes," "Large Extent," "Small Extent," or "No." These questions were:

- Has the program demonstrated adequate progress in achieving its long-term performance goals?
- Does the program (including program partners) achieve its annual performance goals?
- Does the program demonstrate improved efficiencies and cost effectiveness in achieving program goals each year?
- Does the performance of this program compare favorably to other programs, including government, private, etc., with similar purpose and goals?
- Do independent and quality evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

OMB budget examiners assigned scores to programs based, in part, on the answers the program managers in the agencies gave to the PART questionnaire. Through the use of the PART process, OMB highlighted the issue of evaluating program effectiveness. While OMB examiners had raised questions about program effectiveness for many years, the PART tool presented a very explicit, systematic and transparent process to analyze programs across the federal government.

There have been many critiques of the U.S. federal government's use of the PART tool. The two biggest concerns about the use of the PART expressed by OMB examiners were a lack of time to conduct and review the assessments, and conflicting priorities conveyed to them about their responsibilities tied to the budget and their role in PART reviews. Program managers in the agencies also complained about the great demands in terms of time and resources needed to meet the demands of the review processes, and they were also concerned about the level of consistency in the application of criteria to programs across different budget examiners (Newcomer, 2006; Gilmour & Lewis, 2006; Moynihan, 2008; Metzenbaum, 2008; Newcomer and Redburn, 2008).

The message that experimental research (also called random control trials) is the best way to evaluate policies and programs was given throughout government during the Bush administration. But the problem was that there are many circumstances in which experimental protocols are simply inappropriate or impossible, e.g., water quality and foreign relations to name two, and there was inadequate guidance provided on what other evaluation methods could be used to assess program effectiveness. In fact, in March 2008 a task force of experts from the American Evaluation Association reviewed the OMB guidance on evaluation and concluded that "a more balanced and considered presentation of the role of RCTs in assessing the effectiveness of federal programs was needed," and stated that a "balanced presentation of the spectrum of appropriate, rigorous, evaluation methods will improve the likelihood of selecting appropriate measures and methods to assess and improve program performance." Throughout the U.S. and Europe, there is still a paucity of advice on how to evaluate complex public programs and policy interventions when RCTs are not feasible.

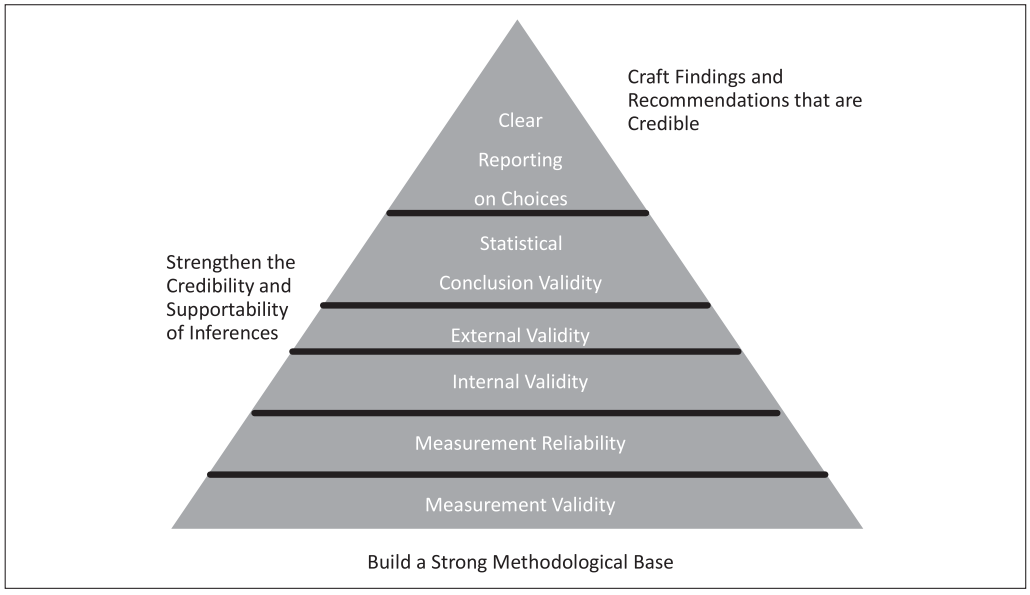
It is also typically very difficult to establish a causal link between policy interventions and behavior change. Numerous factors affect individual and firm behaviors, and trends in life style and mobility offered to citizens are changing faster than ever before. Further constraints on research are presented by inherent ethical prohibitions and logistical impossibilities that do not allow random assignment of subjects in evaluation, particularly when evaluating policies that affect broad swaths of the population.

Given this difficult environment, evaluators need to draw upon social science methods to bolster credibility of research; take advantage of existing wisdom about the art of evaluating complex policies and programs; and build knowledge in this rela-

tively new policy field through careful sharing and testing of methods and measures and practicing transparency in methodological decision making. The credibility of social policy evaluation work is dependent upon the methodological integrity of the work.

Rigorous application of social science methods to evaluate policy interventions in the field is an especially complex undertaking. There are several facets of the integrity of the evaluation enterprise that merit close attention. As shown in Figure 1, the strength of findings, conclusions and recommendations about the implementation and results of policy interventions depend upon well founded decisions made regarding measurement and design of the evaluations. The rest of this chapter addresses those decisions.

Figure 1: The Pyramid of Methodological Integrity



(2) Designing Evaluations to Build a Strong Methodological Base

Methodological integrity is affected by a variety of decisions made about measurement. Strong evaluation work requires clear, valid measures in order to collect evidence in a reliable, consistent fashion. Evaluators must begin with credible measures and strong procedures in place to ensure that measurements are consistent across time and space.

It is important to note that the professional standards and recommendations offered here apply to evaluation work whether it entails qualitative or quantitative research methods. The same criteria for assessing the methodological integrity of the work applies both when qualitative methods to collect data are used, e.g., interviews and focus groups, and when quantitative data are analyzed, e.g., govern-

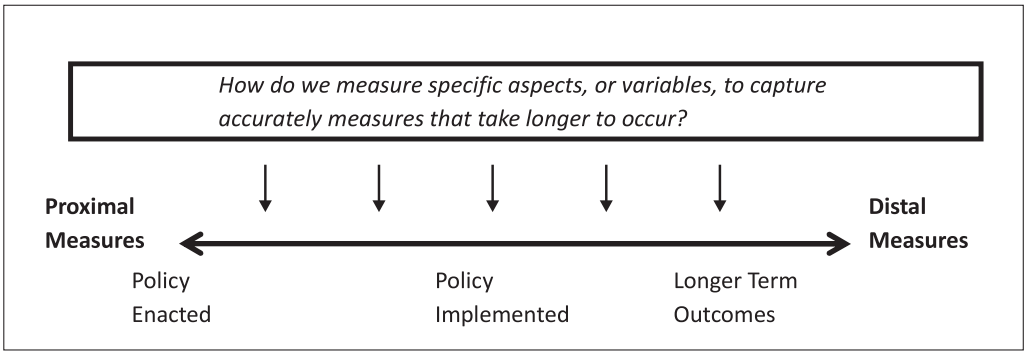
ment records (For more on strengthening evidence in qualitative research, see Mason, 2002). While in the U.S., there was a time when experts argued over whether quantitative or qualitative research methods were better, the prevalent view now is that the evaluation questions should indicate the appropriate research methods, and that multi-methods approaches in evaluation that entails both quantitative and qualitative research methods are probably superior to address the more complex evaluation questions (See Patton, 2008).

Some examples of policy measures include legislation language, presence of funding versus an unfunded mandate, and presence of enforcement strategies. Examples of intended cohesion policy outcomes may be found in employment and educational achievement data systems. Measures may include self reporting instruments, observations, geographic information systems data, and census data. Measures should be both valid and reliable.

Measurement Validity

Measurement validity is concerned with the accuracy of measurement, such that the measure accurately assesses what the evaluator intends to evaluate. Policy interventions are composed of complex sets of changes to the environment, and behavioral outcomes may include individual behaviors, as well as specific public agency and private firm behaviors, and may be viewed as falling on a continuum, with varying proximity from the policy implementation to more long- term outcomes, as seen in Figure 2. Evaluators should consider the timeframe available for evaluation as well as the time needed for specific outcomes to result from a policy intervention.

Figure 2: Measurement Validity and Reliability



Measures may be validated, that is, tested for their accuracy, through several different processes (For more on validity and reliability, see Shadish, Cook and Campbell, 2002; and Wholey, Hatry and Newcomer, 2004). For example, experts may be asked to comment on the **face validity** of the measures. Experts may be asked to respond to questions such as the following:

- Are the measures relevant to the concept being assessed?

- Do other experts in the field use these measures?
- Are newly constructed measures credible?

Experts may also be consulted to comment on the **content validity**, or breadth of coverage, of the measures used. For example: Do the measures selected adequately represent the potential pool of similar measures?

Credibility can also be bolstered through testing the behavior of measures that are selected after data are collected. For example, evaluators can address the following questions with the data:

- Do the measures correlate to a specific agreed-upon standard or criterion measure that is credible in the field?
- Do the measures behave (i.e., correlate) with other measures in ways consistent with existing theory and knowledge?
- Do the measures predict subsequent behaviors in ways consistent with existing theory and knowledge?

Measurement Reliability

In addition to being relevant, measures should be reliably recorded. Measurement reliability refers to the extent to which a measure can be expected to produce similar results on repeated observations of the same condition or event. Having reliable measures means that operations consistently measure the same phenomena and consistently record data with the same decision criteria. For example, evaluators should consider whether survey and interview questions are worded such that respondents of different cultural and or sociodemographic subgroups will respond in a similar fashion to the same question.

In order to strengthen reliability of measures and measurement procedures, evaluators should adequately pre-test instruments and then plan for quality control procedures when in the field. To enhance the measurement reliability of measurement instruments, evaluators should:

- Pre-test instruments with representative samples of intended respondents before going into the field;
- Implement adequate quality control procedures to identify inconsistencies in interpretation of words by respondents in surveys and interviews;
- When problems with the clarity of questions are uncovered, the questions should be revised, and evaluators should go back to re-survey or re-interview if the responses are vital.

Careful training and oversight of trained observers and interviewers is essential. Evaluators should:

- Adequately train observers and interviewers so that they consistently apply comparable criteria;

- Implement adequate and frequent quality control procedures to identify obstacles to consistent measurement in the field; and
- Test coders by asking them to all code a sample of the materials to test levels of consistency.

There are statistical tests that can be used to test for inter-coder and inter-observer reliability, such as Cronbach's Alpha. When statistical tests are desired, research texts or websites should be consulted (For example, see Shadish, Cook and Campbell, 2002).

(3) Strengthening the Credibility and Supportability of Inferences

In order to test the effectiveness of policy interventions, researchers must ensure their ability to make well-founded inferences about (1) relationships between the intervention and the observed effects (internal validity); and (2) generalizability of the results (external validity and statistical conclusion validity).

Internal Validity

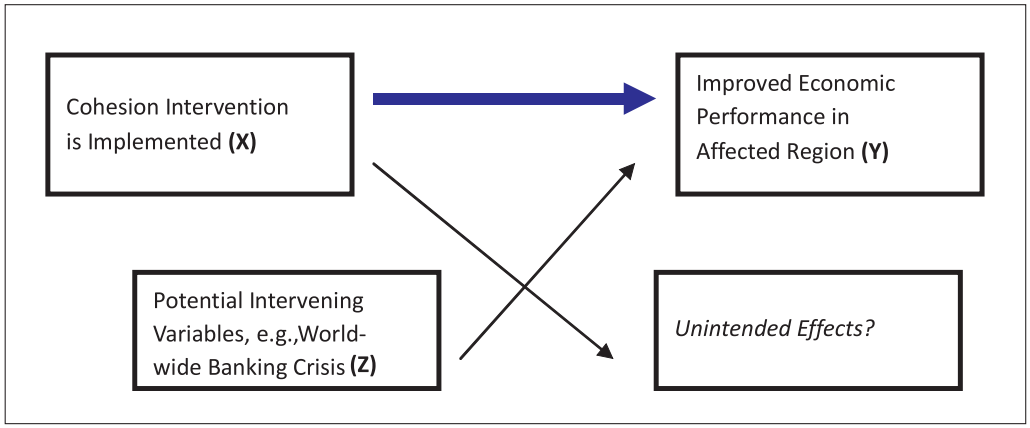
Internal validity is concerned with the ability to determine whether an intervention has produced the intended outcome or result, and to determine the magnitude of the effect. When considering the internal validity of an intervention, the evaluator must assess whether a causal inference can be established between a policy intervention and the desired effect and, perhaps, whether there are unintended effects of the policy.

When making causal inferences, evaluators must measure several elements:

- Temporal order, to ensure the observed outcomes occurred after the policy was implemented;
- Co-variation, which involves observing that after the policy was implemented, changes in outcomes occurred; and
- Mediating factors (or confounding factors), which are other variables that could also have affected desired outcomes.

In addition, observed relationships should be in accordance with expectations from previous evaluation work. See Figure 3 for a diagram depicting the notion of drawing causal inferences.

Figure 3: Making Plausible Inferences About a Policy Intervention



It is very difficult to draw causal inferences. When measuring outcomes, there are several challenges to capturing "net effects" or "net impacts" of the intervention, because there are other events and processes occurring in the neighborhoods or societies that affect achievement of desired outcomes. Time needed for the intervention to change attitudes or behavior may be longer than the time given to measure outcomes. And there may be flaws in the evaluation design or implementation of the intervention that reduce the ability of the intervention to even produce desired outcomes. For these reasons, it may be difficult to ever establish credible causation. It is desirable to use terms such as "plausible attribution" when drawing conclusions about the effects of policies on intended behaviors. When evaluating policies, there are complex sets of interrelated parts; it is incumbent upon evaluators to be thorough when identifying these "moving" parts in order to understand the relative effectiveness of different components. It may be preferable to avoid conducting big multi-regional studies of an intervention, and instead undertake an in-depth regional or local evaluation to purchase more knowledge about an intervention's effects.

In order to strengthen evaluators' abilities to attribute effects to a policy intervention, several steps may be taken:

- Measure the extent to which the intervention was implemented;
- Depending on the intervention being evaluated, ask respondents about other events or experiences they may have that also affected decisions relevant to the policy – before and during the evaluation time frame;
- Given existing knowledge about the likely time period needed to see effects, evaluate whether enough time has elapsed between implementation of the intervention and measurement of intended effects; and
- Review existing evaluation research to identify unintended effects, and build in capacity to measure them.

Generalizability

Evaluators should select sites and individuals that are truly representative of the populations to which they hope to generalize their results. Generalizability includes external validity, or the ability to generalize beyond the groups or context being studied, and statistical conclusion validity, the ability to generalize statistical findings beyond the sample (relevant only to quantitative data).

External Validity

Generalizing results from an evaluation to other sites is especially important in evaluations of policies that may have differential effects on particular subpopulations, e.g., youth, rural, racial/ethnic groups. In order to enhance generalizability, evaluators should make sampling choices that take care to identify subpopulations of interest, such that there are large enough sub-samples of the groups to analyze. However, evaluators should still examine a sample to ensure that it is truly representative of the larger population to which they hope to generalize on demographic variables of interest (e.g., age, ethnic grouping).

Statistical Conclusion Validity

Statistical generalizability, or finding statistically significant findings, is greatly dependent upon the size of the samples used in an evaluation. Estimating the sample size needed to establish statistical generalizability of the results in any evaluation depends on three general criteria:

1. The size of the population to which generalization is desired.
2. The level of confidence desired in results, such as 95 percent or 99 percent.
3. The margin of error acceptable, such as an error band of plus or minus 2 percent, 3 percent, or 5 percent.

Evaluators should also consider desired sub-group populations and ensure sufficient sub-sampling of these groups.

It bears noting that the criterion of statistical significance and the tests related to it have been borrowed from hard sciences, where the concern is to have the highest levels of confidence possible. In a new arena of policy evaluation where there exist many obstacles to obtaining strong evidence, it is reasonable to consider lower levels of confidence. For instance, it may be reasonable to accept a 90 percent level of confidence. It is entirely appropriate to report deliberations on this issue, reasons why a certain level was chosen, and the exact level the evaluation was able to obtain. This is more realistic and productive than assuming that evaluation results will not be discussed unless a, perhaps unrealistically, high level of confidence is reached statistically. Having said that, it is bear noting that in the U.S. evaluation studies that report statistical significance still tend to follow the traditional adherence to a 95% confidence level rule – and many research institutions in the U.S., such as the U.S.

Government Accountability Office, require that such traditional rules be used – in part to preempt any potential for manipulation of findings.

In order to properly report on statistical significance and measured effects, evaluators should report both on the statistical significance of the findings, or whether the sample size allows conclusions to be drawn about the intervention's effectiveness, and the importance and relevance of the size of the effect of the intervention. Since statistical significance greatly reflects the sheer sample size, other pertinent criteria should be identified to characterize the policy relevance of the measured effects.

In order to bolster generalizability of evaluation results, during the design phase evaluators should consider several questions:

- To what groups or sites will generalization be desired?
- What are the key demographic (or other) groups to be represented in the sample?
- What is sample size, with adequate sampling of important subgroups, is needed to make statistical generalizations about the effectiveness of the intervention?
- What aspects of the intervention merit careful measurement to enable generalizability of findings?

Planning Evaluations to Support Inferences

Evaluations should be designed to provide relevant comparisons. Choices about selection of sites and subjects, and the timing of data collection should be made so that findings of an impact will be credible. The set of decisions regarding the type of assignment of units under study employed (i.e., random or not), and the number and timing of observations constitutes the evaluation design.

Evaluation practice offers guidance on designing an evaluation to strengthen inferences made about the impact of policy or program interventions. According to Donald Campbell, usually viewed as the father of research design, the strongest causal inferences will be obtained when implementing a Random Control Trial, or true experimental design, that is, a design that involves both random assignment and the ability to assign randomly subjects to either a treatment or control group. The random assignment with a true control group or site permits needed comparisons about the outcomes of the intervention as compared to outcomes for people or sites that did not have the intervention (or treatment). While this design allows strong inferences to be drawn about the effectiveness of the treatment, random assignment is very difficult to implement, particularly for policy interventions.

At the other extreme, a quite commonly used approach, where-in observations are made about the effects of a policy after it is implemented, also called a One-Group Posttest-Only design (treatment-posttest), is a design in which it is difficult to make inferences about the net effects of a policy on intended outcomes. Adding a pre-

test, or collection of data on intended outcomes before the policy is implemented, creates an improved design. The One-Group Pretest-Posttest design (pretest-treatment-posttest), allows for comparison of baseline data to subsequent outcomes and, thus, permits stronger inferences about attribution of changes due to the policy intervention.

Another strong policy evaluation design is a Time Series Design (multiple pretests-treatment-multiple posttests). This design entails collecting data at multiple time points before and after a policy is implemented, and allows for measurement of trends in outcome data both related and unrelated to the policy intervention, for a period of time before the policy is implemented and for a period after the policy has been in place. And an improvement to this basic time series design is to also collect outcome data for a jurisdiction similar to the jurisdiction being evaluated where the policy was not implemented, to rule out factors in the environment that would have affected the outcomes in both jurisdictions. Multiple time series designs that involve collection of the same data in the treatment jurisdiction and a "control" or comparison jurisdiction, strengthen inferences about the impact of a policy.

Policy evaluations should be designed to permit useful comparisons. When possible, evaluators should collect data on participants, the context in which the policy is implemented, and the intervention itself at more than one point in time (pre- and posttest evaluations) and over a longer period, in order to measure outcomes of interventions fairly. It can be incredibly useful to find a comparison site or jurisdiction where an intervention was not implemented to serve as comparison units. Statistical analyses also may be used to attempt to control for mediating or intervening variables when analyzing outcome data. Sophisticated means for assessing the impact of policy interventions relative to intervening variables, such as the use of instrumental variables and propensity scoring, are being used more frequently. There are many texts that provide guidance on the use of these techniques (For example, see Shadish, Cook and Campbell, 2002).

(4) Crafting Findings and Recommendations that are Credible

In the end, careful planning and reasoned decision making about both measurement and design will not ensure that all evaluations will produce credible results. There are a variety of pitfalls that frequently constrain evaluation findings, as summarized in Table 1. As is noted in the table, inadequate planning (the first five pitfalls) as well as inappropriate or incomplete analysis and reporting (the second five pitfalls) are likely to weaken all of the types of validity and reliability discussed above. Clarity in reporting with open discussion about methodological decisions, and any obstacles encountered during data collection will bolster confidence in findings.

Table 1. Common Pitfalls

	Pitfalls	Measurement Validity	External Validity	Internal Validity	Statistical Conclusion Validity	Reliability
1.	Failure to assess whether the policy or program is evaluable yet.		X	X	X	
2.	Not devoting sufficient time and deliberations to identify criteria for measuring implementation and outcomes.		X	X	X	
3.	Failure to assess the quality, completeness, and accuracy of data.	X	X	X		
4.	Failure to pretest data collection instruments and train data collectors appropriately.	X	X	X	X	X
5.	Failure to collect the appropriate data to support our findings and recommendations.	X	X	X	X	X
6.	Failure to address non-response issues and other sample size issues.		X	X	X	
7.	Applying an analytical technique without meeting important assumptions about the data.			X	X	X
8.	Generalizing beyond the confines of the sample, or the limits of the study sites.		X			
9.	Failure to adequately support findings and recommendations with specific data.		X			X
10.	Poor presentation of the data.					X

The context of the policy intervention should be described in enough detail that others many know whether it is feasible to replicate a specific policy intervention. For example, this means that relevant demographic information about the participants in the evaluation should be included. And the manner in which all key components of the intervention were implemented should be described in enough detail to help clarify why and how the policies produced results.

Transparency about the scope and methods employed is also essential. Evaluation reports should:

- Describe design and data collection with sufficient detail for the intended audience;
- Discuss response rates and potential for misinterpretation due to low response rates, as well as what was done to minimize the effects of any such response problems;
- Describe key decisions that affect statistical results, e.g., omitting outlying values in calculations;
- Address limitations to all four kinds of validity and reliability in user-friendly terms; and
- Clarify how to use and interpret the data, for example, "Under these circumstances, X appears to be linked to changes in Y…"

When describing methodological issues or challenges, conciseness and clarity is especially helpful (See Wholey, Hatry and Newcomer, 2004). And it is always good to include lessons learned from conducting evaluations to assist evaluators in future evaluation work.

Conclusion

The ability to produce credible conclusions about the effectiveness of cohesion policy interventions will be strengthened by careful planning and reasoned decisions about evaluation methods. Evaluators should take care to identify or develop valid and reliable measures, identify the population and sub-populations to which they want to generalize, design sampling strategies that allow for appropriate generalization, and select an evaluation design that is feasible and will permit needed comparisons. Where possible, an evaluation should include baseline (pretest) data, and identify comparison groups or jurisdictions and measure target behaviors/outcomes as many times as possible. Key components of the policy intervention and the degree of implementation should be described in enough detail to help clarify why and how the policies produced results.

And finally, transparency about the methodological decisions made throughout the evaluation will bolster the credibility of the work.

However, in the best of conditions and even with thorough planning, it should be noted that it is extremely difficult to produce clear findings about program or intervention effectiveness. Expectations matter. The expectation that it will be possible to measure precisely the impact of a specific intervention in a region or even a neighborhood is probably higher than warranted for a number of reasons. So much economic upheaval has happened recently – bankruptcies in major sectors of economies, continuing high employment, subnational treasuries in peril, to name a few – that it will be impossible to attribute changes in employment rates

and economic recovery due to one intervention. And while the call for an accounting of an intervention impact may sound simple, it will not be simple to verify how much of economic growth can be traced to specific interventions and how much would have happened anyway as part of a standard rebound from a downturn. Humility is certainly needed for evaluators as they address the many challenges to designing evaluations with extremely sound methodological rigor to produce credible findings.

References

Bachtler, John & Grzegorz Gorzelak. (2007). Reforming EU Cohesion Policy. Policy Studies 28(4): 309–326.

Bachtler, John & Carlos Mendez. (2007). Who Governs EU Cohesion Policy? Deconstructing the Reforms of the Structural Funds. Journal of Common Market Studies 45(3): 535–564.

Bachtler, John & Colin Wren. (2006). Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges. Regional Studies 40(2): 143–153.

Basile, Roberto, Castellani, Davide & Antonello Zanefi. (2007). Location choices of multinational firms in Europe: The role of EU cohesion policy. Journal of International Economics 74(2): 328–340.

Baslé, Maurice. (2006). Strengths and Weaknesses of European Union Policy Evaluation Methods: *Ex-post* Evaluation of Objective 2, 1994–1999. Regional Studies 40(2): 225–235.

Blom-Hansen, Jens. (2005). Principals, agents, and the implementation of EU cohesion policy. Journal of European Public Policy 12(4): 624–648.

Bradley, John. (2006). Evaluating the Impact of European Union Cohesion Policy in Less-developed Countries and Regions. Regional Studies 40(2): 189–199.

Bradley, John & Gerhard Untiedt. (2008). EU cohesion policy and "conditional" effectiveness: What do cross-section regressions tell us? GEFRA Working Paper No. 4. Munster, Germany.

Gilmour, John B. and David E. Lewis. (2006) Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores. Public Administration Review 66 (5): 742–752.

Mairate, Andrea. (2006). The 'Added Value' of European Union Cohesion Policy. Regional Studies 40(2): 167–177.

Mason, Jennifer (2002). Qualitative Researching, 2nd edition, Sage.

Metzenbaum, Shelly. (2009) Performance Accountability: The Five Building Blocks and Six Essential Practices. Washington, DC: The IBM Center for the Business of Government.

Moynihan, Donald P. (2008) The Dynamics of Performance Management.: Constructing Informatin and Reform. Washington, DC: Georgetown University Press.

Newcomer, Kathryn (2006) "How does Program Performance Assessment Affect Program Management in the Federal Government?" Public Performance and Management Review, June 2006, volume 29.

Newcomer, Kathryn and F. Stevens Redburn (2008). ACHIEVING REAL IMPROVEMENT IN FEDERAL POLICY AND PROGRAM OUTCOMES: THE NEXT FRONTIER. Washington, DC: National Academy of Public Administration.

Patton, M. Q. (2008). Utilization-Focused Evaluation. *4th Edition*. Thousand Oaks, CA: Sage Publications.

Shadish, William, Thomas Cook and Donald Campbell. (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin.

Wholey, Joseph; Harry Hatry & Kathryn Newcomer Eds. (2004) The Handbook of Practical Program Evaluation, Jossey-Bass

Directions in Public Program Evaluation in the US

Maureen Pirog, Lanlan Xu

I. Introduction

Program evaluation and social experimentation are essentially about causation. Does some deliberate manipulation, sometimes called a "trial," "treatment" or "independent variable" *cause* a change in specified "outcomes" or "dependent variables" that are theoretically tied to one another? From a public program evaluation perspective, "treatments" are largely comprised of government programs, policies, laws, and/or managerial innovations. For examples, in the realm of economic development, "treatments" might include new tax incentives to attract industries, special interest rates on loans for new or expanding businesses, economic development zones, or the number or volume of micro-credit loans. If one were to evaluate the effects of the "one child policy" in China, "outcomes" might include changes in population growth, the balance of males and females in new birth cohorts, malnutrition, the quality of elder care, abortion rates, adoption rates, or a host of other possibilities.

An example of a typical public program evaluation question is whether or not lowering legal blood alcohol content (BAC) levels for countries within the European Union (EU) to .05 from .08 results in fewer traffic fatalities. To address the EU BAC level and traffic fatality issue or to assess the impact of any other "treatment," a basis of *comparison or counterfactual* is necessary. For example, EU countries with .05 and .08 BAC limits can be compared to one another using cross-sectional data or fatality rates within the same EU country can be compared pre and post-implementation of the lower BAC levels using longitudinal data. Both comparisons between countries at a given time and comparisons of a country to itself over time can be made using panel data (See Albalate, 2008). Similarly, one could estimate the effects on the diversity of college incoming classes of a new law such as Texas H.B. 588 which guarantees admission to any Texas public university for all students who graduate in the top 10 percent of their high school class (Niu and Tienda, 2010). To assess the impact of this law, the authors looked at matriculation rates of students from different racial/ethnic backgrounds who scored just above and below the 10 percent cutoff using a regression discontinuity design.

II. Experimental versus Quasi-Experimental Studies

How one constructs a counterfactual gives rise to the distinction between experimental and quasi-experimental studies. Experiments use random assignment

(e.g., coin toss, die, page of random numbers, or computerized random number generator) to assign units to treatments. Among many other possibilities, units could consist of firms, cities, states, regions or countries in economic development studies; or individuals in social program or health studies. Random assignment is considered the "gold standard" in evaluation research simply because *ceteris paribus* units in the treatment and control groups should be statistically equivalent on both observable and *unobservable* characteristics. In contrast, in quasi-experimental studies some other nonrandom method of selection is used to assign treatments to units: first-come, first-serve; neediness; perceived ability; etc. When nonrandom assignment processes are employed, units are likely to differ on measurable and unmeasurable characteristics. From an evaluation perspective, measurable differences (e.g., age, race, weight, height, number of children, income) are not problematic as they can be controlled for in statistical analyses, although it may take considerable effort to collect the data. However, *unmeasurable* differences (e.g., intelligence, motivation, altruism) are much more problematic because if they are correlated with the outcome measures, they can have large effects on estimates of program impacts. Ignoring relevant unmeasured characteristics normally biases estimates of program impacts. This source of bias is called selection bias or self-selection bias if units volunteer for the treatment.

Think about the effectiveness of job training programs funded by the European Social Fund, for example, for which participants were randomly selected versus those who were so motivated to join that they arrived early to wait in line. Clearly the groups of participants would not be equivalent. They may differ in observed characteristics such as education (which can be measured and controlled statistically) but they also differ in motivation. Those individuals who stood in line and self-selected into the job training program typically will outperform individuals who were randomly assigned to the job training. The magnitude of this self-selection bias is difficult to assess, making it difficult to judge the true magnitude of the programmatic impact. If patronage was used to assign training to applicants, the biases would be different but would still exist. Perhaps those with the best political ties would have better social networks and eventually better access to jobs.

Because of these selectivity issues, randomized experiments are typically preferred to quasi-experimental studies. Additionally, random assignment studies have the advantage that the statistical analysis required is simple and the results are typically easy for the general public, including politicians, to understand.

On the flip side, undertaking randomization requires designing the experiment *prior* to program implementation so that units can be randomly assigned to treatments. Often treatment dosages or interactions between multiple treatments are of policy interest and these factors also need to be taken into consideration prior to program implementation. This level of advance planning is often impractical and as such, evaluators are left to fall back on quasi-experimental designs. Randomization also has to be politically feasible and ethically justifiable. In many European countries

where social cohesion and social integration are highly valued, it may be more difficult to make the case for randomized social experiments.

It should also be noted that while the random assignment of treatments to units may seem straightforward, in practice this process can be complex. The trick is to design a randomization process that cannot be undermined by program operators who have incentives to demonstrate program effectiveness, often by "cream-skimming"; treatment units that may want to opt in or out of treatment; or politicians who may want to ensure that the treatment units in their political districts (e.g., persons, schools, hospitals, water treatment plants) receive what is perceived a priori to be the preferred. At times, implementing randomization may require modification of complex client eligibility determination software packages to ensure that randomly selected applicants are eligible only for the treatment or control group. Some experiments like the Mexican Universal Health Insurance Program have explicitly incorporated some of these pressures which can undermine the randomization process even as part of their initial designs (King *et al*., 2007).

Additionally, while the treatment and control groups may be statistically equivalent at the point when they are assigned to groups, they may no longer be equivalent if different types of persons from each group leave the experiment. For example, a highly motivated person randomly assigned to a "no treatment" group in a job training program might quit the experiment and seek alternative types of treatments. Differential attrition from the treatment and control groups undermines the statistical equivalency that randomization initially achieved. And, it is also possible that the types of individuals who even agree to be randomized are systematically different from individuals who would participate in a program if it were offered more broadly. Think about someone who is critically ill and seeking help. They might not agree to being randomized into a "no treatment" or "likely less effective" treatment, preferring to proactively seek better treatments. When this occurs, estimated treatment effects for the group of individuals included in the experiment may actually be different from the population-wide treatment effects if the program becomes more generally available. This phenomenon is referred to as randomization bias. Both differential attrition and randomization bias can alter estimates of "true" treatment effects.

Finally, it has been argued, although never definitively proven, that randomized experiments are more expensive than other types of evaluation due to the costs associated with implementing and monitoring randomization. However, any additional funds needed to implement and monitor randomization should be weighed against the extra costs of more complex statistical and econometric modeling in quasi-experimental evaluations.

Because of their ease of interpretation and the selectivity issues described above, randomized assignment studies are preferred theoretically to quasi-experimental and non-experimental evaluations. Yet, in practice, at least in the US, quasi-experi-

ments are far more common than experiments with the exception of studies in the health arena where double-blind randomized drug trials are the norm. Recently, as editor-in-chief of the *Journal of Policy Analysis and Management (JPAM)*, I reviewed the evaluations published in 2009. JPAM is known for rigorous policy and program evaluations. However in 2009, all of the evaluations published in JPAM were quasi-experimental. This is a case where theory dictates social experiments or randomized clinical trials (RCTs) but other practicalities lead to a preponderance of quasi-experiments.

Because quasi-experiments are far more prevalent, a vast, complex literature on correcting for selection bias has developed. James Heckman (H) received the Nobel Prize in Economics in 2000 for his pioneering work on correcting for selection bias. Since his early work on the topic in the 1970s, a variety of other methods have been developed to correct for selectivity including propensity score matching (PSM), difference-in-differences (DD) methods, instrumental variables (IV) approaches, and fixed effects (FE) models, each with their own limitations and applicable only under certain circumstances. Regression discontinuity designs (RDD) are also becoming more commonplace as they appear to fairly accurately replicate experimental results when comparing individuals just above and below an eligibility cutoff.

III. Sources of Program Evaluation Research

Program evaluations are conducted for a variety of purposes and by a variety of groups. Advocacy research – research conducted by individuals with a deep, vested interest in a problem – has a long history in the United States. However, since the 1960s, social advocacy groups have increasingly employed questionable tactics to legitimate their work and muddy policy debates. Unreasonable definitions of problems, flawed research methods and misinterpretation of findings sometimes have resulted in exaggerations of the magnitude of social problems that have been picked up by news outlets and television talk shows (Gilbert, 1997). Abortion advocates, right-to-life groups, proponents for the rights of victims of violence, environmentalists, animal rights organizations and similar groups all have vested interests in promoting specific political agendas and hence, there are strong incentives to impress the importance of the issue or problem on policy makers.

Federal government departments and agencies also produce evaluation research, often through the Office of the Assistant Secretary for Planning and Evaluation but also through in-house research shops focused on specific government programs. The US Congressional Research Service (CRS) employs over 700 professional researchers to conduct specific studies on behalf of the members of Congress. The CRS is supposed to provide nonpartisan, unbiased research on the issues confronting Congress.

Not all policy research that is designed to influence the policy process is conducted by government agencies. Sometimes government produced research can be viewed as politically motivated and potentially biased. At other times, the skills needed to conduct evaluations exceed the capacity of government research operations. Some evaluations require very specialized knowledge and field operations that would be more cost-effective if conducted by researchers closer to the data or geographic areas. As such, the US federal government undertakes substantial contract and grant activity with research companies, think-tanks and universities. In Fiscal Year 2007 alone, the National Science Foundation reported that the federal government provided $30.1 billion for university research (National Science Foundation, 2009).

Universities in the US are classified by the Carnegie Foundation for the Advancement of Teaching (2010) into various categories. Doctorate-granting universities, institutions that awarded at least 20 doctorates in 2003-04, are classified by their level of research activity, as measured by research expenditures, number of research doctorates awarded, number of research-focused faculty, and other factors. There are 96 universities that are classified as having Very High (RU/VH) research activity, and these represent approximately 2.2% of all degree granting schools in the U.S. These 96 schools received approximately $24.9 billion in federal R&D funding in 2007. An additional 2.3% of degree granting institutions, or 103 schools, are classified as having High (RU/H) research activity, and these schools received about 3.4 billion in federal R&D funding in 2007 (National Science Foundation, 2009). The federal government also provides grants and contracts to think-tanks, consulting firms and nonprofit research organizations that are not university based. Exact figures for the magnitudes of these activities are not tallied but the level of this activity is very substantial. Because workers in these organizations are not tied to the academy and do not have the obligations of university professors, they can be more flexible in meeting some of the research needs of the federal government. In the area of social service evaluation, several very large groups are notables in this field including, among many others, the MDRC, Mathematica Policy Research, Inc. (MPR), Abt Associates, Inc., the Urban Institute, RAND, and the Brookings Institution.

One of the large differences between universities and non-university research companies is that the evaluations conducted by academics are more likely to be published in scholarly venues as academic researchers are afforded the time and incentives to publish scholarly research. Elite not-for-profit think-tanks like the American Enterprise Institute and the Brookings Institution are rather different. Some of these institutions have considerable foundation support and there is a sharp focus on publishing. With private consulting companies, the final report to the government is more likely to be the end product. Thus, there is a difference in the dissemination of the research findings based on the type of organization conducting the research.

IV. Trends in Evaluations Published in the *Journal of Policy Analysis and Management (JPAM), Evaluation Review (ER)* and *Economic Development Quarterly (EDQ)*

A. Selection Bias

We conducted an eight year retrospective review of articles published in *JPAM, ER* and *EDQ* to determine trends in published evaluation research. We only focused on those articles that did include an evaluation of a program, policy, law and/or managerial innovation. We chose JPAM because it is known for publishing rigorous evaluation research, EQ because its sole focus is evaluation, and EDQ because it is a substantively important field of inquiry that also has a strong evaluation component. A summary of this review is found in Table 1.

Table 1 Summary of Published Evaluation Studies

	Total	JPAM	ER	EDQ
Program Evaluation Articles	225	133	48	44
International	20	10	8	2
Experimental	22	7	15	0
Quasi-Experimental	203	126	33	44
Correct for Selection Bias	137	99	13	25
DD	6	6	0	0
IV	5	5	0	0
H	5	2	0	3
PSM	12	7	4	1
FE	54	41	5	8
RE	7	4	1	2
2SLS	6	6	0	0
RDD	1	1	0	0
Other	5	2	0	3
Multiple	36	25	3	8
None	64	27	20	17

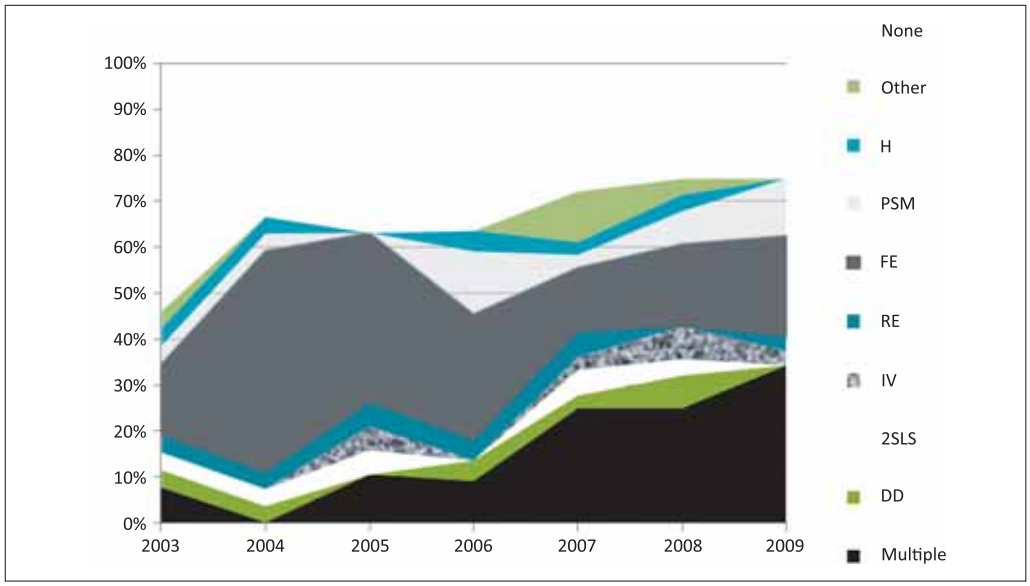
H=Heckman, PSM=Propensity Score Matching, DD=Difference-in-Differences,
IV= Instrumental Variables, FE=Fixed Effects, RDD=Regression Discontinuity Design

One clear trend has been discussed above: While a randomized experiment is considered the "gold standard" of evaluation research, the large majority of published evaluations are quasi-experimental. This extends beyond 2009 in *JPAM* and even when examining other journals that publish evaluation research like *ER* and *EDQ*. In the past eight years, 226 evaluations have been published in *JPAM*, *ER* and

EDQ of which only 22 or 9.7 percent were randomized experiments. Also, roughly two-thirds of quasi-experimental evaluations use one or more corrections for self-selection.

Because of our increasing awareness of the seriousness of selection biases, evaluators are relying increasing on a growing arsenal of statistical corrections. The following graph depicts changes in the use of the various selection correction methods in published evaluations in the past eight years in *JPAM, ER* and *EDQ*. See Figure 1.

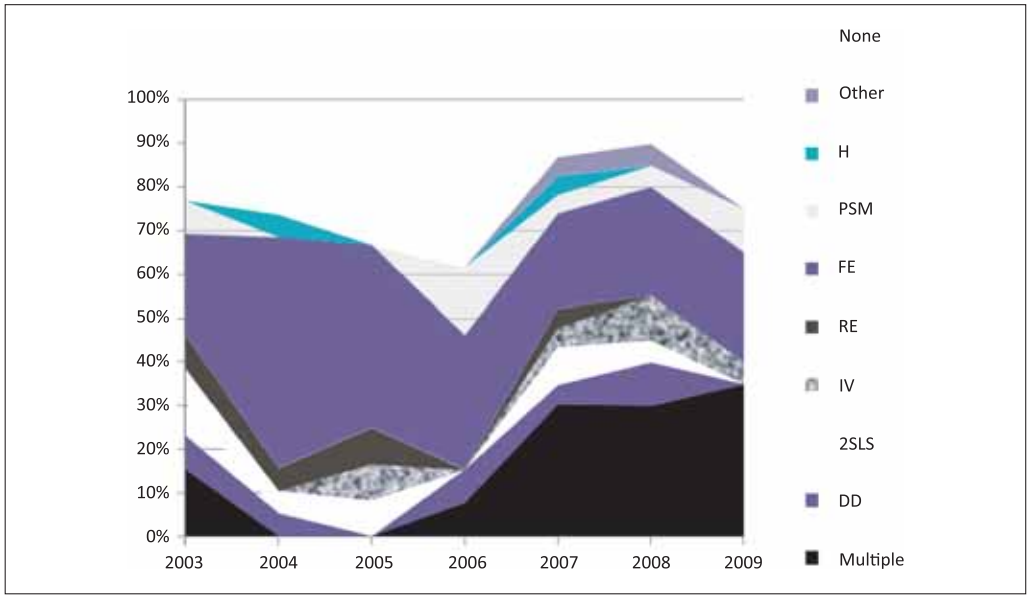
Figure 1: Changes in use of Selection bias correction methods



As is clearly seen in Figure 1, the percentage of published articles using some type of statistical correction for selection bias has increase from 46 percent in 2003 to 75 percent in 2009. Further, the percent of publications using two or more statistical correction methods for selection bias has increased from roughly 7 percent in 2003 to 34 percent in 2009. The growth in the "multiple" category is probably best explained by the desire of many authors to demonstrate the robustness or sensitivity of their findings to changes in their model specifications. It also demonstrates an increasingly better understanding of these methods in the evaluation community. When only one correction method is used, fixed effects seem to dominate followed by propensity score matching. Instrumental variables, difference-in-differences, the Heckman correction, and Two-Stage Least Squares are used alone in relatively small proportions of articles.

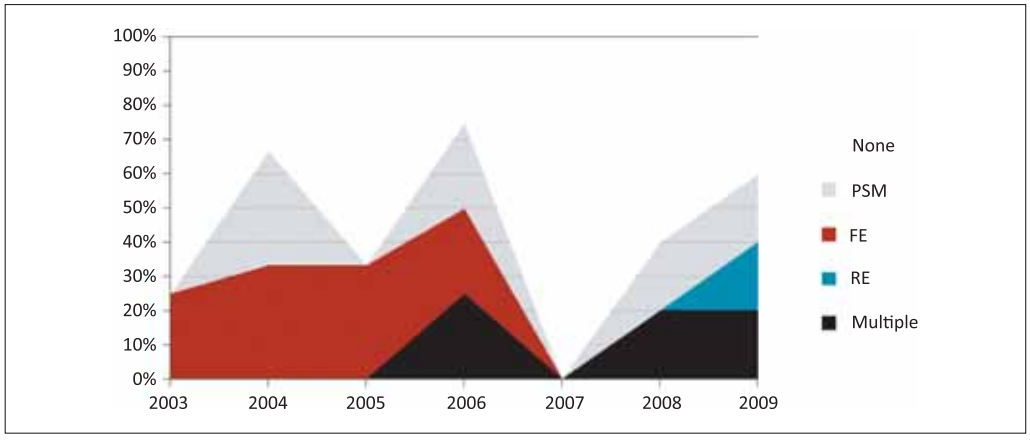
Because *JPAM* published so many more evaluations than either *ER* or *EDQ*, *JPAM* tends to dominate and obscure the journal specific figures. The use of these methods does vary a fair amount across journals, perhaps reflecting some disciplinary differences or variation in the types of data employed. Journal specific trends are depicted in Figures 2–4.

Figure 2: Change in use of selection bias correction methods in JPAM



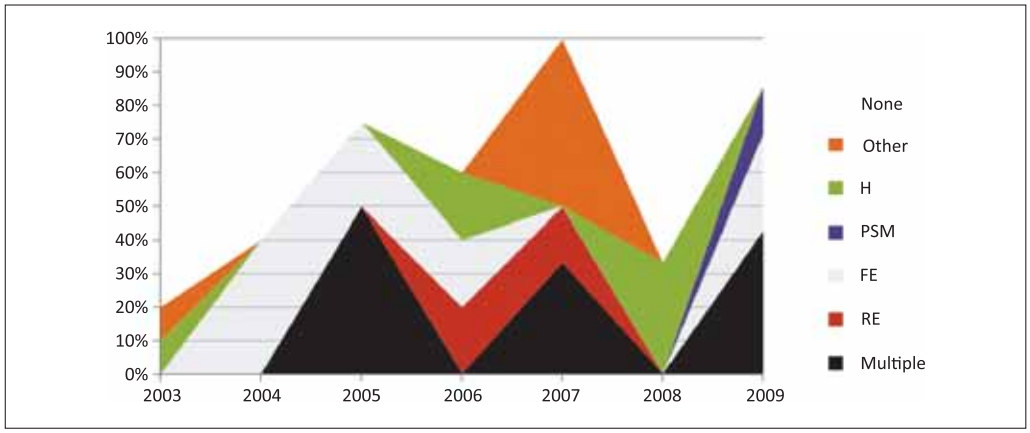
Interestingly, there was less growth in the use of these methodologies in *JPAM*. Seventy-five percent of the published evaluations used one or more of these methods in 2003 and the figure was the same in 2009 although there were some small changes across the intervening years. Again, the one clear change is that many more articles are now using two or more of these methodologies to conduct robustness checks.

Figure 3: Change in use of selection bias correction methods in Evaluation Reviev



The findings for *ER* exhibit more volatility simply because of the smaller number (48) of evaluations published in this journal in the 2003–2009 volumes, inclusive. Nonetheless, fewer articles rely on these statistical methods and when they do, they rely on propensity score matching, fixed effects, random effects or more recently, multiple methods.

Figure 4: Change in use of selection bias correction methods in Economic Development Quarterly



Again, the results for *EDQ* are simply erratic given that only 44 evaluations were published between 2003 and 2009. Divided over 7 years, even small changes look large in percentage terms. Despite this caveat, these methodologies are increasing used in this journal as well. By 2009, only 1 of the 7 of the quasi-experimental evaluations published in this journal did not use some method to adjust for selection bias.

It is unclear how to interpret the increasing use of these methodologies. On the one hand, it appears that they are important in *JPAM* and increasingly important in the other two journals under examination. Multiple methods are being used more often across the board and from a sensitivity analysis perspective, one can hardly argue with this approach. Yet, there is a growing body of evidence on the ability of these methods to actually replicate experimental results (summarized by Pirog, Buffardi, Chrisinger, Singh & Briney, 2009). Sadly, at this point in time, most of these methods do not reliably replicate the "true" experimental impacts. While theoretically possible, it seems that the circumstances and conditions or knowledge needed to actually replicate experimental results are too onerous for most evaluators to implement. This suggests that a lot more work is needed to improve these or yet to develope new methods to deal with selection bias.

B. Types of Data Used in Experimental and Quasi-Experimental Evaluations

Another distinctive difference between experiments and quasi-experiments is that most randomized experiments require original data collection whereas many quasi-experimental evaluations use secondary data already collected from the US Census, the Survey of Income and Program Participation (SIPP), the National Longitudinal Survey of Youths (NLSY), the Panel Study of Income Dynamics (PSID), the *Fragile Families and Child Wellbeing Study*, state or city administrative databases, or other

such sources. In many quasi-experiments, data are combined from multiple sources. For example, in Houston and Richardson's 2006 JPAM study on seat belt use and traffic fatalities, they extracted traffic fatality data from the National Highway and Traffic Safety Administration's (NHTSA) Fatal Reporting System (FARS) for the years 1990–2002. Additional data were obtained from the Federal Highway Administration, the Insurance Institute for Highway Safety, NHTSA's *Traffic Safety Facts*, the US Census, and the US Department of Labor. Similarly Lopoo and Delire's 2006 study of the impact of welfare reform on teen fertility used natality reports from the National Center for Health Statistics, demographics from the US Census data, birth data from the National Vital Statistics Reports, and state level data from a variety of additional sources. These studies are not exceptional in this regard.

Because the quality of the research depends heavily on the quality of the data used, this topic deserves some elaboration. International evaluations appear to also rely heavily on secondary data sources and combined data from multiple sources. Recent international evaluations have focused on the effects of nonprofit status of child care providers on the quality of child care in Canada (Cleveland & Kranshinsky, 2009); deforestation in Bolivia (Andersson & Gibson, 2006), public management reform in the UK (Walker & Boyne, 2006), pollution in Canada (Harrison & Antweiler, 2003), pollution, economic growth and national debt in Latin America (Aubourg, Good, & Krutilla, 2007), Columbian sickness funds (Trujillo & McCalla, 2004), drug use surveys in Australia (White, Hill & Effendi, 2004), the effectiveness of adult drug courts in New South Wales (Shanahan, Lancsar, Haas, Lind, Weatherburn, & Chen, 2004), and union formation in Canada (Harknett, 2006). For example, the article on debt, democratization and development (Aubourg, Good & Krutilla, 2007) utilizes data on 29 Latin American and Caribbean countries from the 2001 World Bank Development Indicators CDROM with data inconsistencies being completed or corrected by data published by Haiti's Central Bank, the Caribbean Center for Monetary Studies at the University of the West Indies in Trinidad and Tobago, and the Center for Latin American Monetary Studies in Mexico. The Albalate (2007) study on blood alcohol content limits and traffic fatalities in the former EU15 countries relied on the European database Community database on Accidents on the Roads in Europe, Eurostat, WHO Europe, World Bank Development Indicators and World Road Statistics.

Many US focused quasi-experiments also rely on large databases such as the Census, the Survey of Income and Program Participation (SIPP), the Panel Study of Income Dynamics (PSID), the National Longitudinal Survey of Labor Markets – Youth Cohort (NLSY), Fragile Families, National Surveys of American Families, among others. The US Census provides data on numerous measures including demographics, income, manufacturing establishments, and poverty. Data are released in the aggregate, by states, counties, Census tracks, as well as individual responses. Data are available on CDROM and on-line. Because of the fundamental nature of the data provided by the Census, they are used frequently in evaluation research. Recent studies

exemplify the diversity of topics that can be addressed using Census as a primary or secondary data source: the impact of urban enterprise zones on manufacturing establishments (Greenbaum & Engberg, 2004); the effects of full-day kindergarten (Cannon, Jacknowitz & Painter, (2006); the effects of contracting out on public sector employment (Fernandez, Smith & Wenger, 2006); immigration and welfare reform (Haider, Schoeni, Bao, Danielson, 2004); and, the effect of state R&D tax credits on private sector R&D (Wu, 2005).

Similarly, state-level and city-level databases have proven to be a goldmine for longitudinal assessments of state programs and policies as well as interstate comparison of public policies where there is variation in policies across states. State administrative databases have been used widely in education studies (Wilde & Hollister, 2007), evaluation of state children's health insurance programs (Marton, 2007), child support programs, welfare reforms (Page, Spetz and Millar, 2007; Loeb, Fuller, Kagan & Carrol, 2003), smoking laws (Stolzenberg & D'Alessio, 2007), and traffic laws (Stolzenberg & D'Alessio, 2003). Securing confidential data from state agencies can be challenging as well as the sheer size of many of these datasets. There were over 46,000 observations in the study on Marton's 2006 study of Kentucky's children's health insurance program. Even city level studies can produce large datasets. For example, when examining the impacts of subsidized rental housing on property values in NYC, data were amassed on 432,984 property sales (Ellen, Schwartz, Voicu & Shill, 2007). In the same city, Joyce, Gibson and Colman (2005) examined differences in birth weights for 811,190 singleton births comparing infant birth weights for women who were on and off of the Women, Infants and Children (WIC) program.

The federal government has also supported the development, fielding and administration of several very large surveys, usually multi-year panels, of teens and families that have been enormously rich sources of evaluation data for researchers. There are typically enough observations in these databases that cross-state comparisons can be drawn about the effects of variation in state policies on individual or family outcomes. Examples of these surveys include the SIPP, PSID, NLSY, Fragile Families, and National Surveys of American Families. These surveys are used extensively by researchers. For example, the NLSY website (http://www.bls.gov/nls/nlsbib.htm) lists over 6,000 NLS-based journal articles, monographs, working papers, and dissertations. The PSID website (http://psidonline.isr.umich.edu/Publications/Bibliography/default.aspx) lists over 2,800 journal articles, books and book chapters, and conference proceedings. Again, there are numerous publications in top journals such as *JPAM* and *ER* that utilize these data.

To summarize, in the United States most published evaluation studies are quasi-experimental and rely on large secondary data sources. Most of these secondary data sources consist of large federal databases such as those constructed and maintained by the Census Bureau, the US Depart of Labor, the Environmental Protection Agency, the Department of Justice and so on. State and city agencies can also provide detailed data to researchers. Also, large special databases, usually

individual and family surveys, are collected for the specific purposes of conducting behavioral and policy research. These databases are heavily utilized in academia as well as by many think-tanks.

V. Overarching Issues in Program Evaluation

A. Independence of Public and Published Evaluation Research?

Questions have been raised in the evaluation literature (Reingold, 2008) as to the independence or unbiasedness of evaluation research funded by government grants and contracts. Some firms as well as universities rely heavily on government funding. As such, it is in the interests of researchers to maintain good relationships with funding agencies who may have a vested interest in finding particular programs to be effective or ineffective. Sometimes, grant and contract sponsors ask researchers to modify their study samples to include particularly successful programs or ignore poor performers, sometimes under the guise of particular programs being too new to evaluate or too atypical to study. Also, funding agencies usually have an opportunity to comment on draft final reports and can make recommendations to researchers to frame their findings in a more positive light.

When all else fails, government funding agencies can sequester or shelve findings for so long that the results are no longer interesting if and when they are released to the public. For example, Former Surgeon General Richard H. Carmona told a Congressional panel that top Bush administration officials repeatedly tried to weaken or suppress important public health reports because of political considerations. He was not allowed to speak on issue reports on a variety of topics including stem cells, emergency contraception, sex education, or prison, mental and global health issues. Additionally, top officials delayed for years and tried to "soften" a seminal report on secondhand smoking that concluded that brief exposure to cigarette smoke could cause immediate harm (New York Times, 2007). These practices are not restricted to government research shops. A recent *JPAM* article (Trenholm, Devaney, Fortson, Clark, Quay & Wheeler, 2008) by researchers from a large, evaluation corporation was released by the US Department of Health and Human Resources after considerable delay because it found that abstinence-only sexual education was largely ineffectual – a finding that was politically unpopular with the Bush administration. A report by one of the authors for this article (Pirog & Kioko, 2010) was sequestered for years by the Inter American Development Bank because of political considerations in the country in which the research was conducted.

While all of these activities exist in the US, there is a real debate about how common these practices are and whether or not they really bias the ultimate findings of researchers.

B. Informed Consent and Other Ethical Considerations

Another issue in the evaluation research receiving press regards the extent to which participants in experiments and other evaluations are treated ethically – particularly as some "treatments" can be harmful (Curry, 2001; Brody, 2002), despite full compliance with requirements established by Institutional Review Boards in universities and other firms. These boards have established tests for researchers and detailed guidelines for the conduct of research designed to protect human subjects from unethical or negligent behavior that could harm study participants and subject the research institution to legal action. Some argue, however, that these boards that were set up in response to blatant disregard by researchers for human subjects have not imposed unreasonable restrictions on researchers. Others debate what ethical standards should be used in evaluation research and if there should be formal mechanisms to ensure that evaluations are conducted ethically (Nathan, 2005; Blustein; 2005; Barnow, 2005, Rolston, 2005; Schochet, 2005).

VI. Implication for Program and Policy Evaluators

To facilitate the growth and healthy development of a strong evaluation research community that can help inform public policy making, several recommendations can be drawn from this review.

Encourage government agencies to collect and make available datasets on the basic characteristics of the population, poverty, income, employment, economic growth, pollution, industries, and other policy relevant data series. Even though these datasets are initially intended for administrative purposes, see if individual-specific identifiers can be placed on each record that are the same across programs and government agencies. This will allow merging of administrative data to give fuller pictures of the effects of changes in one program or a mix of programs on a variety of outcomes.

- Encourage the exploration of country, region or even city-specific data maintained by government agencies. Try to establish guidelines whereby researchers can remove safeguard of the personal information in these files and yet use them for research purposes. And, again, the same issue of individual identifiers that allow the linking of large administrative datasets is critical.
- Encourage the expansion of funding for the collection of large, panel data on persons and families to facilitate research on a broad array of social and economic programs. While making data comparable across EU countries is difficult work, it is possible as aptly demonstrated by the Luxembourg Income and Wealth Studies, and continued efforts in this area will facilitate comparative evaluations since similiarly designed programs may differ in their effectiveness in different cultures or settings.

- Politics play a role in formulating evaluation research in the US and in its dissemination. Researchers should establish the intellectual property rights of these evaluations as their own so as to avoid sequestering and other political maneuvering, if possible. Try to establish professional guidelines that will minimize political pressures on the research community.
- A consortium of universities or professional organizations need to establish guidelines for evaluators for the ethical treatment of human subjects.
- Workshops or symposia on selectivity correction and other evaluation issues should be offered for evaluation researchers as well as Master's and doctoral students. Better funding for research in developing and improving these methodologies as well as research on the circumstances under which different selection bias correction methods are more or less effective is also needed. This is essential to producing credible evaluation results and being competitive in publishing in the academic arena, particularly given reliance of the evaluation field on quasi-experimental methods.

References

Albalate, D. (2008). Lowering Blood Alcohol Content Levels to Save Lives: the European Experience. *Journal of Policy Analysis and Managemen*t, 27(1): 20–39.

Andersson, K., & Gibson, C. C. (2006). Decentralized governance and environmental change: Local institutional moderation of deforestation in Bolivia. *Journal of Policy Analysis and Management*, 26(1), 99–123.

Aubourg, R.W., Good, D. & Krutilla, K. (2007). Debt, Democratization, and Development in Latin America: How Policy Can Affect Global Warming *Journal of Policy Analysis and Management* 27(1), forthcoming.

Barnow, Burt S. "The Ethics of Federal Social Program Evaluation: A Response to Jan Blustein," *Journal of Policy Analysis and Management*. Autumn 2005, 846–848.

Blustein, Jan. (2005). "Toward A More Public Discussion of the Ethics of Federal Social Program Evaluation," *Journal of Policy Analysis and Management*. Autumn 2005, 824–846.

Blustein, Jan. (2005). "Jan Blustein's Response," *Journal of Policy Analysis and Management*. Autumn 2005, 851–852.

Brody. "Research on Children and Lead." *The Chronicle of Higher Education*. March 15, 2002.

Cannon, J. S., Jacknowitz, A., & Painter, G. (2006). Is full better than half? Examining the longitudinal effects of full-day kindergarten attendance. *Journal of Policy Analysis and Management*, 25(2), 299–321.

Carnegie Foundation for the Advancement of Teaching. (2010). University categorizations were accessed on 3/31/2010 at: http://classifications.carnegiefoundation.org/descriptions/basic.php.

Cleveland, G. & Krashinsky, M. (2009). The nonprofit advantage: Producing quality in thick and thin child care markets. *Journal of Policy Analysis and Management*, 28(3): 440–462.

Curry. "Court Condemns Hopkins-Approved Study That Exposed Children to Lead Dust," *The Chronicle of Higher Education*. Sept. 7, 2001, 1–4.

Ellen, I. G., Schwartz, A. E., Voicu, I., Schill, M. H. (2007). Does federally subsidized rental housing depress neighborhood property values? Journal of Policy Analysis and Management, 26(2), 257–280.

Fernandez, S., Smith, C. R., & Wenger, J. B. (2006). Employment, privatization, and managerial choice: Does contracting out reduce public sector employment? Journal of Policy Analysis and Management, 26(1), 57–77.

Gilbert, N. (1997). Advocacy Research and Social Policy. Crime and Justice 22, 101–148.

Greenbaum, R. T., & Engberg, J. B. (2004). The impact of state enterprise zones on urban manufacturing establishments. *Journal of Policy Analysis and Management*, 23(2), 315–339.

Haider, S. J., Schoeni, R. F., Bao, Y., Danielson, C. (2004). Immigrants, welfare reform, and the economy. *Journal of Policy Analysis and Management*, 23(4), 745–764.

Harknett, K. (2006). Does receiving an earnings supplement affect union formation? Estimating effects for program participants using propensity score matching. *Evaluation Review*, 30(6), 741–778.

Harrison, K., & Antweiler, W. (2003). Incentives for pollution abatement: Regulation, regulatory threats, and non-governmental pressures. *Journal of Policy Analysis and Management*, 22(3), 361–382.

Houston, D. J., Richardson, L. E., Jr. (2006). Reducing traffic fatalities in the American states by upgrading seat belt use laws to primary enforcement. *Journal of Policy Analysis and Management*, 25(3), 645–659.

Joyce, T., Gibson, D. & Colman, S. (2005). The Changing Association Between Prenatal Participation in WIC and Birth Outcomes in New York City. *Journal of Policy Analysis and Management*, 24(4), 661–685.

King, G., Gakidou, E., Ravishankar, N., Moore, R. T., Lakin, J., Vargas, M., Téllez-Rojo, M. M., Ávila, J. E. H., Ávila, M. H., & Llamas H. H. (2007). A "politically robust" experimental design for public policy evaluation with application to the Mexican universal health insurance program. *Journal of Policy Analysis and Management*, 26(3), 479–506.

Loeb, S., Fuller, B., Kagan, S. L., Carrol, B. (2003). How welfare reform affects young children: Experimental findings from Connecticut- A research note. *Journal of Policy Analysis and Management*, 22(4), 537–550.

Lopoo, L. M. & DeLeire, T. (2006). Did Welfare Reform Influence the Fertility of Young Teens? *Journal of Policy Analysis and Management*, 25(2), 275–298.

Marton, J. (2007). The impact of the introduction of premiums into a SCHIP Program. *Journal of Policy Analysis and Management*, 26(2), 237–255.

Nathan, Richard P. "The Ethics of Federal Social Program Evaluation." *Journal of Policy Analysis and Management*. Autumn 2005, 823–824.

National Science Foundation (2009). *Academic Research and Development Funding Expenditures*. Accessed on March 30, 2010 at http://www.nsf.gov/statistics/nsf09303/pdf/nsf09303.pdf.

New York Times. (July 10, 2007). Surgeon General Sees 4-Year Term as Compromised. Accessed on August 25, 2007 at: http://www.nytimes.com/2007/07/11/washington/11surgeon.html?ex=1200974400&en=7c098e508d090c29&ei=5087&excamp=mktat6

Niu, S. X. & Tienda, M. (2010). "The Impact of the Texas Top Ten Percent Law on College Enrollment: A Regression Discontinuity Design" *Journal of Policy Analysis and Management* 28(1): 84–110.

Page, M. E., Spetz, J., & Millar, J. (2007). Does the minimum wage affect welfare caseloads? *Journal of Policy Analysis and Management*, 24(2), 273–295.

Pirog, M.A., Buffardi, A.L., Chrisinger, C.K., Singh, P. & Briney, J. (2009) Are the alternatives to randomized assignment nearly as good? Statistical corrections to nonrandomized evaluations. *Journal of Policy Analysis and Management* 28(1): 169–172.

Pirog, M. A. & Kioko, S.N. (2010). Evaluation of the Education Sector Enhancement Program in Barbados. *International Public Management Journal.* 13(1): 72–99.

Reingold, David A. (2008). Can government-supported evaluation and policy research be independent? *Journal of Policy Analysis and Management*. 27(4): 934–941.

Rolston, Howard. "To Learn or Not to Learn," *Journal of Policy Analysis and Management*. Autumn 2005, 848–849.

Schochet, Peter Z. "Comments on Dr. Blustein's Paper, Toward A More Public Discussion of the Ethics of Federal Social Program Evaluation," *Journal of Policy Analysis and Management*. Autumn 2005, 849–850.

Shanahan, M., Lancsar, E., Haas, M., Lind, B., Weatherburn, D., Chen, S. (2004). Cost-Effectiveness of the New South Wales Adult Drug Court Program. *Evaluation Review* 28(3): 3–27.

Stolzenberg, L., & D'Alessio, S. J. (2003). "Born to be wild": The effect of the repeal of Florida's mandatory motorcycle helmet-use law on serious injury and fatality rates. *Evaluation Review*, 27(2), 131–150.

Stolzenberg, L., & D'Alessio, S. J. (2007). Is nonsmoking dangerous to the health of restaurants? The effect of California's indoor smoking ban on restaurant revenues. *Evaluation Review*, 31(1), 75–92.

Trenholm, C., Devaney, B., Fortson, K., Clark, M., Quay, L. & Wheeler, J. (2008). Impacts of abstinence education on teen sexual activity, risk of pregnancy, and risk of sexually transmitted diseases. *Journal of Policy Analysis and Management* 27(2): 255–276.

Trujillo, A. J., & McCalla, D. C. (2004). Are Colombian sickness funds cream skimming enrollees? An analysis with suggestions for policy improvement. *Journal of Policy Analysis and Management*, 23(4), 873–888.

Walker, R. M., & 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*, 25(2), 371–393.

White, V. M., Hill, D. J., & Effendi, Y. (2004). How does active parental consent influence the findings of drug-use surveys in schools? *Evaluation Review*, 28(3), 246–260.

Wilde, E. T., Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management,* 26(3), 455–477.

Wu, Y. (2005). The effects of state R&D tax credits in stimulating private R&D expenditure: A cross-state empirical analysis. *Journal of Policy Analysis and Management*, 24(4), 785–802.

Theory-Driven Evaluation: Tracing links between assumptions and effects[1)]

Karol Olejniczak

1. Introduction

"There is nothing more practical than a good theory". This remark made by Kurt Lewin (1952, p.169) – founder of social psychology and organizational development studies– became common sense wisdom.

The statement is also true for our undertakings in the field of public policy. Relating our projects and programmes to the socio-economic theories gives us two advantages. Firstly, it shows us the "bigger picture", helping us to discover the pattern behind our everyday actions. Secondly, it connectsthe experience and initiatives of our organizations with a rich pool of earlier knowledge, research and experience coming from the particular field.

These two things in turn substantially change the type of reflection we undertake in reference to our actions. We switch from merely considering "are we doing things right?", which is a narrow technical efficiency perspective, to the effectiveness question of "are we doing the right things?", which is a critical, strategic reflection on the direction and real utility of our actions. The classic literature on organizational learning labels these two perspectives as single-loop vs. double-loop learning (Argyris, 1977). The former is purely mechanical while the latter is the core of an organization's survival and adaptation to the changing environment (Morgan, 2006, p.83–86).

The evaluation of public interventions is commonly defined as a tool of public sector learning (Dahler-Larsen, 2005; Furubo *et al*., 2002). In the case of summative (effects) evaluations, both Cohesion Policy regulations (Council of the European Union, 2006, art.47.2) and the literature (e.g. Bachtler, Wren, 2006; Ferry, Olejniczak, 2008, p.24) underlay its function as a strategic reflection and lessons learning for the future. Thus, the primary aim of evaluation of effects is double-loop learning.

The objective of this chapter is to discuss the practical approach that facilitates this type of strategic reflection and learning. We argue that this suitable approach is

[1] This chapter is an updated version of the paper presented at The Sixth European Conference on Evaluation of Cohesion Policy, Warsaw, November30 – December 1, 2009. Author would like to thank all the participants of the Workshop 3 *Revealing the causal chain* that provided him with valuable comments to this paper, in particular to chairperson prof. FransLeeuw from Maastricht University and Laura Polverari from European Policies Research Center.

Theory-Driven Evaluation (TDE). The chapter consists of four parts. It begins with an introduction of American origins and definitions of the main TDE concepts. In the second part, the practical application and toolbox is discussed. The TDE analytical sequence for cohesion policy programmes is also proposed.[2)] The third part gives a real life example of TDE in the context of structural funds in Poland. This is a case study of an *ex post* evaluation of the Neighbourhood Programme INTERREG-TACIS CBC Poland-Ukraine-Belarus 2004–2006. The last part is devoted to an assessment of the advantages and limitations of Theory-Driven Evaluation for the Cohesion Policy programmes.

2. Origins and definition of the Theory-Driven Approach

"Theory-Driven evaluation" (TDE), also called "Theory-based evaluation", initially appeared in the United States in the middle of 1980s as a reaction to the growing disappointment with the narrow, bureaucratic evaluation of inputs-outputs (Bickman, 1987; Chen, Rossi, 1983).[3)] The main promoter of this approach, Huey-Tsych Chen, pointed out that evaluation studies were avoiding any references to theories or even wider concepts when formulating elements of intervention, its rationality or causal relation. In his opinion this narrow practice was limiting the utility of studies' results (Chen, 1994, p.17–18).[4)] Simple input-output comparison told little about the mechanism of change and ignored side effects (that is, effects not included in the formal list of objectives) as well as the organizational and political context. What is more, decisions made on the basis of this type of study about the future of the programmes in question would be highly exposed to both type I and type II errors.[5)]

The main weakness of these evaluations was that they never explained "why" the success or failure took place. Was it the wrong concept of the programme, its scale, its implementation mechanism or maybe changes in the context? This diagnosis presented by Chen for the American evaluation in the mid-1980s seems surprisingly valid for current Cohesion Policy studies (compare: Olejniczak, 2009).

---

2 In this chapter I refer mainly to the programme level since most European Union evaluation activities are conducted for programmes or their parts. However, it has to be noted that the presented TDE approach can be equally well applied to different levels of public interventions, such as single projects, group of projects, thematic research or even whole policy.

3 As Leeuw points out, the origins of the core concept "programme theories" can be traced back to the classical sociology works of Karl Mannheim in the 1930s and Jacques van Doorn in the 1950s (2003, p.5-6).

4 Chen supported his thesis with the results of the meta-evaluation of 175 studies published in the 1980s in leading research quarterlies (Lipsey *et al*., 1985). Those American evaluations turned out to be atheoretical and without deeper analytical structure.

5 The terms are borrowed from statistics. Type I error is a situation in which the evaluator concludes that the programme caused the effects while in fact the positive change occurred as a result of contextual factors (the programme itself has no impact). Type II error is a situation in which the researcher concludesthere was an absence of programme effects, while in reality the effects appear but they were misplaced in time or space (Daponte, 2008). Obviously in both situationsa decision made on the bases of these wrong judgements could be socially harmful. In the first case the ineffective programme will continue and funds would be lost. In the second case a programme that works and changes the social situation could be terminated.

As an alternative solution Chen proposed an approach that would allow structuring a systemic thinking about the programme (both questions and research design) in the form of theory of public intervention.

TDE has been dynamically developing since the mid-1990s and is currently one of the main evaluation approaches. There have been a number of follow-ups to the original concept that enriched it with interpretation in different thematic contexts (compare: special edition of New Directions for Evaluation – Petrosino *et al*., 2000). Apart from authors such as Chen (2004) and Donaldson (2007), TDE is used by "classics" of American evaluation – C.H. Weiss (1997), M.Q. Patton (2008) and P.H. Rossi (Rossi *et al*., 1999). References to TDE can also be found in the work of European authors (e.g. Górniak&Keler 2008, p.117; Leuuw, 2003) and promoters of realistic evaluation (Pawson, 2002).

Moving to the nuts and bolds of TDE, the term clearly suggests that "theory" plays a central role in the approach. The first straightforward interpretation is that evaluation research should relate its content and findings to wider socio-economic theories. While this is true, TDE goes much further. In fact it treats the whole intervention as a theory – a hypothesis yet to be tested and verified in a real-life situation.[6)] This view is in line with classics of policy-decision makings that perceive policies as trial and error problem-solving processes (Bardach 2006, p.350).

Theory-Driven Evaluation is "*a contextual or holistic assessment of a programme based on the conceptual framework of programme theory. Its purpose is to provide information on not only the performance or merit of the programme but on how and why the programme achieves such a result*" (Chen, 2005, p.415). To put it simply, "Theory-Driven Evaluation" not only shows the real effects of the programme but also explains why they appeared. Thus, TDE is useful both for effects assessment (accountability function) as well as for learning about the mechanisms behind an interventions' success or failure (cognitive function of evaluation). In fact, the real value of TDE, especially for decision-makers, is its explanatory side. It reveals how a programme works, with whom and under what circumstances (Astbury&Leeuw 2010, p. 365), and explains the mechanisms that led to the success or failure of the intervention.

As pointed out by Weiss (2004), the paradox is that so far evaluators have not developed a concise and unified understanding of the "programme theory". Patton (2008, p.339) indicates that the interpretation of what "programme theory" is depends on the context of a given research, the scale of the intervention in question and scientific background of the evaluator.

---

[6] It is worth noting that the term "Theory" comes from Greek and has been in the English language since the 16th century. In its first meaning it is "the analysis of a set of facts in their relation to one another", while the second meaning is "an abstract thought – a speculation, an unproved assumption" (compare: Encyclopaedia Britannica & Merriam-Webster Dictionary).

Based on international literature (Chen, 2005; Donaldson, 2007; Pawson, 2009), the following clarification of terms and definitions can be proposed for the purposes of the Cohesion Policy. Each cohesion policy intervention consists of four types of elements and assumptions:

Context – it is a socio-economic, institutional, legal and political environment in which a programme operates (Pawson 2009).

Underlying Theories – consists of knowledge, earlier experiences, assumptions, as well as socio-economic theories that became inspirations (consciously or intuitively) for the people or group of stakeholders involved in the creation of a programme.

The theory of change is also called "conceptual theory" (Chen, 2004, p.257) or "programme impact theory" (Donaldson, 2007, p.25). It is an assumption about the causal relationship between strategy and its expected effect. This is how planned actions and outputs bring the expected socio-economic change (results and impact) at which the programme aims. It is usually described in the form of a statement "if ….then…".

The theory of Implementation is also called "intervention process theory" or "action theory" (Bamberger *et al*., 2006, p.175; Donaldson, 2007, p.25). It is an operational strategy of changing inputs into outputs. In other words, it is an assumed sequence of technical, day-to-day implementation activities and institutional arrangements required to fulfil the programme's goal.

Each of these three theories has to be indentified, defined, applied and then verified in theparticular socio-economic **context** in which the intervention operates. The main message is that "*a programme is liable to have mixed effects patterns which will comprise the intended and unintended consequences of programmes. Programmes will work for some – in a certain situation, time and place, while fail for others*" (Pawson 2009).

All four of the above elements are part of "programme theory". To sum up, TDE suggests that every programme can and should be described by the following logical chain of statements:

1. take into account the situation and environment in which we operate [Context]…
2. be based on our earlier experiences, current knowledge and findings of modern socio-economic theories [Underlying Theories]
3. assume that IF we do X THEN we will get a certain Y effect that will fulfil our needs/solve our problems [Theory of Change].
4. The most efficient way to move from X to Y will be the implementation of a sequence of actions A, B and C [Theory of Implementation].

The example of this "template" of thinking is presented in the exhibit below.

Exhibit 1. Thinking template of TDE – an example of the Regional Innovation System project

One of the key problems identified in Polish region X is its low competitiveness. The in-depth diagnosis revealed that the main root of the problem is low innovativeness of local firms. Based on modern regional development theories of Regional Innovation Systems and clusters (Cooke, 2004; Porter, 1998), as well as practical experiences of other European regions that implemented Regional Innovation Strategies [Underlying Theory], decision-makers assigned money for the Regional Innovation System project. They assumed that if RIS will strengthen trust, interactions and cooperation between firms and research institutions then the innovativeness of companies will increase, which in turn will lead to an increase of regional competitiveness [Theory of change]. Furthermore, it has been decided that the most efficient RIS construction should have the following sequence of actions [Theory of Implementation]:

1. To conduct together with key regional stakeholders an analysis of regional innovation potential;
2. To develop a joint innovation strategy and choose the priority sectors;
3. To establish an internet platform and other simple ways of communication between key participants of the process: companies, research institutions and regional authorities.

3. Toolbox of Theory-Driven Approach

It is important to understand that Theory-Driven Evaluation and its "programme theory" is not a specific method of evaluation but an approach to thinking and organizing an evaluation research – its questions, research strategy and methodology (Frechtling, 2007, p.5). The main advantage of this approach is that, in principle, it is method-neutral (Donaldson, 2007, p.11). It stays outside of the paradigm war between positivists and constructivists (see: Christie, 2008). It can be a guiding rule both for quantitative or qualitative strategies, quasi-experimental or mix-method designs. The choices depend on: the type of evaluation (*ex ante*, on-going or *ex post*), the scale of the intervention as well as the researcher's background and stakeholder preferences.

Regardless of methodological choices two things stay valid for all the options. First is the use of logic models – a tool crucial for the clarity of Theory-Driven Evaluation. The second is a sequence of analysis. Both of these are discussed in this section.

Logic models

Logic models are simply a graphical illustration of the causal chain assumed in the particular intervention (Frechtling, 2007, p.1) and a concise description of how a programme is supposed to work. "*They provide a clear roadmap to a specific end*" (Knowlton, Phillips, 2008, p.5).

The origins are usually traced back to the work of Suchman in the 1960s, in which he recommended to present public interventions as a chain of three types of objectives – immediate, intermediate and ultimate goals to which three types of effects

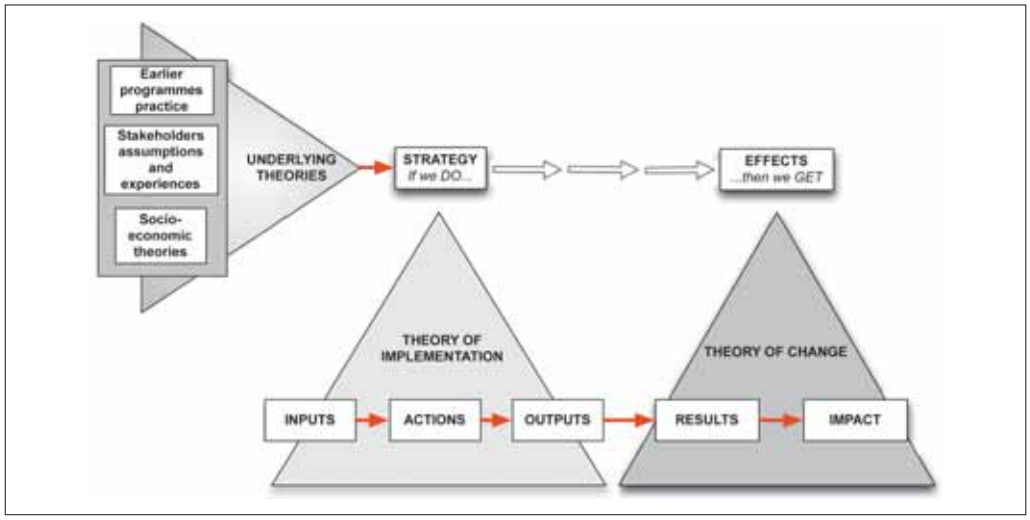
could be assigned: outputs, short-term outcomes and long-term impacts (Patton, 2008, p.340; Rogers, 2005, p.232).

The first models showed a simple path from inputs to outputs (inputs → actions → outputs → outcomes) (United Way, 1996). However the most popular interpretation of the basic model used to have two clear blocks (what vs. why) and two types of effects (outcomes and impacts) (Kellogg Foundation, 2004).[7)]

This logical chain remains the current standard of logic models, though over time some modifications appeared. For example, EU models (EC 2000) added two elements from the beginning: a starting point and responding strategy (Needs/Problems → Objectives → Inputs → Operations → Outputs → Results → Impacts). Some American authors explicitly add "customers" after outputs and distinguish three types of effects (short-term, intermediate and log-term outcomes) (Mc Laughlin, Jordan, 2004, p.9).

The latest literature clearly connects logic models with Theory-Driven Evaluation and, moreover, postulate in the case of complex interventions to apply two separate models i.e. those that show the bigger picture – the strategic assumption (Theory of Change) and those that illustrate the details of a plan (Implementation Theory) (Bamberger *et al*., 2006, p.175–179; Knowlton, Phillips, 2008).

Exhibit 2. Using the logic model for Theory-Driven Evaluation



Source: modified models from Knowlton & Phillips (2008) p.23-23, p.36, p.45

The simplified way of combining logic models for Theory-Driven Evaluation has been presented in Exhibit 2. Of course it has to be noted that the schemes are always simplified. The models built in practice do not have to follow one graphical pattern.

[7] The first block was a process one. It was a planned work and it included resources/inputs and activities. The second block was on intended effects and it consisted of outputs, outcomes and impact. These two blocks correspond to the "Theory of Implementation" and "Theory of Change" (Knowlton, Phillips, 2008, p.5).

Their shape, structure and level of complexity are determined by the type and purpose of the evaluation as well as the characteristic of the evaluand.[8)]

Analytical sequence

American authors propose a simple three-stage analytical sequence of "Theory-Driven Evaluation" (Donaldson, 2007, p.20–49; Chen, 2004):

Step 1: Developing "programme theory" and presenting it in a graphical form –usually in two models: the model of "Theory of Change" and the model of "Theory of Implementation";

Step 2: Based on models created, formulating a set of evaluation questions and then prioritizing them;

Step 3: Addressing the questions of an evaluation, that is conducting empirical research (data collation, analysis and assessment and conclusions).

The initial step – building programme theory – is crucial in the whole procedure. It becomes a roadmap for later analysis. Patton distinguishes three approaches to programme theory development (2008, p.344–346):

- A deductive approach – building on scholarly theories and earlier research results from academic literature;
- An inductive approach – doing fieldwork on a programme to generate a grounded theory (usually as part of evaluability assessment);
- A user-focused approach – working with intended users to extract and make explicit their implicit underlying theories as well as their theory of change.

When comparing the pros and cons of each approach, Patton points out out the third one as most effective. Donaldson (2007, Part 1) – despite his strong inclination for using scholarly theories, also underlines the importance of stakeholder involvement.

It seems that in the case of Cohesion Policy interventions this order cannot find a direct application for two reasons. First of all, in the case of *ex post* evaluation the process of "building the programme" is done in reverse order. We do not build theory from scratch; we reconstruct it on the basis of things already done, that is *post factum*.

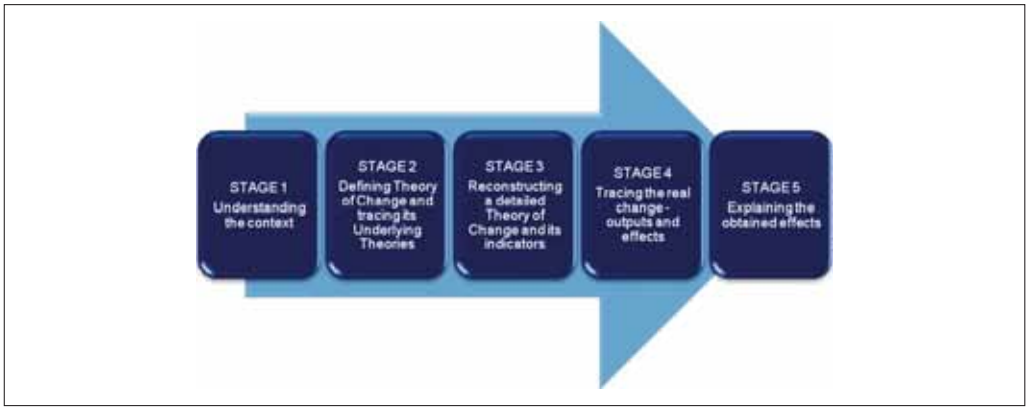
Second, the complexity of cohesion policy programmes requires a clear split between the main, strategic assumption of causality and the set of detail "theories of change". The former is "if/then" logic of strategic objectives (the structure of

[8] Evaluand is a term introduced by Michael Scriven to describe any object of an evaluation study. Thus, „evaluand" may be an idea, group of persons, process, project, programme or even policy (Mathison, 2005, p.139).

Actions/Priorities of the programme), while the latter is an assumption within each of the Programme priorities.

Therefore, it seems rational to adapt TDE to anex post assessment of cohesion programmes by applying a modified, five stage procedure. This is summarised in Exhibit 3 and discussed in the following paragraphs.

Exhibit 3. Five stages of Theory-Driven Evaluation for cohesion programmes



Stage 1: Understanding the context

The starting point is to understand the context of the intervention – the policy field, spatial and socio-economic situation in which the particular programme has been operating. This means addressing two questions:

- What do wider socio-economic theories tell us about the field in question?
- What was the political, institutional and socio-economic context in which the programme has been operating?

In order to address the first question, a systematic, in-depth literature review of the particular field is required. The procedure for the literature review follows the standards of social science research (see for example: Hart, 1998), though the time pressure is much greater. Some technical solutions can facilitate this process, such as academic databases of scholarly periodicals (e.g. EBSCO, JSTOR, Emeralnd), bibliographical software for sources collection (e.g. Endnote, Bookends, Sente) or mindmapping tools for analysis and grouping of the literature sources, motives and ideas (e.g. MindManager, NovaMind, Freemind).

In order to address the second question, evaluators have to conduct desk research (statistical data, earlier research, documents and newspapers articles related to the programme or policy field in concern),[9)] as well as conduct interviews with

[9] A very good source of spatial data presented in a comparative EU-wide perspective is ESPON observatory: http://www.espon.eu/

key programme stakeholders. It has to be noted that analysis conducted should be dynamic, not static – meaning it should payspecial attention to trends and changes that took place in the analysed context over the period of the programme's implementation.

Stage 2: Defining the main Theory of Change and tracing its Underlying Theories

At this stage evaluators try to grasp and articulate the strategic assumption of the programme (*if we do… then we get…*), and then try to identify the assumptions or inspiration that became a cornerstone of this strategy [Underlying Theories]. Three main questions that have to be addressed are:

- What was the expected socio-economic change that the programme was suppose to bring?
- Why have strategic objectives been chosen and structured in this particular way?
- How do these choices correspond to what we have learnt from wider socio-economic theories?

The methods used at this stage mainly include desk research of the programme documents, interviews with key decision-makers and the stakeholders involved in the creation of the programme and negotiations of its objectives, a critical comparative analysis between data collected and findings of the literature review from Stage 1. The presentation of a programme's strategic assumption in the form of a logic modelhelps to facilitate interviews and confront the picture coming from the documents with the assumptions taken by stakeholders.

Stage 3 – Reconstructing a detailed Theory of Change and its indicators

This is a crucial stage for the whole research. Evaluators try to clearly define the causal assumptions of the programme strategy. The key research questions include:

- How could the effects of the programme in each field of its activity (i.e. priorities, actions)be defined with measurable indicators and assessment criteria?[10)]
- How have the expected effects been defined and described in terms of concrete indicators by programme authorities?

The main toolst hat are usefulat this stage of the research are logic models, experts' knowledge and a review of the programme monitoring system. The rules of logic modelling are well discussed in literature (Frechtling 2007; Knowlton & Philips 2008),

[10] This is the crucial challenge of describing the socio-economic change in a measurable way. In practice, what helps evaluator to pin down "change" is to address the question *"How will I know it when I see it?"*.

but it is good practice to underlay a central one – the model needs to keep a balance between information coverage (parts of the model should be comprehensive and self-explanatory) and visual clarity.[11] When building logic models, evaluators should remember that for *ex post* studies they should rather examine the value of the programme by investigating what the programme has really done rather than what it has been trying to do (so-called: *goal-free evaluation*)[12] (Mathison 2005, p.171). In practice this means that evaluator should not to stick too closely to the formal detailed objectives and indicators of the programme but rather treat them as indicators of the fields of interventions and instead, develop his own ways of grasping the real effects. Thanks to this, evaluators avoid narrowing down their perspective and focus on real, observable effects, instead of only the manager's assumed intentions.

Patton (2008 p.336) also points out that at this stage it is absolutely necessary to clearly define all the terms used in the programme documents that describe the effects. Special attention should be paid to terms indicating value or improvement (e.g. good, appropriate, sufficient, etc.).[13]Evaluators often find out that there has been no clear consensus in programme documents on what the particular terms really mean. This requires an additional interview or even workshops with key stakeholders on clarifying their understanding of the expected change.

Stage 4 – Tracing real change – outputs and effects

At this stage two general questions are addressed:

- What are the real (that is obtained) outputs of the programme?
- What are the real (expected and unexpected) outcomes of the programme?

For the first question, analysis is based on the summary of monitoring data. For the regional development programmes, products (the number and structure of projects and beneficiaries) should preferably be presented on both charts and maps. The second tool allows to dwell on spatial considerations of the cohesion policy. Of course the smoothness and quality of this analysis relies on the quality of the monitoring data.

---

[11] The clarity of the graph can be enhanced though the use of advanced software (e.g. Visio, Omnigraffle). The latest innovation in this field, introduced by Claremont Graduate University (Azzam, 2009) include an interactive, animated multilayer scheme (developed with the use of Swishmax2 or Adobe Macromedia Flash), which allows to apply a "nested approach" that moves between different levels of details, zooming in and out onto the complex parts e.g. Actions, Priorities, groups of projects.

[12] This term has been developed and popularised by Michael Scriven.

[13] For example, statements such as "substantial increase of the innovation potential", "better jobs", „improvement of the chances at the labour market" require thorough definitions. What should the expected characteristic be of a "better job?" What elements consisted of innovation potential and how much do they need to increase in order to accept it as "substantial". What is hidden under "market chances" in the context of this particular labour market? Does it cover only basic IT skills or also psychological support and improvement in self-esteem?

Tracing effects (results and impact) is the most challenging and time-consuming part of evaluation. Evaluators try to find real-life proofs for the Theory of Change, that is a causal relationship that connects actions and outputs with results and the wider, structural change (impact).

Data collection strategy depends on the topic. However in practice, what works best is a down-to-earth "follow the money" strategy. This means local visits of the projects sites, interviews and, if projects were implemented on a larger scale, social surveys. In the case of programmes aiming at spatial impact, it is worth conducting a review of popular regional or local press and case studies of particular communities. This strategy has two advantages. First it allows for assessing programmes even when the monitoring data are limited. Secondly, it encourages thinking outside of the box of the theory of change assumed by programme stakeholders. In other words, it allows for discovering side effects of the intervention.

For the purposes of the analysis, Chen suggests quantitative methods, putting a special emphasis on quasi-experiments (Chen, 2004). Of course this type of procedure usually requires a "before and after" comparison. A number of measurement strategies can be applied here (see for example: Russ-Eft, Preskill, 2001, p.147–176; Reichardt, Mark, 2004).

The scale, type of intervention and data availability determine the choice of the particular research plan. It is worth pointing out, however, that based on U.S. and EU practices, control groups are applied mainly for the human resource programmes and support of enterprises.[14)]When it comes to the situation when the subject of intervention was a spatial unit (region, community) the rational strategy seems to focus on final beneficiaries or thematic parts (segments) of the interventions, analyzing desegregated segments of the programmes and then trying to combine them at the final stage of the research.

Stage 5 – Explaining the determinants of change

The question addressed at this stage is straightforward: what are the factors and mechanisms that explain the obtained effects of the programme?

Usually four sets of hypotheses need to be taken into account:

- Context
- Logic of assumptions about change
- Delivery mechanism (Theory of implementation)
- The characteristics of beneficiaries

The context hypothesis includes previously discussed elements of the programme environment (the socio-economic situation, political constraints (Bamberger *et al*.,

[14] The conditions of quasi-experiments and counterfactual analysis are discussed in depth in a later section of this book, in the chapter written by Professor Alberto Martini.

2006, p.175–176). All these issues can influence the causal chain. The data used for the hypothesis testing include all the information collected in the first stage of research. Some authors point out that even the relation between output and result does not have to be direct (especially in the case of big programmes). They introduce the terms mediators and moderators. The former are factors that influence the implementation of its effects and can be modified by the intervention itself.[15)]The latter are qualitative (e.gsex, origins) or quantitative (level of motivation, experience) factors that impact the direction and strength of the relations between cause and effect (Donaldson, 2007, p.27–32).

In many cases programmes do not work as expected simply because the assumption of obtaining a certain change with a certain product was wrong from the start. This is especially the case with more complex programmes involving a number of objectives. This issue can usually be traced by use of logic models. Looking at the graphical representation of the needs – actions – effects tree allows us to trace illogical connections and doubtful assumptions or even contradicting actions.

The delivery mechanism hypothesis is a verification of the extent to which the Theory of Implementation worked in practice – meaning how inputs have been transferred into products. This includes the analysis of all the information about programme promotion, application process, implementation procedures, institutional settings and real everyday proceedings. Apart from documents, the main information comes from interviews with programme and project managers. Surveys of potential beneficiaries and unsuccessful applicants can also be a valuable source.

The characteristic of beneficiaries, their institutional and organizational constraints and capabilities are the last explanatory factor. Again the key sources of data are surveys combined with comparative analysis of project documents (project application and final reports). The use of case studies may not be taken as reliable proof for the programme's general conclusions, but can nevertheless bring very revealing insight into the beneficiary perspective, their contextual constraints and implementation mechanisms.

The discussed 5 stage procedure can be treated as a roadmap, a template to be adapted for the particular real-life situation. An example of its application is given in the next subchapter.

[15] Bamberger (2006, p.176) gives an example of micro-credits for women as part of a small business development programme that aims at activating housewives. The mediator in this case is the attitude of husbands. If they are open to the idea of wives working outside of the home and be supportive for their idea of business, the effects of the programme will be stronger.

4. TDE in practice – An example fromPoland

In this section the application of Theory-Driven Evaluation approach is presented using the example of *ex post* evaluation of the Neighbourhood Programme INTERREG-TACIS CBC Poland-Ukraine-Belarus 2004–2006.

The study was contracted out by the Polish Ministry of Regional Development, Department of Territorial Cooperation. This institution was themanaging authority of the programme. A study[16)] was executed by the company EGO s.c. (Evaluation for Government Organizations) over the period of 5 months (July-November 2009). The main objective of the research was to assess the impact of the programme on economic, social and territorial cohesion at the cross-border area covered. This included an explanation of the mechanism responsible for the scale and quality of this impact. The contracting authority also clearly expressed that the lessons learnt will be applied in the process of designing the new cooperation forms.

Before we move to the presentation of TDE application it is worth adding a few details about the evaluand (the programme in question). The Poland-Belarus-Ukraine (PBU) programme aimed at improving the quality of life and socio-economic integration of the neighbouring regions. It consisted of 3 priorities. The structure and financial allocations are presented in the table below.

[16] The research team was led by Karol Olejniczak, PhD and Bartosz Ledzion. The team consisted of: Anna Domaradzka-Widła PhD, Elżbieta Kozłowska, Katarzyna Krok, Andrzej Krzewski, Adam Płoszaj, Łukasz Widła-Domaradzki, Katarzyna Wojnar, Michał Wolański and Katarzyna Zalewska.

Exhibit 4. Structure and funds of Poland-Bielarus-Ukraine Neighbourhood Programme 2004–06

Programme's Priorities and Measures	ERDF allocation (mln EUR)	Tacis CBC allocation (mln EUR)
Priority 1 Increasing the competitiveness of the border area through the modernisation and development of the cross-border infrastructure	**21,447**	
Measure 1.1 Modernisation and development of the existing cross-border transport network to increase accessibility	11,111	
Measure 1.2 Developing a common cross-border natural environment protection system	6,649	
Measure 1.3 Developing tourism: business and infrastructure	3,717	
Priority 2 Developing human capital and institutional cross-border cooperation including security at the European Union's borders	**14,068**	**8.0**
Measure 2.1 Strengthening of cross-border institutional cooperation and enhancing the quality of human capital	7,034	
Measure 2.2.: Support for local communities initiatives (Micro-project Fund)	7,034	
Priority 3 Technical assistance	2,272	
Measure 3.1.: Management, implementation and control	1,515	
Measure 3.2.: Programme information, publicity and evaluation	0,756	
TOTAL	**37,818**	**8.0**

Source: Neighbourhood Programme Poland-Belarus-Ukraine INTERREG IIIA-TACIS CBC 2004-06 – Programme Complement (version 20.12.2006), p.60-61, www.interreg.gov.pl

The programmes covered the border regions of Poland (8 NUTS III regions), Ukraine (3 regions) and Belarus (also 3 regions). The list of potential beneficiaries included local authorities, public services (e.g. health service, fireman, police, border guards) educational and cultural institutions, as well as NGOs. Polish participants were co-financed from INTERREG funds while Ukrainian and Belarusian partners were supported from very modest TACIS-CBC resources.

When evaluation started, the programme was in its final stage (all of the funds had been allocated) although at the project level some of the activities were still running. By the time of the research, 173 major projects were implemented (including 80 hard-infrastructure oriented projects, 84 "soft"-institutional projects and 9 umbrella micro-projects funds). The vast majority of them were projects with Polish

institutions as leaders. Within the umbrella projects, there was a population of over 300 successfully implemented micro-projects.

As we can see the programme is a model example of the Cohesion Policy initiative: it has a complex, multi-sectoral structure of objectives, it aims at very wide and diverse beneficiaries, it is spatially extensive and funds are relatively limited considering the large geographical areaand ambitious objectives.

In order to tackle this complex subject in a structured and organized way, evaluators decided to apply Theory-Driven Evaluation. The summary of the approach, together with the main findings is presented in Exhibit 5. These five stages directly corresponded to the management of the research as well as the structure of the final report, though the final report did have one additional chapter with suggestions for the future.

5. Advantages and limitations of Theory Driven Evaluation

In this final part of the chapter, the advantages and limitations of using TDE in cohesion programmes are summarized.

When it comes to advantages, we can distinguish six main arguments that support the utility of Theory-Driven Evaluation in the studies of the effects of cohesion policy interventions.

First, Theory-Driven Evaluation articulates well the rationality of the programme (or weakness in it) – the main assumption about the particular social change (Weiss, 2004). This is especially valuable for cohesion programmes that have complex structures of objectives developed by a number of institutions. The description of the programme in terms of hypothesis introduces some clarity in the thinking about a programme and its functions.

Second, TDE structures thinking about a programme as a subject of the research and provides solid foundations for the conceptual part of the work. This enables:

- putting forward a clear research problem (the verification of the programme – hypothesis),
- maintaining scientific discipline and focus during the research because TDE gives a guiding structure for data collection and analysis,
- creating a clear narrative for the final report (5 main chapters that explore following issues: exploring context, revealing underlying inspirations, reconstructing expected change, verifying the real changes, explaining the observed situation, success or failure).

Exhibit 5. Application of TDE in the *ex post* evaluation of Poland-Belarus-Ukraine Neighbourhood Programme 2004-06

Stage of the research	What was done? – Questions addressed	How did we do it? – Methods	What did we find out? – Findings
Stage 1: Understanding the context	What do theories tell us about cross-border cooperation? What type of border is it – looking from an EU-comparative perspective? What was the scale of the intervention in comparison to other financial incentives available?	Literature and research review on CBC cooperation and theories of regional development and peripherality Review of the findings of earlier CBC evaluations and expertise Analysis of general statistics (GDP, population, employment, infrastructure network, border flows, financial statistics) Review of the socio-political situation	Cooperation depends on the permeability of a border, the difference in the development of partners as well as their complementing parts. On highly peripheral areas, when all partners are weak and the border is a barrier, cooperation does not take place. In such as case only support of indirect determinants of territorial cooperation (such as good governance, local civil society and joint search for complementarities) can slightly increase its intensity. The PBU border area is a highly peripheral space, underdeveloped and isolated, with limited development potential and substantial deficits of basic infrastructure. The existing border between regions is a clear barrier (visa requirements and limited capacity of border crossings). The PBU programme was a minor financial impulse for Polish regions (only 2.2% of EU support), while for UA & BY it was one of the few external resources. Earlier research on similar interventions indicate that such a scale of investment would have no regional impact. Effects should be visible and traceable only at the lowest local level.
Stage 2: Defining the main Theory of Change and tracing its Underlying Theories	What strategic objectives have been designed? What were the assumptions and sources of inspiration for these decisions?	Programme documents analysis Interviews with key programme stakeholders involved in the programme design	A dual objective was set: "IF we act in five thematic fields THEN the quality of life AND socio-economic integration will be improved". This assumption followed the INTERREG philosophy as well as theories of cross-border cooperation, which state that in highly peripheral areas a certain level of development must first be reached by all partners in order to allow integration. Set of 5 actions –fields of interventions was chosen (see Exhibit 4 – prog. measures). This rationale was in line with the condition of the border as well as experiences coming from earlier CBC programmes (e.g. Poland-Germany).

Stage of the research	What was done? – Questions addressed	How did we do it? – Methods	What did we find out? – Findings
Stage 3: Reconstructing a detailed Theory of Change and its indicators	How would evaluators define the change in each of the fields targeted by the programme? How did the programme designers define these changes (sets of indicators)?	Logic models developed for each thematic group: infrastructure, environment, tourism, institutional development (for example, see Annex 1). Review of the system of indicators used for Programme monitoring (SMART method)	Planned effects and side-effects have to be traced in 3 dimensions: · Thematic – groups of projects (infrastructure, environment, tourism, institutional development, local initiatives) · Relational – partnerships of leaders and their partners · Territorial – selected local communities with the biggest accumulations of projects Since the financial impulse was small and visible only locally, the effects should be researched project by project and then aggregated. Each project effect has to be assessed on two scales: the quality of life vs. integration objective. Criteria assessment differs with the thematic orientation of the projects, but a unified scale has been developed by each thematic expert
Stage 4: Tracing the real change – outputs and effects	For outputs: What were the funds and projects' spatial distribution? How many partnerships were established? For effects: What were the effects of the groups of projects from each thematic group in terms of quality of life and integration (thematic dimension)? What were the effects on the functioning of partnerships (relational dimension)? What was the impact on local communities with the highest project accumulation (territorial dimension)?	Maps of project locations (communities and projects topics), geographical distribution of funds (per capita in communities) For the thematic dimension: · Local visits of all project sites · Survey of project beneficiaries · Review of the local press For relational dimension: · Survey of all partnerships (leaders and their partners) · Network analysis of selected partnerships For the territorial dimension: · Social survey of twincommunities (pairof communities close to the border with/without INTERREG projects, and a pair far from the border with/without projects) · Interviews with local authorities	The main conclusion is that the programme has mainly local effects and it was rather close-to-border than cross border. However, in a given context (the border as a strong barrier) every joint project has to be assessed undoubtedly as a success. More detailed findings of the analysis can be summarised in five main points: 1. There was a disproportion between the activity of Polish, Ukrainian and Belarusian beneficiaries. This is visible in terms of the projects' spatial location as well as the composition of partnerships. 2. The programme in practice focused to a large extenton improving the quality of life. 3. Quality of life improvements were mainly local and "close to border". 4. The main integration effect was institutional. It was brought about mainly by projects from Priority 2. The integration of wider groups i.e. local societies was observed to a much lesser extent. 5. More balanced effects in terms of quality of life and integration was provided by the tourism development projects and projects that aimed at border security. These projects were often bridging hard infrastructural initiatives with soft micro-projects.

Stage of the research	What was done? – Questions addressed	How did we do it? – Methods	What did we find out? – Findings
Stage 5: Explaining the obtained effects	A set of "Why" questions arising from the 5 main findings: 1. Why was the spatial pattern of project distribution unbalanced? 2. Why did the majority of the projects obtain objectives for quality of life and to a lesser extent integration? 3. Why did the majority of the projects have local and close-to-border effects? 4. Why was the main integration effect institutional in nature and to a lesser extent societal? 5. Why were tourism development projects and border security projects the most balanced in terms of quality of life and integration effects?	Each of the question was tested with 3 explanatory factors: · The role of the context · The role of delivery mechanism (implementation system) · The characteristics of potential beneficiaries Methods and sources included: · Surveying unsuccessful programme applicants and communities – potential beneficiaries that never applied · Structured interviews with Programme management teams · Discussion with key Programme Stakeholders · Feedback from the qualitative and quantitative data collected during the cause of the research · Experts' brainstorming and assessment	1. The disproportion between the 3 partners is clearly explained by the context –the initial EU decision on unbalanced allocation between Poland, Ukraine, Belarus (37 mln vs. 8 mln EUR) as well as different implementation regimes applied by EU for INTERREG and TACIS (e.g. for most of the time TACIS CBC could not finance the infrastructural projects). An additional, though smaller factor was the greater experience of Polish institutions in applying for EU funds. 2. The obtained focus on quality of life is again explained mainly by the context in which the programme operated – the strong peripheriality and huge developmental needs of local actors from the area concerned. An additional factor was the delivery mechanism. It was an absence of strong demarcation line for Polish partners between PUB and other available EU funds. This resulted in attempts to "repack" the same ideas of local projects for the purposes of different programmes. 3. The third issue (close-to-border effects) can be explained again by a mixture of delivery mechanism and contextual factor. The funds and implemented investments were too small and too spread out to make a real transnational impact over a highly secured border. 4. The explaining factors of institutional integration effect come from programme delivery and beneficiary characteristics. Firstly, the selection criteria gave a strong, positive impulse for applicants to establish partnerships. Secondly, beneficiaries have limited trust to the partners. Thus, they preferred to work on smaller, less risky initiatives (Priority 2 instead of 1 – infrastructural). Thirdly, most of the projects in Measure 2.1. were an expert-type, thus, they integrated institutions and narrow target groups. Fourthly, micro-projects were in many cases created as top-down (initiated by local authorities) not bottom-up initiatives (local civil society organizations) due to the weakness of the latter. Thus, their follow-up or even identification was limited in the local societies. 5. The final finding is explained by the nature of the projects and delivery mechanism. Small infrastructural projects in the field of tourism or culture allowed to combine them with cultural micro-projects (which brought multiplier effects). Security projects hadan initial condition of a signed cooperation agreement between particular public services (police, fire brigades, etc.). This guaranteed the use of a projects' effects on both sides of the border.

Source: Final report from ex post Evaluation of Poland-Belarus-Ukraine INTERREG/TACIS CBC Neighbourhood Programme 2004-06

Thanks to this the findings of the study are based on a stronger, more coherent and transparent picture. At the same time the presentation of findings is not a mere description, but an elegant, logical story that explains the chain of actions, effects and their determining factors.

This is followed by a third argument. The logical chain promoted by TDE compels focusing on effects and treats issues of implementation system (procedures, processes, institutional competences) as only one explanatory factor. This is especially important for Cohesion Policy evaluation, which in practice tends to dwell on easier, procedural issues at the cost of more demanding analysis of effects.

Fourth, TDE facilitates connecting particular analysis with a wider body of knowledge and theories of socio-economic development. As a comparative analysis of European Social Fund Evaluations shows, studies clearly lack reference to this type of "bigger picture" and scientific approach. They are mainly technical and administrative exercises despite the scale of interventions and abundance of resources (see: previous chapter by Olejniczak in this book). Referring explicitly to theories and tracing the theoretical concepts and experiences that underlie practice would improve the quality of future evaluation as well as the depth of the public debate on the rationality of policies and programmes. Thus, TDE would clearly support the recent move towards evidence-based policies.

Fifth, it can also be stated that TDE helps to explore all three fields of evaluation of effects: verification of initial objectives, analysis of real effects (both planned and side-effects) and finally a tentative explanation of the causes of the observed effects.

Finally, TDE makes possible to some extenta discussion on causal relations, which is crucial for the studies on effects, even in the case of complex multi-objective programmes and topics, when the use of control groups of quasi-experiments is not feasible. By showing that a programme proceeded in line with a certain planned sequence of actions and that the reaction of beneficiaries was as expected, the evaluator can connect (in a tentative manner) actions with products and effects (Weiss, 2004).

Of course, Theory-Driven Evaluation has its limitations. Three of these are especially challenging for cohesion policy programmes.

Firstly, the mechanical, rigid use of one model can result in "tunnel vision" – that is, narrowing down a perspective of the research and excluding from the analysis any phenomenon that was not embedded in the initial model. However, Bamberger (2006, p.173) reassures that TDE usually manages to grasp the side-effects in the initial stages of the program theory reconstruction, while during the research it is just a matter of the evaluators staying alert and open-minded, especially during study visits.

Secondly, in the case of the complex interventions with multiple objectives there is always a trade-off between taking into account most of the factors and keeping the model clear and concise. What is more, for this type of programme the standard research plans and tools for causal relations have very limited application. A possible solution to this could be focusing the research onto separate, smaller parts of the programme (Actions, Priorities, sub-priorities, etc.). On the one hand this solution substantially increases the costs and time of research because every selected segment becomes a separate "evaluation", however, it does assist in the exploration causal considerations.

A final challenge has been pointed out by Patton (2008, p.358), which is that there is a risk of losing a balance between theoretical considerations and practice. TDE could change evaluation into scientific research, detach it from information needs of the concrete stakeholders and limit its utilitarian character. What is more, Scriven points out that too much theory (especially introduction of wider, scholarly theories) can polarise and discourage stakeholders. It can undermine their own programme theory that is not grounded in sublime and sophisticated theoretical constructs but in real life practice and common sense. It seems that only the evaluators can overcome this last challenge. In order not to alienate stakeholders, the evaluator has to establish good communication with them (in terms of regularity of contact as well as clarity of message) and most of all remember the simple rule that evaluation is a utilisation-focused exercise.

Despite these three challenges it can be stated that Theory-Driven Evaluation seems to be a promising and potentially useful approach for the evaluation of effects of cohesion policy programmes. Of course its potential needs further exploration in practical research. The closure of the 2000–2006 interventions provideda good opportunity to evaluate the utility of this approach in different regional, national and sector-specific situations.

Bibliography

Argyris, C. (1977) Double-loop learning in organizations. *Harvard Business Review*, 55(5), pp.115–125.

Astbury, B. &Leeuw, F.L. (2010) Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation*, 31(3), 363–381.

Azzam, T. (2009) "Using Technology to Enhance Applied Research and Evaluation", *Claremont Professional Development Workshop Series*, Claremont Graduate University, Claremont, August 26, 2009.

Bachtler, J. & Wren, C. (2006) Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges. *Regional Studies*, 40(2), pp.143–153.

Bamberger, M., Rugh, J. & Mabry, L. (2006) *Real World evaluation: working under budget, time, data, and political constraints*. Thousand Oaks, Calif.; London: Sage.

Bardach, E. (2006) "Policy Dynamics"; in: Moran, M., Rein, M. &Goodin, R.E. (eds.) *The Oxford Handbook of Public Policy*, pp.336–366. Oxford, New York: Oxford University Press.

Bickman, L. (1987) Using program theory in evaluation. *New Directions for Program Evaluation*, 33, pp.5–18.

Chen, H.T. (1994) *Theory-Driven Evaluations.* Newbury Park CA: Sage Publications, Inc.

Chen, H.T. (2004) *Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness*. Thousand Oaks: Sage Publications, Inc.

Chen, H.T. (2005) "Theory-Driven Evaluation"; in: Mathison, S. (ed.) *Encyclopedia of evaluation*, p.415–419. Thousand Oaks, Calif.; London: SAGE Publications.

Chen, H.-T. & Rossi, P.H. (1983) Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7, pp.283–302.

Christie, C.A. (2008) "What have we learned about RCTs, Gold Standards, and credible evidence?", *European Evaluation Society Conference*, Lisbon, 2.10.2008.

Council of the European Union (2006) Council Regulation (EC) No 1083/2006 of July 11, 2006 laying down the general provisions for the European Regional Development Fund, the European Social Fund and the Cohesion Fund. 1083/2006,

Dahler-Larsen, P. (2005) "Evaluation in Public Management"; w: Ferlie, E., Lynn, L.E. & Pollitt, C. (red.) *The Oxford handbook of public management*, p.615–639. Oxford: Oxford University Press.

Daponte, B.O. (2008) *Evaluation Essentials: Methods For Conducting Sound Research (Research Methods for the Social Sciences)*. San Francisco: Jossey-Bass.

Donaldson, S.I. (2007) *Program Theory-Driven Evaluation Science: Strategies and Applications*. New York: Lawrence Erlbaum.

European Commission (2000) *Methodological Working Paper 3 – Indicators for Monitoring and Evaluation: An indicative methodology*, The New Programming Period 2000–2006 Methodological Working Papers; Brussels: European Commission – DG-Regio.

Ferry, M. & Olejniczak, K. (2008) *The use of evaluation in the management of EU programmes in Poland*, Warsaw: Ernst & Young – Program "SprawnePaństwo": http://www.ey.com/PL/en/Industries/Government---Public-Sector/Report_The-use-of-evaluation

Frechtling, J.A. (2007) *Logic modeling methods in program evaluation*. San Francisco, Calif.: Jossey-Bass.

Furubo, J.-E., Rist, R.C. &Sandahl, R. (2002) *International atlas of evaluation*. New Brunswick, U.S.A.: Transaction Publishers.

Górniak J., Keler K., (2008) Ewaluacja i wskaźniki w cyklu planowania realizacji polityk publicznych. *Zarządzanie Publiczne*, 1(3), pp.109–123.

Hart, C. (1998) *Doing a literature review: releasing the social science research imagination.* London: SAGE.

Kellogg Foundation (2004) *The logic model development guide*, Battle Creek: W.K. Kellogg Foundation.

Knowlton, L.W. & Phillips, C.C. (2008) *The Logic Model Guidebook: Better Strategies for Great Results*. Los Angeles: Sage Publications, Inc.

Leeuw, F.L. (2003) Reconstructing Program Theories: Methods Avaliable and Problems to be Solved. *American Journal of Evaluation*, 24(1), pp.5–20.

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers by Kurt Lewin*. London: Tavistock.

Lipsey, M.W., Crosse, S., Dunkle, J., Pollard, J. &Stobart, G. (1985) Evaluation: The state of the art and the sorry state of the science. *New Directions for Program Evaluation*, 27, pp.7–28.

Mathison, S. (2005) (ed.) *Encyclopedia of evaluation* Sage, Thousand Oaks, Calif.; London.

Morgan, G. (2006) *Images of organization*. Thousand Oaks; London: SAGE.

Olejniczak, K. (2009) *Praktyka ewaluacji efektów programów rozwoju regionalnego – studium porównawcze*, Raporty i analizy EUROREG 2/2009; Warszawa: EUROREG-Uniwersytet Warszawski: www.euroreg.uw.edu.pl

Patton, M.Q. (2008) *Utilization-focused evaluation. 4th edition.* Los Angeles, London: Sage Publications.

Pawson, R. (2002) Evidence-based Policy: In Search of a Method. *Evaluation*, 8(2), pp.157–181.

Pawson, R. (2009) *"Introduction to Realist Evaluation and Realist Synthesis", AkademiaEwaluacjiProgramówRozwoju Społeczno-Gospodarczego,* EUROREG – UniwersytetWarszawski, Warszawa, 7 February 2009: www.ewaluacja.edu.pl

Petrosino, A., Rogers, P., Huebner, T. & Hacsi, T. (2000) Program Theory in Evaluation: Challenges and Opportunities. *New Directions for Evaluation*, 87(Fall),

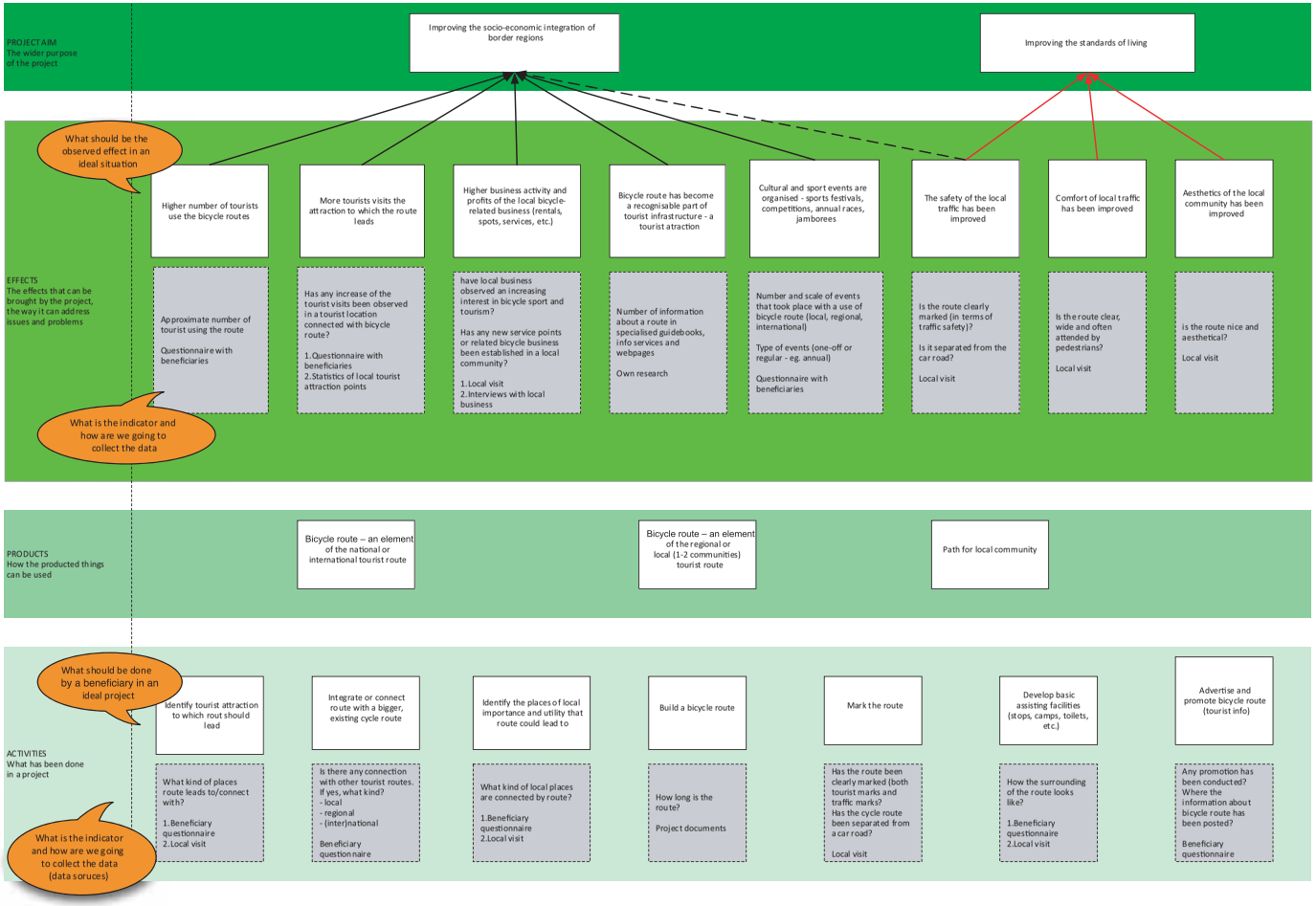
Rogers, P. (2005) “Logic model”; in: Mathison, S. (eds.) pp.232–235. Thousand Oaks, Calif.; London: SAGE Publications.

Rossi, P.H., Freeman, H.E. &Lipsey, M.W. (1999) *Evaluation: a systematic approach. 6th edition*. Thousand Oaks, Calif.; London: Sage Publications.

Weiss, C.H. (1997) How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), pp.501–524.

Weiss, C.H. (2004) On Theory-Based Evaluation: Winning Friends and Influencing People. *Evaluation Exchange*, IX(4), pp.2–3.

Annex 1: An example of a logic model –the tourism sector – bicycle routes



OPPORTUNITIES AND LIMITATIONS OF COUNTERFACTUAL IMPACT EVALUATION OF STRUCTURAL FUNDS

Alberto Martini[1)]

We all desire to have an impact. We also tend to attribute the results we obtain to our actions. But if we want to improve what we do, we must learn which results can be really attributed to us and must distinguish them from what would have happened anyway.

1. A tale of two questions

This simple wisdom from above also applies to public policies and provides the logic of impact evaluation. We can start by assuming that policymakers (either elected or career officials) *desire to have* an impact on the problems facing their constituency or client base. They also tend to presume that all desirable changes are a consequence of their own policies. However, if policymakers must decide whether to expand, contract or maintain a program, or simply want to improve it, they need more than accountability information, they need to *learn what does and does not work and why.* Thus, evaluating the impact of (cohesion) policy involves a variety of cognitive tasks, with varying degrees of complexity. These are schematically illustrated in Table 1.

Table 1. Different cognitive tasks of evaluation

Main purpose of evaluation	Accountability for results	Learning for better decisions	
Main evaluation question	What has been achieved? How much progress has been made?	What works in solving the problem?	Why does it work? Why does it not?
Main analytical challenge	Finding evidence on the contribution of the policy	Establishing a causal link between policy and outcomes	Identifying causal mechanisms
Main analytical tool	Indicators and targets	Counterfactual-based methods	Theory-based methods

When the purpose of evaluation is *accounting for results*, analyzing impacts means establishing the "contribution" of the policy to the progress made toward the objectives. An excellent example of such a quest for "contribution" can be found in the

[1] The author is Director of Progetto Valutazione and Associate Professor of Statistics and Program Evaluation at the Universitŕ del Piemonte Orientale, Italy. He would like to thank Fabrizio Barca, Veronica Gaffey, Daniel Mouque, Barbara Romano and Kai Stryczynski for useful discussions on the role of impact evaluation in cohesion policy.

following quotation. *"A priority of the new approach to evaluation in the 2007–2013 period is to assess the contribution of cohesion policy to the achievement of the Lisbon goals and to make that contribution more visible."* (DG Regio, 2007, page 11).

When the purpose is not to be accountable for past results, but to support future decisions, the evaluation must go beyond "contribution" to determine "attribution" – the causal link between the outcomes and the intervention – and "mechanisms" – the channels through which the intervention produces its outcomes. In this paper we are mainly concerned with the first of these challenges (contribution), but we start by clarifying the difference with respect to the second.

When we want to evaluate the impact of (cohesion) policies with the goal of improving them, we must deal with two distinct questions: the *quantification* of effects – their sign, size and with their substantive as well as statistical significance, and their *explanation* – which mechanisms produce the effects. The most effective way to distinguish these two cognitive tasks is to say that one tries to understand *what works*, the other tries to understand *why it works.*

The *what works* question should take precedence – but not prominence – with respect to the *why* question: if we do not know the direction and size of an effect, we do not know what to explain. Explaining the *presence* of an effect involves different mechanisms than explaining its *absence*. That said, it must also be recognized that estimates of the impact's direction and size, without knowledge of the underlying mechanisms, are more difficult to interpret and harder to use for policy-making. Without sound explanation, even a solidly demonstrated lack of impact can be largely inconclusive in terms of policy implications. Lack of impact can be grounds for either reducing or eliminating the intervention ("it is simply ineffective"), expanding it ("it is ineffective because it is underfunded") or simply taking a harder look at its implementation.

An example from cohesion policy

Let us take a program providing *subsidies to increase R&D expenditures* among small and medium enterprises (SMEs).[2)] From a simple (simplistic) accountability perspective, the take-up rate is a measure of impact. A very low or nil take-up rate would certainly imply lack of impact. How about a high take-up rate: does it also imply a positive impact? The answer is no. If we want to understand whether these subsidies work, we need to establish whether they increase the average R&D expenditure *with respect to what would happen in the absence of subsidies.*

[2] Another example is offered by training programmes, evaluated for their effects on the post-program earnings and employment retention of participants. While it is important to know the average effect and how it varies with the characteristics of participants, it is also important to understand how programme effects are generated – by increasing the productivity of those trained (human capital explanation) or by revealing those who are more productive among the wider pool of potential trainees (signaling explanation). If training fails to produce impacts, is this due to the wrong design of the training programs, to the inadequacy of the participants, or simply to adverse economic conditions?

Let us consider the case in which – with one of the methods we will present in Section 3 – we find that the average impact is zero, motivating the search for an explanation of this disappointing result. The explanation turns out to be that the subsidies are too small to induce any firm to change its investment strategy: the take-up rate might well be 100 percent, but the impact is zero. The implication is that the program should be terminated because it only generates a transfer to enterprises, without changing their behavior.

An alternative explanation for the zero impact finding could be that firms, after receiving the subsidy, adjust their portfolio of other R&D projects – closing or slowing down non-subsidized projects. The policy implications of this mechanism are different than before: the composition of the R&D portfolio is altered, perhaps promoting better quality projects. The program should not be terminated, although it has zero impact on the main outcome of interest.

Let us change scenarios. The impact analysis now reveals that subsidies have a positive effect –they do increase R&D expenditures with respect to the counterfactual situation. Different mechanisms may generate this result. It could be that the subsidies induce firms to undertake riskier projects, which they would deem too risky had they to bear the full cost. Another explanation of the positive effect could be that the subsidy eases existing credit constraints, thus allowing firms to undertake worthy projects they could not afford. One mechanism reduces the average quality of the R&D portfolio, the other does not. The policy implications differ in both cases.

The questions we raise (what are the effects? what explains them?) are indeed formidable questions, and the available analytical methods provide at best tentative and incomplete answers to most of them. Thus it is important to clarify which methods can answer which questions and under which circumstances. We find it useful to distinguish between two sets of methods:

- methods *primarily* devoted to establishing *whether* a given intervention produces the desired effects on some pre-established dimension of interest. The overarching goal is to answer the question "*does it make a difference?*" by identifying and estimating causal effects through *counterfactual methods*. We speak of *Counterfactual Impact Evaluation (CIE)*;
- methods *primarily* devoted to understanding why an intervention produces intended and unintended effects. To answer the "*why it works?*" question one needs to identify the theory of change behind the program and to compare theory with actual implementation. Here we are speaking of *Theory-Based Impact Evaluation (TBIE)*.

We want to stress the term "primarily". Identifying and estimating causal effects requires *some* theory, while comparing theory and implementation requires *some* quantification. However, quantification of effects and explanation remain two distinct tasks. It would be counterproductive at this stage of the development and

utilization of these methods to force a synthesis between the two sets of questions and related methods. Rhetorical claims about the alleged superiority of one set of methods over the other should be left to the bygone era of the fruitless "paradigm wars". What the two camps mostly have in common is *how little* they truly understand the effects of public policies.

A clear cut distinction between these two sets of methods helps in preventing antagonism, which tends to thrive when proponents of alternative evaluation methods vie for the attention of the same policymakers and compete for the same resources. Such antagonism should be discouraged by rewarding those who are intellectually honest and admit the drawbacks, limitations and pitfalls of the analytical tools each side is able to deploy in answering questions about effectiveness of policy.

Counterfactual Impact Evaluation

The question CIE poses – how much difference does a treatment make – produces answers that are typically *numbers*, most often in terms of *differences*, to which we may able to give a plausible *causal interpretation* based on empirical evidence and assumptions. Is the difference *observed* in the outcome after the implementation of the intervention *caused* by the intervention itself or by something else? Answering this question in a credible way is very challenging because it involves one of the most difficult intellectual tasks – establishing causality in a rigorous way.

The CIE results are useful for many policy decisions because: (i) it provides easily interpretable information; (ii) it is an essential ingredient for cost-benefit and cost-effectiveness calculations; (iii) it can be broken down in separate estimates for subgroups, provided that the subgroups are defined in advance. To sum up, "how much difference does a treatment make" is an important, relevant, and methodologically sound evaluation question. Yet it remains extremely challenging to answer, as this paper openly argues. And it is certainly *not* the only question.

In this chapter, we deal with the methods devoted to quantify whether a given intervention produces the desired effects on some pre-established dimension of interest. The essence of the problem has to do with the "attribution" of the observed change to the intervention. Is the change due to the policy or would the change have occurred anyway? Answering this question is not as straightforward as it might seem. The challenge for quantifying effect is finding a credible approximation to what would have occurred in *the absence* of the intervention, and to compare it with what actually happened. The difference is the estimated effect, or impact, of the intervention on the particular outcome of interest (be it per capita GDP, R&D expenditure, housing values or employment levels).

The importance of TBIE stems from the fact that a great deal of other information, besides quantifiable causal effect, is useful to policymakers in making decisions on program design and targets. The question of *why* a set of interventions produces intended as well as unintended effects is as relevant and important as the "made a difference" question, and equally if not more challenging.

This approach does not produce a number, *it produces a narrative*. Therefore, it cannot be used for cost-benefit calculations, it is not communicated as quickly and schematically, and it is not backed by a comparable set of statistical tools. Thus it appears to some observers as less scientific, less "objective". But it can provide a precious and rare commodity, insight into why things do or do not work. Above all, it is based on the idea that the essential ingredient is not a counterfactual ("how things would have been without") rather a *theory of change* ("how things should logically work to produce the desired change"). The centrality of the theory of change justifies calling this approach *theory-based* impact evaluation.

While the two approaches should be kept separate methodologically, policymakers should use the results of both sets of methods as they see fit. While joint utilization is up to the user of the information, it does not imply joint production.

2. Common (but avoidable) fallacies about evaluating impacts

The concept of impact is surrounded fascination, fallacies and misconceptions. We focus on three of these: (i) the inept distinction between effects and impacts; (ii) the illusory decomposition of gross impacts; (iii) the insatiable appetite for impact indicators.

2.1. Effects and impacts: a distinction without a difference

The first fallacy we consider is the attribution of different (and often interchangeable) meanings to "impact" and "effect". In the most popular definition, "effects" are seen as immediate results for the direct beneficiaries and "impacts" as effects in the long-run and/or applicable to a wider, usually undefined population.

We follow the literature on causal inference in treating impacts and effects as perfect synonyms. There is no truly meaningful difference between the two terms within this approach. To underline this equivalence, we speak of "causal" effect/impact – the difference between the outcome occurred after an intervention has taken place and the outcome that would have occurred in the absence of the intervention. Whether it is the short run or long run, concerning direct or indirect beneficiaries, or has important consequences for the way the evaluation is performed, the distinction between effects and impacts is not much use by itself.

2.2. From “gross” to “net” impacts through “deadweight”

It is faithfully repeated document after document but never translated into credible estimates: we refer to the decomposition of “gross impact” (or, equivalently, “gross effect”) into “deadweight effect”, “substitution effect”, “displacement effect” and “net effect”. Behind its captivating appeal, this decomposition hides a fundamental lack of analytic rigour and clarity of purpose.

The two components related to non-beneficiaries, substitution and displacement, can be invoked on theoretical grounds, but are extremely difficult to quantify for the simple reason that the individuals who are “victim” of substitution and displacement are not known. We can presume their existence, but cannot measure their status (typically, their employment status), l*et al*one determine how the policy has affected their status. This “small” detail is systematically ignored by the proponents of the “gross” decomposition.

But even the remaining “deadweight” component cannot simply be subtracted from the observed change. Before it is subtracted, it must be quantified. It turns out that the deadweight is a special case of the counterfactual, thus it is equally difficult to identify and estimate. The superficial treatment of the concept of “deadweight” employed in evaluation literature is unfortunate: the idea that deadweight can simply be “measured” has reinforced the notion that “net effects” can also be simply derived subtracting observed quantities. *This is clearly not true*. Effects are not observable; however, they can be inferred given the proper data and the assumptions we make on the data. If counterfactuals are not observable, neither are deadweights.

One might ask why the concept of “deadweight” is employed at all, instead of using the broader and more universal notion of counterfactual. The only reason we can see is that “deadweight” is utilized for a family of interventions that consist in *paying out* a subsidy in order to induce certain individuals to perform a desired action. The common problem with such policies is that some eligible units might have adopted the action anyway, even without the subsidy. Thus, the money spent on them is ineffective: with hindsight, it could have been saved. Deadweight seems the apt name for this type of waste. However, labeling it nicely does not in any way help in obtaining credible estimates. What would the recipients have done in the absence of the subsidy? How many firms would have carried out their R&D projects without the subsidy? That is, what would have been the counterfactual situation? Changing names does not solve the problem.

Despite its weaknesses and lack of empirical applications, the decomposition of “gross impact” remains the only conceptualization of the impact evaluation problem for most European evaluation literature, particularly that dealing with Structural Funds. This is unfortunate.

Rather than attempting to decompose change over time, almost all empirical methods used by CIE define effects by comparing *beneficiaries and non-beneficiaries*. The outcomes observed for non-beneficiaries – with the proper adjustment – are used to approximate the counterfactual. This forces the evaluator to deal with the selection bias problem (either with statistical adjustment or with some form of randomization).

This implies that there is a mismatch between the way the problem is conceptualized (mainly in EU documents) and the way it is solved empirically (by the program evaluation community). This mismatch has several negative consequences. On one hand, the proponents of decomposition of change tend to invent all sorts of "effects": not being compelled by the need to find an empirical counterpart to their inventions, they find little constraints for their imagination. On the other hand, the proponents of decomposition have little appreciation for the way counterfactual analysis is actually done, by comparing beneficiaries and non-beneficiaries. This gap must be closed, realigning conceptualization and empirical solutions.

2.3. Impact indicators: an oxymoron?

Another example of misalignment between theory and methods refers to the much invoked use (and abuse) of impact indicators. What is a purely descriptive tool – an indicator, the quantification of phenomenon – with the addition of the word "impact" becomes magically a tool to establish causality. On this topic, we fully endorse the position of the Barca Report: *"Outcome indicators and impact evaluation respond to two radically different tasks: the first represents a tool to focus policymaker and public attention on objectives and to monitor what happened to objective-indicators while interventions are being implemented; the latter represents a tool to learn about whether specific interventions have had an effect on a given dimension. Unlike what has happened so far, these two functions must not be confused."* (Barca, 2009).

Outcome indicators have an important role to play in the management of Structural Funds in tracking progress toward established objectives. This is a worthy goal, and it requires a good system of performance monitoring. In tracking progress, no causality is involved nor needs to be established. It is enough to gather data on a baseline and set targets along the way. To be sure, the specter of causality never stops haunting the evaluator: while doing performance analysis, he or she will invariably be asked whether the "results" are "due" to the policy. However, the answer will be a polite "in this context we do not really do that". One needs a different type of analysis to establish causality. The very idea of an impact indicator is an oxymoron: an indicator describes something, it cannot imply a causal relationship.

"Impact" indicators often represent the social/economic problems the program intends to affect. These indicators can be compared across programmes, and particularly across time for the single programme, to show whether the problem worsens

or improves. There is a need for a baseline, explicit targets and periodic monitoring. But there is no need for causal inference.

3. The basics of counterfactual impact evaluation

Three basic approaches can be used to approximate the counterfactual and then to quantify program impacts: (i) using as counterfactual the outcome observed for beneficiaries *before* they are exposed to the intervention (*before-after comparison*); (ii) asking beneficiaries what they *perceive* is the counterfactual or simply the effect of the intervention (*beneficiary surveys*); (iii) using as counterfactual the outcome observed for non-beneficiaries that are similar – or made similar – to the beneficiaries (*comparison group analysis*). While the third approach, with its many variants, is the most widely used, every approach has strengths and weaknesses, and caution must always be used in giving a *causal interpretation* to any result.

Nevertheless, a plausible causal interpretation is important for decision makers who are considering whether the program should be replicated/maintained/brought up to scale or terminated. To understand "what works" one should rely on robust causal estimates, not simple association between outcomes and treatment. For example, does the difference in R&D expenditures between subsidized and unsubsidized firms reveal the effect of the subsidy received, or does it simply reveal the way firms self-select when deciding whether to apply for the subsidy? Impact evaluation is about *interpreting differences*. The challenge facing the evaluator is to avoid giving a causal interpretation to differences that are due to factors other than the intervention. *This is the essence of counterfactual impact evaluation.*

3.1. Before-after comparisons among beneficiaries

Let us take the first of the three basic approaches, the before-after comparison. *When the same units* are observed both before and after they are exposed to an intervention, the fundamental evaluation problem is that the observed change could be due to the intervention as well as to other changes occurring during the same period. Without other information, we cannot separate the two components.

The problem can be illustrated by the following formula:

(1) $$\Delta_{B\text{-}A} = E + D_{B\text{-}A}$$

where

$D_{B\text{-}A}$ is the average difference in the outcome before and after the intervention among beneficiaries

E is the true effect of the intervention on the outcome, averaged across beneficiaries

D_{B-A} is the "natural dynamics" of the outcome over time among beneficiaries[3)]

The social or economic phenomena public policies are trying to affect do evolve naturally over time in ways that are independent of the policies. To stay with our usual example, we would compare average R&D expenditures the year before the firms applied for the subsidy and the year after the subsidized project was completed (to avoid counting the subsidy a part of the expenditure). Did receiving the subsidy have a long-lasting effect on the firms propensity to invest in R&D? Or are the firms with increasing expenditure in R&D those which apply for the subsidy?

The credibility of the effects estimated from before-after comparisons is generally very low. Other than assuming away the problem – assuming temporal stability, that is, that there is no maturation or natural evolution – there is often little that can be done. Before-after differences do not reveal the true effect of the intervention, unless we assume complete stability – that is, the absence of "natural dynamics".

3.2. Asking beneficiaries about their perception of the impact

An approach quite popular in the evaluation of Structural Funds consists simply of asking a representative sample of beneficiaries how they perceive the impact of the intervention upon themselves. The answers are then aggregated, yielding an estimate of the impact. In our example, this strategy would entail asking a sample of recipients their expenditure in R&D the year preceding and following the subsidy and whether the subsidy was instrumental in increasing it. Alternatively, the question could be formulated with a counterfactual flavor: "if your firm had not received the subsidy, would your expenditure in R&D have changed anyway?".

For each firm we would then have the following information:

Δ_{B-A} is the average difference in the outcome before and after the intervention among beneficiaries

Q_P is the answer to the question "was the subsidy a decisive factor in increasing expenditures?"

The average effect of the subsidy among beneficiaries is estimated by:

(2) $$E_R = \Delta_{B-A} * (\Sigma Q_P / \text{beneficiaries})$$

where SQP is the number of those who respond positively to the question "Was the subsidy a decisive factor …?". If weighted by the take-up rate, we obtain the average effect of the subsidy among those eligible.

---

[3] Note that (1) bears some resemblance with the "gross impact = net impact + deadweight" decomposition so popular in EU documents. However, (1) is more general than the deadweight decomposition because it applies to all types of outcome variables, while the deadweight idea applies only to outcomes that consist of well-defined activities, such as investment projects, hiring of disadvantaged workers, or adoptions of energy saving practices. The concept of deadweight does not apply when the outcome is a behavior or a performance, such as expenditure in R&D. The idea of deadweight is only applicable to subsidized activities: unlike the concept of counterfactual, that of deadweight is meaningful only when referred to a programme.

(3) $$E_E = \Delta_{B\text{-}A} * (\Sigma Q_P \text{ /beneficiaries}) * (\text{beneficiaries/eligible})$$

The weaknesses and advantages of this approach are immediately apparent. Among the latter, the fact of requiring only one interview, limited to the beneficiaries, for whom program administrators typically have tracking information. Among the former, the fact of being based on perceptions, of imposing a substantial cognitive burden on the respondents, and the danger of eliciting biased responses, when the beneficiaries perceive it as being in their interest to provide favorable responses.

3.3. Comparing beneficiaries and non-beneficiaries: the most promising (yet challenging) strategy

The most common strategy to estimate the causal effect of an intervention is to exploit the fact that some "units" (individuals, families, firms, places) have been exposed to the intervention and some have not, according to some *(self)selection process.*

For example, eligible enterprises may or may not apply for state aid to finance R&D projects; unemployed workers may or may not participate in a retraining program after a plant closure; urban neighborhoods may or may not receive funding for urban renovation projects. Although the existence of universal policies cannot be ignored, they are relatively rare in the case of cohesion policies. In most instances, it is possible to find units that are not exposed to the policy. For simplicity, we consider only the case of a simple binary treatment, where the units either receive the treatment implied by the policy or they do not.

The outcomes observed among beneficiaries can be compared to those observed among non-beneficiaries; however, this difference does not by itself reveal the true effect of the intervention on the outcome. It cannot necessarily be *interpreted* in a causal sense. The causal interpretation depends on the nature of the process that leads some units to be exposed to the intervention, while others are not. Ideally, the selection process must be independent of the outcome, which is guaranteed only when the selection is random.

If the selection is not random, the *observed* difference can always be thought as the sum of two components: the true effect of the policy and the difference in outcomes due to the selection process itself. Neither one can actually be observed, we can only make "educated guesses" about them. The following decomposition is fundamental to show the logic behind the impact evaluation methods most widely used in the literature.

(4) $$\Delta_{T\text{-}NT} = E + S_{T\text{-}NT}$$

where

$\Delta_{T\text{-}NT}$ is the average difference in the outcome observed between beneficiaries and non-beneficiaries of an intervention

E is the true effect of the intervention on the outcome averaged across eligible units

$S_{T\text{-}NT}$ is the selection-generated difference between beneficiaries and non-beneficiaries. Keeping with the existing terminology, we will refer to ST-NT as ***selection bias***.

For example, in the case of the subsidies given to firms to invest in R&D, the differences between the performance of subsidized and non-subsidized firms can be decomposed into the true causal effect (possibly zero) of the subsidy and the differences due to the selection process that sorts eligible firms into recipients and non-recipients. Subsidized and unsubsidized firms would most likely differ in terms of R&D expenditures even if the former had not received the support. $S_{T\text{-}NT}$ is not zero.

How does one eliminate selection bias? Eliminating selection bias represents the major challenge in conducting impact evaluations and it has received a lot of attention from statisticians, economists and sociologists. A range of methods and techniques are available to deal with it. Knowledge of the selection process is crucial in order to choose the best methods.

The ideal strategy to avoid selection bias: randomization

The ideal strategy to eliminate selection bias is to randomly select who becomes a beneficiary and who becomes a non-beneficiary. In this case we know selection bias is zero. Unfortunately, randomization is not always a feasible option. Randomization produces impact estimates that are internally valid, but are *difficult to generalize*. Such generalization is key to the usefulness of the result for policy-making. Experiments are often costly and require close monitoring to ensure that they are effectively administered. The potential for denying treatment can pose ethical questions that are politically sensitive. They may reduce the chances of an experiment being considered as a means of evaluating a programme and may also increase the chances that those responsible for delivery of the programme will be reluctant to cooperate.

Randomization requires carefully planning interventions, an early involvement of the evaluator, and a degree of stability of the environment in which the experiment is taking place. All these features are rarely present in the public sector of EU Member States. Randomization requires that the intervention is fairly simple, while cohesion policies are traditionally complex, because they insist on multifaceted/multilevel problems. While complexity is an overall obstacle to evaluation, and to knowledge more generally, in the case of randomization the clash between methods and circumstances is particularly striking.

Another problem with experiments is *non compliance*, which takes the double form of the *no-shows* (those assigned to treatment who drop-out before it is completed, sometimes even before it starts) and of the *cross-overs* (those assigned to control

who manage to receive treatment anyway). With both no-shows and cross-overs, non-experimental methods can be used to retrieve the desired parameters. However, this is a less than optimal position since experiments are designed specifically to avoid this sort of adjustment. To conclude, any credible strategy for evaluating the impact of cohesion policy must include in its arsenal a number of non-experimental methods (also referred to as "quasi-experimental").

The role of non-experimental methods in correcting for selection bias

We briefly illustrate the four main non-experimental strategies to reduce the presence of selection bias. We focus on approaches that have emerged in recent years as the most robust and credible. In brief, the four approaches are based on **matching** (finding comparable units), **double difference** (exploiting parallel trends), **discontinuity** (comparing marginal individuals around a threshold), and **instrumental variables** (using natural experiments). They have one feature in common: they all rely on some fairly plausible but non-testable assumptions. Let's examine each of them in more detail.[4)]

a. The **matching approach** is based on the idea of finding non-beneficiaries that "look like" beneficiaries in all (observable) characteristics. Once such "look alikes" are found, the effect is estimated by the average difference in the outcome variable(s) between beneficiaries and the "look alike" beneficiaries, under the assumption that matching has eliminated the differences due to selection. This assumption cannot be tested, but it becomes more credible as more variables that might have affected the selection process are used in the matching.

   In the example of R&D subsidies, the evaluator needs to know the characteristics that influence the decision on the part of the eligible firms to apply for the subsidy. Using these characteristics, the "propensity (probability) to apply" is constructed for every firm in the sample. Then each firm that received the subsidy is matched to firms that did not receive the subsidy but have a very similar value for the propensity (probability) to apply. The effect of the subsidy is then estimated by comparing R&D expenditures in the two groups.

b. The **difference-in-differences (double difference) approach** is based on the precondition that outcome data are available for beneficiaries and non-beneficiaries both *before and after* the intervention. The non-testable assumption is that the trend over time in the outcome observed for non-beneficiaries is a reasonable approximation of the *counterfactual trend* (what would have happened to beneficiaries *over time* had they not received the subsidy). If it

[4] The literature on this topic is quite extensive. See Angrist J.D., Pischke J,S. [2008] for an unusual and comprehensive treatment of the subject; Bertrand M., Duflo E. and Mullainathan S. [2004] on Difference-in-Differences; Caliendo M., Kopeinig S. [2008] on Propensity Score Matching; Cook T. [2008] and Trochim W. [1984] on Regression-discontinuity Design; Angrist J.D., Krueger A [2001] on Instrumental Variables; Einiö E. [2009] and Gadd H., Hansson G. and MÍnsson J. [2008] for recent applications of these methods to Structural Funds.

is reasonable to assume that non-beneficiaries move in a parallel fashion to beneficiaries, the difference-in-differences or double differencing impact estimate is obtained by subtracting the pre-intervention difference in outcomes from the post-intervention difference.

In the example of R&D subsidies, we do not need the characteristics that influence the decision of the eligible firms to apply for the subsidy, but we need two observations of R&D expenditures for the same firm. The average pre-post difference among subsidized firms, once the average pre-post difference for non-subsidized firms has been subtracted out, represents an estimate of the impact of the subsidy.

c. The **discontinuity approach** is based on the idea of *discontinuity in treatment around a threshold*, which mainly applies to those situations in which some units are made eligible for the intervention while others are made ineligible by some well defined rule, typically some administrative rule. The two groups are similar in other respects, but they are (sharply) divided according to their position with respect to a threshold, with those on one side of the threshold exposed to the policy and those on the other side not exposed. The effect of the treatment (around the threshold) is obtained by the difference in outcomes around the threshold. The identifying assumption (more credible than most) is that selection bias is zero around the threshold. The essential idea is that around the threshold we have a situation similar to randomization. It should be noted that the estimated effect is a *local* effect: it is more credible (better internal validity) but less generalizable (worse external validity).

   Going back to the usual example, the possibility of applying the discontinuity design depends on whether the selection for the R&D subsidy consisted in ranking the applicants according to a continuous score. By comparing the marginal firms, those just included and those just excluded, one obtains a rather credible, albeit local, estimate of the effect of the subsidy.

d. The **instrumental variables approach**, based on the idea of *natural experiment*: those situations in which the receipt of treatment is partially determined by an extraneous factor. This strategy is notably difficult to explain intuitively without algebra, as well as requiring two identifying assumptions. The first is that the extraneous factor has some influence on the participation in the policy. The second assumption is that the extraneous factor has no direct effect on the outcome other than through its effect on participation. This assumption is not testable and its plausibility depends on the nature of the selection process. Thus, the extraneous factor induces two effects: one on the outcome and one on participation. Neither effect is of much interest from a policy perspective, but we are interested in the effect of participation. It is possible to show that such an effect can be estimated by the *ratio of the two effects* of the extraneous factor.

In the case of the R&D subsidy, let us assume that only a fraction of eligible firms received the official announcements of the availability of the subsidy, and this partial coverage was random due to administrative mishaps. All eligible firms can apply, regardless of whether they received the announcement or not. However, the announcement has the effect of increasing participation from, say, 10 percent (among those who did not receive the announcement) to 20 percent (among those who did receive it). A comparison of all the participants with all the non participants would be affected by selection bias. The IV strategy (Instrumental variables approach) consists of computing the difference in the outcome between the recipients and non recipients of the announcement and dividing it by the difference in participation between the recipients and non recipients.

4. Is the counterfactual perspective applicable to cohesion policy?

This section explores the issue of applicability of CIE to cohesion policy. There are two main dimensions to this issue. One has to do with the relevance of the findings about impacts for the decision process: is impact evaluation able to produce findings that are relevant for future decisions? The other issue has to do with the *compatibility* between the technical requirements of the methods and the characteristics of the policies. To our knowledge, no systematic discussion of this issue has ever been conducted, so our attempt is entirely exploratory.

Relevance for decisions

We single out two features that affect the relevance of counterfactual impact evaluation: the prevailing type of motivation of the policy (the traditional redistributive motive vs. a behavioral modification motive) and the replicable nature of the realization of the policy.

We argue that to the extent to which cohesion policies are dominated by a redistributive motive, other things being equal, their impact is less worth evaluating for the simple reason that the impact of redistribution is much less uncertain than the impact of attempts at **modifying the behavior** of social and economic actors. The Barca report makes a strong case against a purely redistributive motive (Barca, 2009): "*The redistribution of resources among places is not a sufficient condition for pursuing either the efficiency or the equity objectives set out in the Treaty when calling for a reduction of disparities. It might be part of the means of such a pursuit – some places receiving more from interventions than they contribute through taxation – but the main purpose of cohesion policy is not redistribution but to trigger institutional change and to break inefficiencies and social exclusion traps through the provision of public goods and services*." To the extent that the purpose of cohesion

policy is to trigger behavioral change, we see a role for impact evaluation because behavioral change is typically hard to trigger, especially in the intended direction.

The second dimension has to do with the degree to which the policy generates initiatives and projects that are uniquely tied to a specific context or situation, as opposed to **being replicable** for other beneficiaries, in another time period or location. Large infrastructure projects tend to be highly idiosyncratic, and estimating their impacts – assuming it is possible – produces knowledge that has limited use. This can be said, to a much lesser extent, of urban regeneration projects. On the opposite side, we find interventions that are replicable across successive cohorts of beneficiaries, such as enterprise support or human capital interventions, or replicable across locations, as in the case of or for energy efficiency. Impact evaluation is useful for decision making as long as its findings are applicable to future decisions, not simply to past achievements.

Compatibility

As far as the compatibility between methods and policies is concerned, there are two key issues. The first has to do with the homogeneity of the treatment, the second with the availability of large numbers of comparable treated and non-treated units.

Counterfactual methods are well suited for relatively **homogeneous treatments**. Most applications focus on the case on binary treatment: it is either received (the same level for everybody) or it is not. Extensions to continuous treatment variables are still on the frontier of method development. Methods that allow for multiple treatments are still rare. Urban regeneration projects are a typical case in which bundles of treatments are provided. The problem is in the interpretation of the results: how would one interpret the finding that urban regeneration projects are effective, when each city receives a different mix of interventions? Even worse, how do we handle an impact of zero? Many components might work, but their contributions get lost in the shuffle.

Counterfactual methods, like all quantitative methods, work well with **large samples**. Some policies are characterized by homogenous treatments administered to large number of subjects. Moreover, these subjects come in successive cohorts, generated by the same underlying phenomenon. Training programmes represent the archetypal situation in which these methods have been developed. Interventions that target SMEs are similarly suited; although the successive cohorts feature is missing, there is a stock of firms of varying size and age. Transport projects and the production of renewable energy are further removed from the archetypal situation, while infrastructure projects, like high-speed railways or seaports, are at the opposite end of the spectrum.

Table 2 summarizes these arguments by assigning (subjective) scores to the four dimensions for six types of cohesion policies.

Table 2. Factors affecting applicability of CIE

Type of policy	Support for R&D projects	Investment support	Renewable energy	Urban renewal	Transport infrastructure	Human capital investment
Behavioral (vs. redistributive) motive	++	++	++	+	+	+
Replicable nature (vs. idiosyncratic)	++	++	+	-	-	++
Homogenous treatment (vs. composite)	++	+	+	- -	+	+
Large numbers of eligible units	+	++	+	-	-	++
APPLICABILITY OF CIE	HIGH	HIGH	MIXED	LOW	MIXED	HIGH

Legend: ++ positive contribution, + moderate contribution, – limited obstacle, – - serious obstacle

5. The case for a prospective approach

In principle, counterfactual evaluation can be conducted at any time after the policy is implemented. Historians now use counterfactual logic to assess the importance and the impact of events and decisions that "shaped history". What was the impact on the process of Italian unification when Napoleon sold the Republic of Venice to the Austrian Empire in 1797? What was the impact of the Gdansk strikes on the disintegration of the Soviet Empire? These are extreme cases of retrospective evaluation: retrospective not just because it looks at the past, but because it has no interest in shaping the present and the future.

The main purpose of the evaluation dictates its timing. If the purpose is to learn what does and does not work, some convincing arguments can be made for designing the evaluation parallelly with the design and implementation of the policy itself. The arguments have to do with data availability, increased utilization and the "disciplinary role" the evaluation can have on the policy itself.

The argument on data availability borders on the obvious: most non-experimental methods require some pre-intervention data and not just post-intervention outcome data. The latter are sufficient only if randomization has been used to allocate beneficiaries and non-beneficiaries, in which case pre-intervention data are not strictly needed, although they are useful to verify that the control group is genuinely indistinguishable from the treatment group. They are so in the case of matching (the two groups must be aligned on pre-intervention characteristics), double differencing (in this case pre-intervention *outcome* data are needed), and discontinuity design (the forcing variable must exist before the intervention is delivered). When pre-intervention data are needed (which is almost always), data collection should

be performed before the intervention takes place. In turn, this implies that the evaluation, or an important part of it, must be designed at the same time as the policy. We recommend that every major Structural Funds program be accompanied by an impact evaluation design that specifies which evaluation questions need to be answered, which data are needed, who is responsible for data collection and analysis, and how the evaluation results are to be used during the implementation of the policy.

Such a prospective approach to impact evaluation would produce two potential benefits. First, early impact findings, made possible by a timely evaluation, would feed into the reallocation of resources that typically takes place midway each programming period. Nowadays, such reallocation decisions are based essentially on monitoring data, which say nothing at all about the effectiveness of the interventions, but say a lot about the distance between what has been spent and what has been allocated. We break no great secret in claiming that no real evidence on impacts ever feeds into the Structural Funds allocation process.

The second benefit of a prospective approach to impact evaluation is the disciplinary effect it might have on the decision process. The Barca report is an adamant supporter of this idea, and we let the following quotation carry the argument:

"When it is prospective, i.e. it is designed together with the intervention, impact evaluation can have a strong disciplinary effect. First, it can help focus the attention of both policy-makers and beneficiaries on objectives. Secondly, it creates an incentive to assemble the information necessary to assess results. Thirdly, it brings to light the criteria by which beneficiaries are selected, which is a delicate problem in all development policies since there is a risk of policy being captured by pressure groups; prospective impact evaluation can, accordingly help to give transparency to the selection procedure. The place-based development approach offers a policy space in which prospective impact evaluation can take place and the relevant information extracted through cooperation between evaluators, policy-makers and beneficiaries."

To be sure, the adoption of prospective impact evaluation is likely to create tensions: to divert substantial resources from a limited budget toward impact analysis and to do it at the beginning of the programming period, one must decide in advance what the most relevant evaluation questions are. This creates the need to think harder about the design of interventions from the very beginning, making some commitment toward learning what works. *And why, of course*.

References

Angrist J.D., Krueger A. [2001] “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments”, in «The Journal of Economic Perspectives», Vol. 15, n. 4

Angrist J.D., Pischke J,S. [2008], Mostly Harmless Econometrics, Princeton University Press, NJ.

Bertrand M., Duflo E. and Mullainathan S. [2004], How Much Should We Trust Differences-in-Differences Estimates?, in «The Quarterly Journal of Economics», 2004, vol. 119, n. 1.

Caliendo M., Kopeinig S. [2008], Some Practical Guidance for the Implementation of Propensity Score Matching, «Journal of Economic Surveys», 22(1).

Cook T. [2008], Waiting for Life to Arrive: a History of the Regression-discontinuity Design in Psychology, Statistics and Economics, in «Journal of Econometrics», vol. 142, n.2.

Barca, F. [2009], An Agenda for a Reformed Cohesion Policy, Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy.

DG Regio, [2007], “Indicative Guidelines on Evaluation Methods: Evaluation during the Programming Period”, WD No. 5.

Einiö E. [2009], The Effect of Government Subsidies on Private R&D: Evidence from Geographic Variation in Support Program Funding, Discussion Paper n. 263 / May 2009, Helsinki Center of Economic Research, University of Helsinki.

Gadd H., Hansson G. and MÍnsson J. [2008], Evaluating the Impact of Firm Subsidy Using a Multilevel Propensity Score Approach, Centre for Labour Market Policy Research (CAFO), School of Management and Economics, Växjö University.

Leeuw F. [2003], Reconstructing Program Theories: Methods Available and Problems to be Solved, in «American Journal of Evaluation», n. 24(1).

Mayne J. [2001], Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly, in «Canadian Journal of Program Evaluation», n. 16(1).

Trochim W. [1984], Research Design for Program Evaluation: The Regression-discontinuity Approach, Beverly Hills, CA, Sage Publications.

White H. [2009], Theory-Based Impact Evaluation: Principles And Practice, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi.

The measurement of changes in traffic flow on sections of roads under public intervention

Michał Wolański
Dominika Wojtowicz
Łukasz Widła-Domaradzki

The increasing role of infrastructure investments – both in the context of the international competitiveness of Poland and the amount of EU support addressed to this sector in the years 2007–2013 – entails the necessity to seek effective tools for evaluating their efficiency. A main challenge in this respect is to evaluate external benefits, and therefore the long-term impact of the project. This evaluation should be based not only on theoretical data, but also on actual experience gained through previously-implemented investments and be carried out not only *ex-ante* but also *ex-post*.

Therefore, the objective of the research project discussed in the following chapter was the practical verification of a method which will allow better evaluation of the impact of implemented infrastructural projects, so making up for the deficiencies of existing solutions indicated in the previous evaluation research and referring directly to their recommendations. The research project was developed under the Grant Competition "Structural Funds on National Cohesion Strategy level" organized by the Ministry of Regional Development co-funded by the European Regional Development Fund within Operational Programme Technical Assistance 2007–2013.

The adaptation of the selected efficiency evaluation method regarding public interventions fits the actions undertaken to enhance the evaluation potential of public investments in transport infrastructure. As the results of evaluation research in this regard show, even minor streamlining of the efficiency of investments in transport infrastructure may be of fundamental importance for quality improvement, enhancing their effect and increasing the adequacy of financial resources allocated to this field.

This text presents the results of **the measurement of changes in traffic flow on road sections with the use of data derived from satellite navigation systems based on assumptions of the net effect concept** with a detailed description of the method and methodological conclusions, *i.e.* practical recommendations for future research.

Research objective

Impact evaluation assuming the net effect of a given project/programme to be estimated is one of the most advanced, effective tools for evaluating the results of state intervention, since it helps to assess its actual effect, separating out the impact of external factors. So far this methodology has not been applied in the infrastructural sector due to a range of problems related to difficulties with determining reliable success rates (e.g. in the case of the construction of junctions), collecting data, and finally selecting, so-called counterfactual situations. The latter term is construed as entities or populations which have not been under intervention, being as similar as possible to a given one. Their mutual comparison allows the determination of changes other than those being the effect of actions evaluated.

The methodological approach of impact evaluation is consistent with the approach of the European Commission, stressing that "when we say that certain effects were produced or caused by a programme, this means that if the programme had not been there or had been there in a different form or degree, those effects would not have occurred, or would not have occurred at the same level. This means that it is important to have a clear idea of what would have happened without the programme."[1)]

The research on the net effect based on the example of traffic flow is a direct continuation of the project: "Adaptation of the methodology for measuring the net effect of public interventions to the needs of the transport infrastructure sector"[2)], which successfully adapted this method to the measurement of the impact exerted by the rebuilding of junctions on road traffic safety. Already then the need was indicated for including the second rate of investment efficiency, illustrating changes in traffic flow. However it was significantly hindered, *inter alia*, due to the lack of historical data, making the assessment of the initial situation impossible.

It turned out that the solution to the problem was the use of data from satellite navigation systems monitoring the travel time on individual sections of roads, so that on this basis the fastest route could be determined and suggested to users. Though limited, these systems currently have the ability to access selected historical data, whereas in the future they may allow complete "photographs" of the network to be taken in order to fully monitor real travel times within the national road network.

---

1 *Evaluating EU expenditure programmes: A Guide*. European Commission, January 1997, p. 39

2 This research project was also realized by Koźmiński University under the Grant Competition „Structural Funds on National Cohesion Strategy level" implemented by the Ministry of Regional Development and funded by ERDF within Operational Programme Technical Assistance 2007–2013. The final reports presenting the results of both projects (in Polish version) are available in the evaluation research database http://www.ewaluacja.gov.pl/Wyniki/Strony/Wyniki_badan.aspx

Table 1. The construction of pilot research on measuring the net effect of changes in traffic flow on the sections of road under public intervention

Evaluation problem	How do road infrastructure modernisation projects (especially those relating to the modernisation of existing roads) influence traffic flow?
General research problem	How easily can data on traffic flow on various road sections be obtained – including historical data?
Research method	The measurement of the net effect with the use of the PSM method, and input data from the satellite navigation system collecting information on travel times from road users having GPS
Detailed research problem	How useful can data from the satellite navigation system be for evaluation research?
Method advantages	A strong probability of obtaining highly-useful results Innovative data source
Method disadvantages	Success of the research dependent on a data provider (satellite navigation system operators) Similar (counterfactual) sections may not be found for every researched item Population of GPS users is continuously growing, which challenges comparability of data between different time periods

Research method

Within the initial stage of the project, the providers of satellite navigation systems were identified and contact was established with the operator functioning the longest in Poland. Researchers were notified that data are available for the 2 previous years only. This fact crucially limited the selection of projects that could be included in the pilot research into method adequacy – within 24 months the following actions were supposed to be included: the measurement of traffic flow prior to and following the investment (for the purposes of data comparability – during the same months) and the process of investment implementation itself.

It was established that such conditions are met only by projects involving road repairs and rebuilding implemented from domestic funds, with the objectives being: the improvement in safety of all road traffic users, improvement in traffic flow and improvement in the operation of the road embankment drainage.

Based on the list of investments received from GDDKiA (General Directorate for National Roads and Motorways) the following was specified: possible duration of measurement prior to the investment (01.01.2008-28.02.2008), duration of measurement following the investment (01.01.2009–28.02.2009) and four projects for which it will be possible to conduct research, i.e.:

- located on the longest road sections possible (over 7.5 km);
- commenced after 1.02.2008 and completed before 1.12.2008;
- located on roads, in the vicinity and along the line of which no other investments influencing traffic volume were implemented (a counterexample may be national road No. 10, which after the opening of the A1 motorway,

took over a part of the Warsaw – Gdańsk traffic, which might have caused an increase in traffic and a reduction of its flow).

Consequently, the list of investment sections examined was specified, included in Table 2.

Table 2. Investment sections, on which the net effect of changes in traffic flow was examined

Road No.	Section beginning		Section end		Section length in km	Number of journeys	
	Location	Chainage*	Location	Chainage		I-II.2008	I-II.2009
20	Kłobuczyno	271+000	Egiertowo	281+500	10.5	812	1594
74	Bałtówka	165+600	Annopol	174+900	9.3	307	911
79	Magnuszew	57+436	Ryczywół	65+053	7.6	408	1014

** Chainage – exact location of a point on the road, given as distance from the beginning of the road. For example chainage 271+000, given for the begin of first section means, that it is located 271 km 000 m from the beginning of the road no. 20.*

It should be stressed that the initial list of investment sections to be examined, along with the reserve ones, consisted of 7 projects. Unfortunately, during analyses it turned out that in the case of three sections, the number of journeys is relatively low (especially in 2008)[3)], whereas two of them at the later stage failed to assign a sufficient number of similar sections (also referred to as alternative or counterfactual ones) – ref. Table 3.

Table 3. Investment sections, on which the net effect failed to be measured

Road No.	Beginning chainage	End chainage	Section length	Location	Reason
12	460 +700	468+260	7.6	Gielniów – Pomyków	Too few vehicles
32	93+700	99+800	6.1	Kopanica – Żodyń	Difficulty in finding similar sections
53	58+800	67+000	8.2	Świętajno – Występ	Too few vehicles; difficulty in finding similar sections
32	1+200	32+100	30.9	Gubinek – Połupin	Too few vehicles

Research was conducted in the following stages:

- The creation of a database of potentially similar road sections;
- The selection of similar sections from the database;
- Similarity verification through field visits;
- An analysis of the traffic flow rate for similar junctions (of the average travel time and the distribution of travel times – in dynamic terms);

[3] Yet it should be emphasised that in the period 2008 – 2009, based on 17 investment and potential reference sections, the number of satellite navigation system users applied increased twofold.

- An analysis of the possibility to apply the tested method in the future – preparation of recommendations.

The database of potentially similar road sections was created based on readily available cartographic sources (e.g. road atlases of Poland, Google Earth – data on road geometry) and the results of the General Traffic Survey (GTS 2005) conducted by the General Directorate for National Roads and Motorways in 2005. The database included, besides the investment sections, 450 potentially similar sections – only single-roadway sections of national roads located outside urban areas. Each section was ascribed attributes, presented in Table 4.

Table 4. The attributes of investment and potential reference sections

Attribute	Source
Section length	Maps, GTS** 2005
Existence of a hard shoulder	Maps
Number of junctions with national roads	
Number of junctions with Voivodeship roads	
Number of junctions with other hard-surface roads (local)	
Number of junctions with loose-surface roads (local)	
Approximate share of built-up areas	Maps, Google Earth
ADT* – total motor vehicles	GTS 2005
ADT – motorcycles	
ADT – passenger cars and minibuses	
ADT – delivery trucks	
ADT – lorries without a trailer	
ADT – lorries with a trailer	
ADT – buses	
ADT – tractors	
ADT – bicycles	

** ADT – average daily traffic of vehicles of a specific category*
*** GTS – General Traffic Survey by GDDKiA*

Alternative sections were selected with the use of a logistic regression model, which assumed the inclusion in the group of investment sections to be a success rate. Out of the above-mentioned variables, the following were recognised as predicators: percentage of built-up area, number of junctions with national and voivodeship roads, and ADT – in total, for passenger cars, lorries (in total), delivery trucks, other (in total).

On these grounds, an attempt was made to select several reference sections out of previously-chosen potentially similar sections for each investment section. Finally, each of these sections was assigned three subsequent reference sections, rejecting proposals suitable in terms of the statistical model, yet distinguished by either a low number of journeys (in case of 2 out of 18 potential reference sections visited)

or the fact that the field visit showed other rebuilding undertaken there or in the vicinity (this involved also 2 out of 18 sections – one of them was partially repaired – ref. photo. 1, on the other an important junction was moved – ref. photo. 2 and 3). The final arrangement of sections is presented in Table 5.

Table 5. The selection of reference sections

Investment section	Reference sections (road number, length)		
Bałtówka – Annopol	Bielsk Podlaski – Boćki (DK 19, 13.1 km)	Lipnica – Wrocki (DK 15, 12 km)	Kamień Kraj. – Sępólno Kraj. (DK 25, 9.9 km)
Magnuszew – Ryczywół	Lipnica – Wrocki (DK 15, 12 km)	Firlej – Lubartów (DK 19, 9.4 km)	Zamarte – Kamień Kraj (DK 25, 8,5 km)
Kłobuczyno – Egiertowo	Lipnica – Wrocki (DK 15, 12 km)	Kowalewo Pomorskie – Lipnica (DK 15, 9 km)	Firlej – Lubartów (DK 19, 9.4 km)

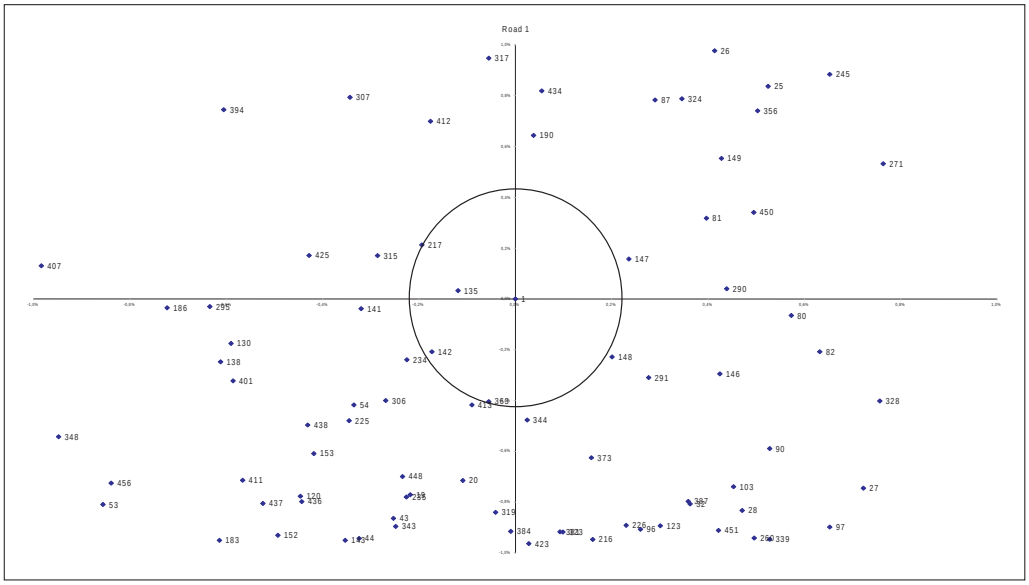
Photograph 1. The potential reference section Brzeźnica Nowa – Radomsko on DK 42 was eliminated as a result of the field visit due to the partial replacement of road surface.

Photograph 2. The section DK 62 Wyszków – Łochów was eliminated due to the opening of the ring road of Wyszków (S8), which significantly changed the traffic arrangement. Vehicles turning in the direction of Warsaw and Wyszków can use a new interchange equipped with a flyover, located between Wyszków and Łochów.

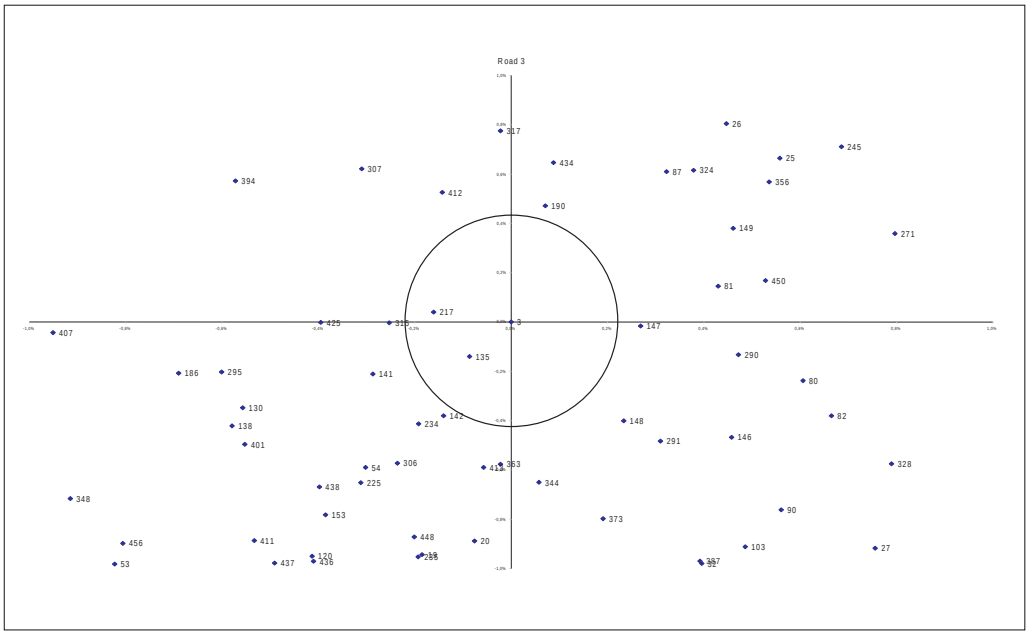
Photograph 3. Instead of an old, one-level junction in Wyszków, which generated substantial traffic jams (the necessity to cross Warsaw – Białystok traffic to turn in the direction of Warsaw). Additionally, data obtained could be distorted due to the construction of the Wyszków ring road itself.

Obviously, individual reference sections were characterised by a different similarity to investment sections. Similarity was estimated based on distances on maps illustrating two key dimensions: traffic volume (axis X) and other factors (axis Y). Three maps for three selected road sections are presented below. Owing to the length of names and number of sections initially selected, for the purposes of reference, maps included only number codes. Maps were rescaled so that a given investment road section could be located in the centre.

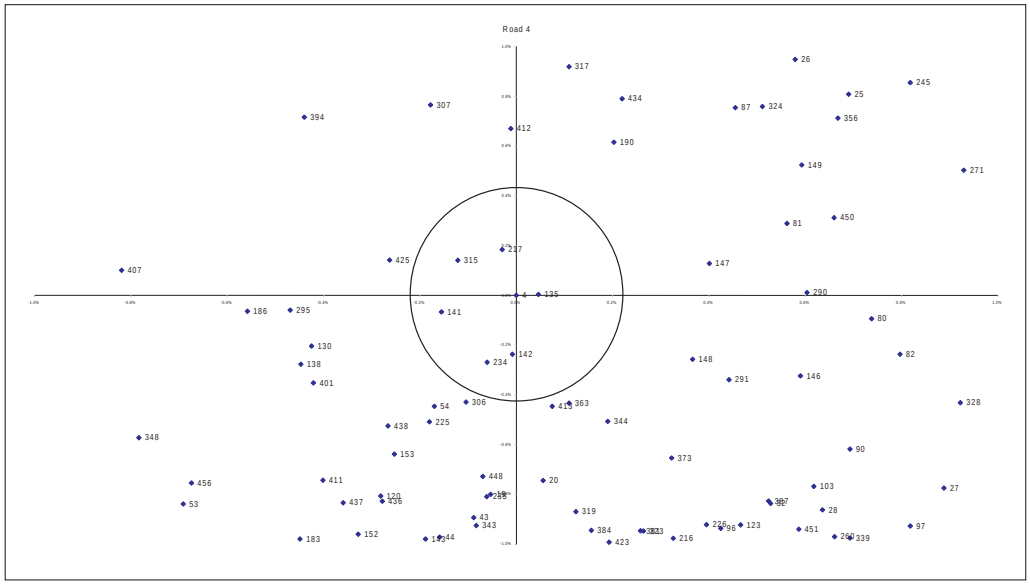
Map 1. Similarity of potential reference sections to the Bałtówka – Annopol section



Map 2. Similarity of potential reference sections to the Magnuszew – Ryczywół section



Map 3. Similarity of potential reference sections to the Kłobuczyno – Egiertowo section



Although reference sections were selected based on many factors, at the beginning we assumed that traffic flow on a given section would be a success rate. After maps were constructed, we realised which reference sections ”are located in the vicinity” of a given investment section.

Next, in order to specify one rate for a counterfactual state, a weight, being a derivative of this section’s location relative to an investment section on the above-presented maps, was determined. The weight was constructed in two stages. Firstly, based on the above-presented maps, it was stated whether a given counterfactual section could be covered faster or slower than a reference section. These weights had little value since, for the purposes of pairing up, those sections were selected that were located closest on the map. Table 5 presents the values of weights, with the use of which, data obtained from the satellite navigation system were calibrated.

The weights used in the calibration described below (a list of weights is presented in Table 5) were created based on the distance on axis OX only. Since an investment section always constituted point 0, points illustrating counterfactual sections were projected onto axis OX. Measures obtained were interpreted as distances of counterfactual sections from investment sections on the traffic volume dimension. Those values are presented in Table 6.

Table 6. The location of reference sections relative to an investment section

Investment section	Reference sections / location					
Bałtówka – Annopol	Bielsk Podlaski – Boćki	**–0.05%**	Lipnica – Wrocki	**–0.17%**	Kamień Kraj. – Sępólno Kraj.	**0.20%**
Magnuszew – Ryczywół	Lipnica – Wrocki	**–0.01%**	Firlej – Lubartów	**–0.12%**	Zamarte – Kamień Kraj.	**0.40%**

Investment section	Reference sections / location					
Kłobuczyno – Egiertowo	Lipnica – Wrocki	**–0.47%**	Kowalewo Pom. – Lipnica	**–0.61%**	Firlej – Lubartów	**–0.58%**

At the second stage, final weights were constructed showing how close (in absolute terms) a given reference section is located to an investment section. In other words, the closer (in absolute terms) a reference section was located to an investment section, the higher a weight it was ascribed. The weights of individual sections in relation with investment sections are presented in Table 7.

Table 7. The weights of individual reference sections

Investment section	Reference sections / weight					
Bałtówka – Annopol	Bielsk Podlaski – Boćki	**43.6%**	Lipnica – Wrocki	**29.8%**	Kamień Kraj. – Sępólno Kraj.	**26.6%**
Magnuszew – Ryczywół	Lipnica – Wrocki	**49.2%**	Firlej – Lubartów	**38.6%**	Zamarte – Kamień Kraj.	**12.2%**
Kłobuczyno – Egiertowo	Lipnica – Wrocki	**36.0%**	Kowalewo Pom. – Lipnica	**31.5%**	Firlej – Lubartów	**32.5%**

On these grounds, it was possible to determine the gross and net effects of a change in the average travel speed on investment sections and to analyse changes in the distribution of travel times.

The results of the pilot research

After a model was constructed, it was supplemented by data from the satellite navigation system, regarding the average travel time for all sections analysed in January-February 2008 and in the corresponding period of 2009. The data obtained (information on the average travel time calculated in minutes) were rescaled so that information on the average speed reached on a given section could be obtained, taking into account the length of individual sections. The average speed for a given section was calibrated based on weights in Table 6. The average weighted[4)] speed (km/h) is presented in Table 8.

Table 8. The values of the average weighted travel speed on individual sections [km/h] in I and II.2008 as well as I and II.2009

Section	Bałtówka – Annopol		Bielsk Podlaski – Boćki		Lipnica – Wrocki		Kamień Kraj. – Sępólno Kraj.	
Year	2008	2009	2008	2009	2008	2009	2008	2009
Vav	90.5	96.3	82.0	78.0	78.7	76.5	93.7	96.6
Section	Magnuszew – Ryczywół		Lipnica – Wrocki		Firlej – Lubartów		Zamarte – Kamień Kraj.	

[4] Calibration involved the calculation of the absolute distance of a given reference section from individual investment sections. It might have happened that a given section, occurring several times, each time had a different speed specified: tables present calibrated values for a given section.

Year	2008	2009	2008	2009	2008	2009	2008	2009
Vav	80.0	81.3	78.6	76.3	94.3	88.6	74.7	67.2
Section	Kłobuczyno – Egiertowo		Lipnica – Wrocki		Kowalewo Pomorskie – Lipnica		Firlej – Lubartów	
Year	2008	2009	2008	2009	2008	2009	2008	2009
Vav	73.6	71.1	78.9	76.7	87.8	79.3	94.8	89.0

Having taken into account the weights presented in Table 6, relative increases in the average speed reached on investment and reference sections were calculated, according to the formula:

$$(V_{av2009} - V_{av2008}) / V^{av2008}$$

The results are presented in Table 9.

Table 9. Changes in the average speed on investment and reference sections

Sections	Relative change in Vav
Bałtówka – Annopol	+6.0%
Reference (weighted average)	-2.1%
Magnuszew – Ryczywół	+1.5%
Reference (weighted average)	-5.3%
Kłobuczyno – Egiertowo	-3.5%
Reference (weighted average)	-6.6%

Final net effects for individual road sections evaluated were easily derived from such calculated relative increases. The results are presented in Table 10.

Table 10. Changes in the average speed on investment and reference sections

Section	Net effect
Bałtówka – Annopol	+8.1%
Magnuszew – Ryczywół	+6.8%
Kłobuczyno – Egiertowo	+3.1%
Arithmetic average	+6.0%

Additionally – apart from the analysis of the net and gross effect for the average speed – based on data obtained from the satellite navigation system, a compilation concerning the distribution of speed (and therefore of travel times, but in a way

comparable between sections of a different length) on individual sections was also prepared.

Firstly, the deciles of the travel time distribution for investment sections in 2008 and 2009 were compared, without having them compared to reference sections. Already owing to that, it was possible to see how the investment influenced the distribution of travel times. This information is useful in the extent that, for example in the case of an increase in speed, it allows the evaluation of whether it occurred due to the acceleration of the slowest vehicles (hence those that prior to the investment were slowed down by traffic jams) or the acceleration of the fastest vehicles – posing, as a rule, a threat to traffic safety, even in conditions of lesser traffic volume.

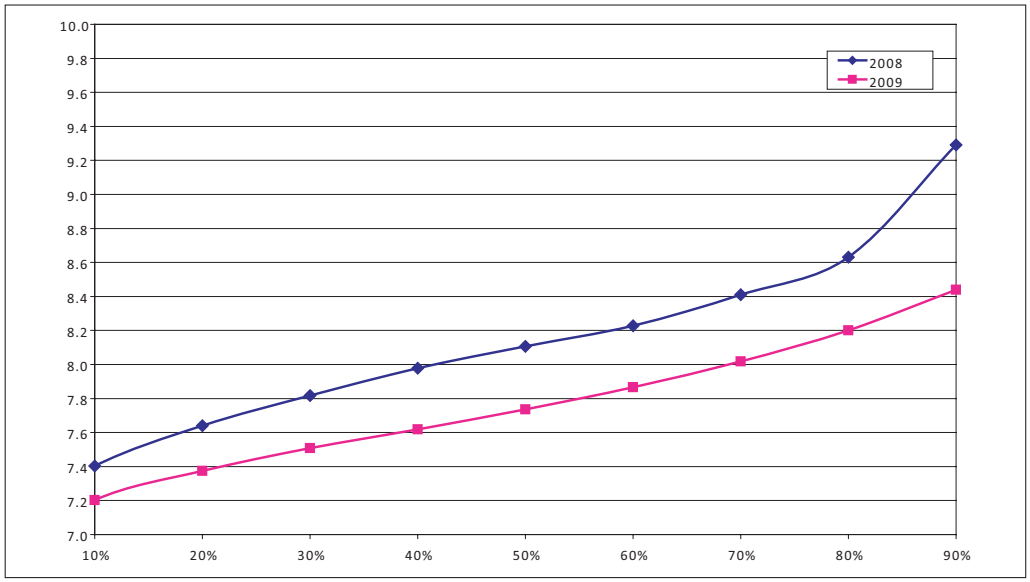
At the second stage, the results for investment sections were compared with corresponding reference sections, for which data were aggregated in such a way that it would be possible to assess traffic flow on investment sections in the context of data for similar sections.

It was decided to omit extreme deciles in the analysis, according to the principle of rejecting diverging cases (so-called outliers) so that rare incidents do not distort statistical analysis based on probability calculus. An accident or breakdown of a vehicle monitored on any section was sufficient for the measurement of its journey to distort the whole of the statistics – especially at the ends (within first and last 10%).

The subsequent diagrams present the average travel times (vertical axis, expressed in minutes) for individual deciles of vehicles (horizontal axis).

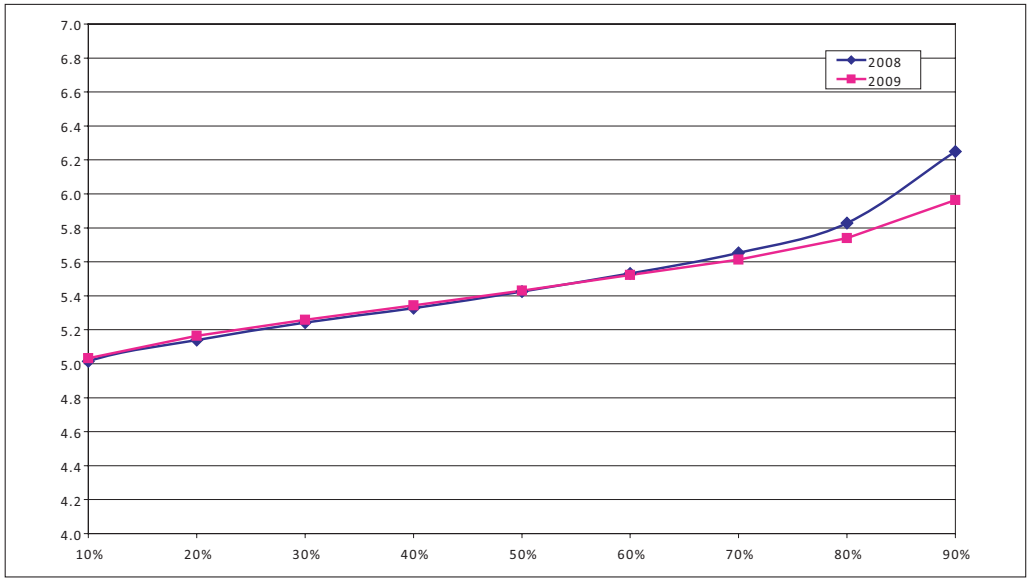
In the case of the Bałtówka-Annopol section (ref. diagram 1), the travel time was shorter than in 2008 for all vehicles. In 2008 half of the vehicles covered the Bałtówka-Annopol section on average within approximately 8.1 minutes, whereas, in 2009 the average time for half of the vehicles amounted to slightly more than 7.5 minutes. The ninth decile was at the level of 8.4 minutes in 2009, while in 2008 the result was almost a minute longer.

Diagram 1. The travel time distribution [min] for the Bałtówka – Annopol section before and after modernisation



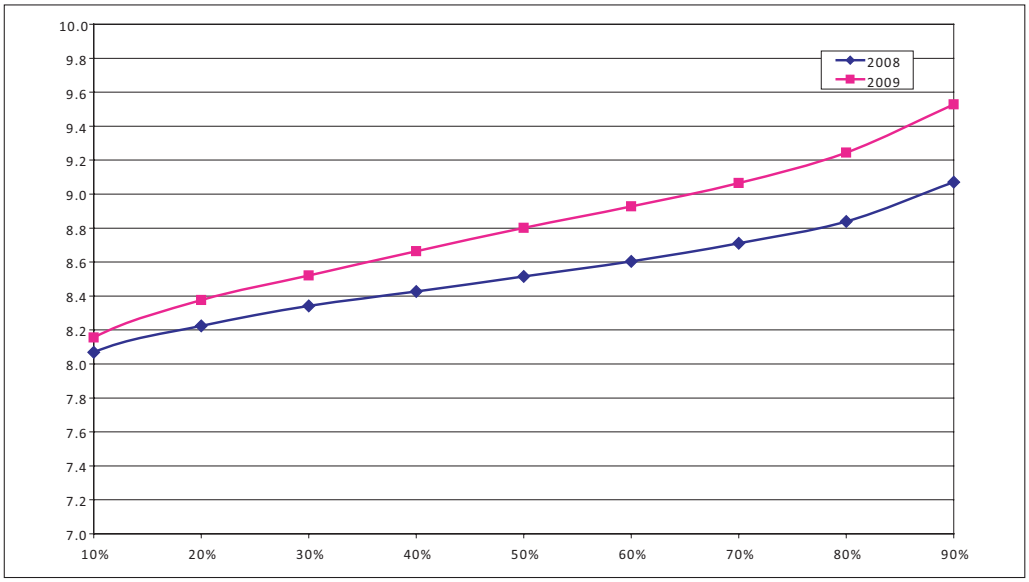
No evident differences are visible between the travel times on the Magnuszew-Ryczywół section (ref. diagram 2). Half of the vehicles examined left the section after 5.4 minutes – both in 2008 and also a year later. One distinct difference was recorded at the end of the scale: in 2009 90% of all vehicles left the section in question 12 seconds faster than in the previous year.

Diagram 2. The travel time distribution [min] for the Magnuszew – Ryczywół section before and after modernisation



In the case of the last of the sections concerned – Kłobuczyno-Egiertowo, an increase in the average travel times was observed between 2008 and 2009. While in 2008 50% of vehicles covered the section within 8.5 minutes, in 2009 this time amounted to 8.8 minutes. A slowdown is especially visible in the case of the examination of 90% of vehicles: in 2008 the average travel time amounted to slightly over 9 minutes – in 2009 it already amounted to over 9.5 minutes.

Diagram 3. The travel time distribution [min] for the Kłobuczyno – Egiertowo section before and after modernisation



The subsequent diagrams compare changes in the travel time distribution on investment sections with changes in the travel time distribution on reference sections. For the purposes of comparability, they were compared to the standard time of a fluid, safe journey determined during the field visit.

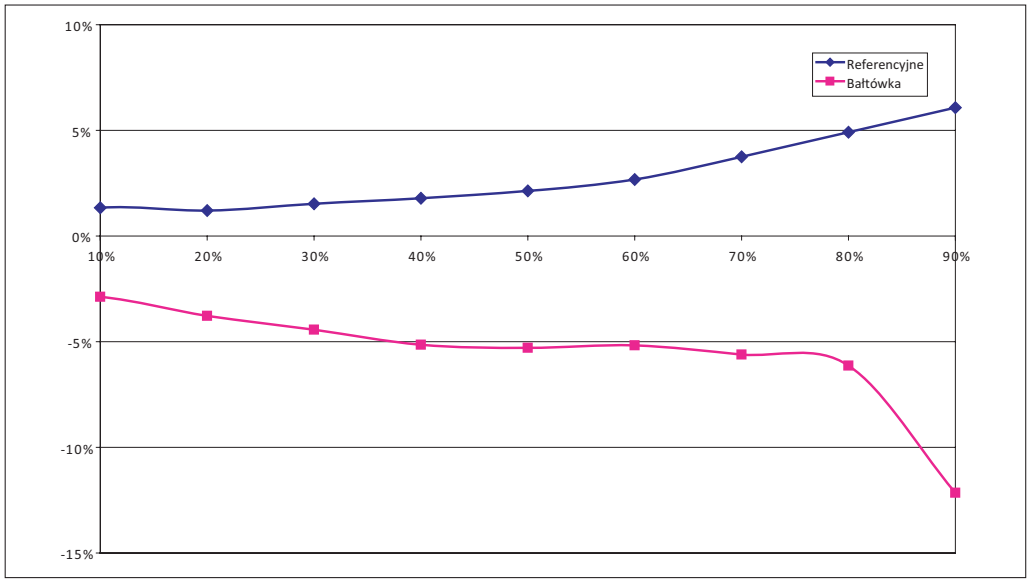
In the subsequent diagrams (4-6), the horizontal axis shows the deciles of the journey of vehicles. Whereas the horizontal axis illustrates an increase/decrease in the average travel time in relation to the standard (i.e. determined during the field visit concerning the time of a fluid, safe journey on a given section). It was calculated as a ratio of the travel time for a given decile in 2008 and in 2009. While calculating the relation, the following formula was used: $(\frac{\text{Time 2009}}{\text{Standard}} - \frac{\text{Time 2008}}{\text{Standard}})$. This allowed for the direction of relation to be determined: regardless of the standard travel time we can compare how much longer/shorter the travel time was for a given decile in 2009 than in 2008.

The value –10% means that the travel time in 2009 was 10% shorter than the travel time in 2008. A positive value on the OY axis indicates an increase in the travel time. The travel times were scaled so that the OY axis was simultaneously the standard

time, which allows us to determine whether a journey on a given section is longer or shorter than the standard one.

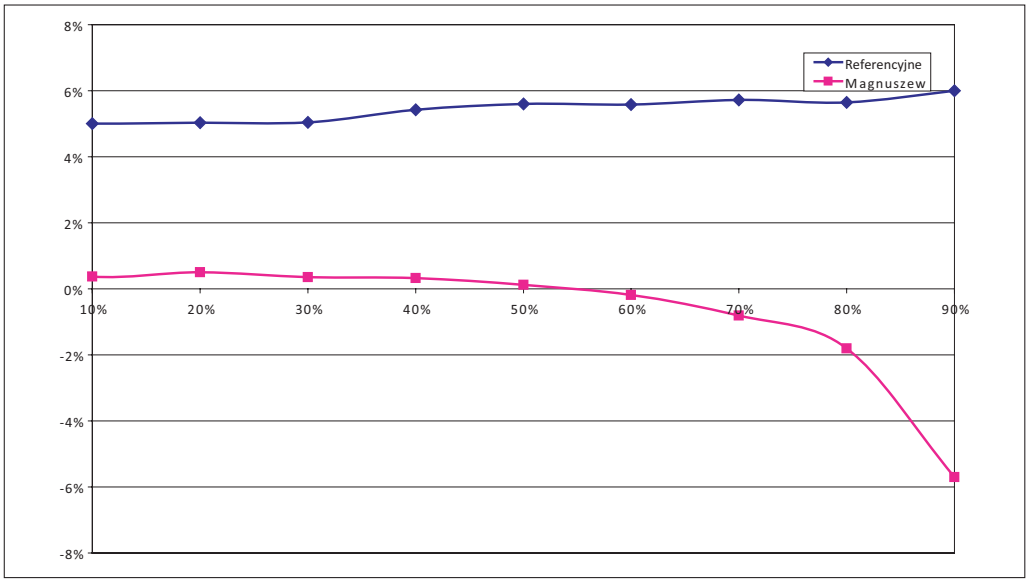
The most interesting situation can be observed in the case of the road section Bałtówka-Annopol (ref. diagram 4). The travel time for all vehicles was considerably shorter than not only the travel time on a counterfactual section, but its drop was increasingly greater – it declined with every subsequent decile. All vehicles travelled by over 30% faster in 2009 compared to 2008. In case of counterfactual sections, a real slow down occurred – by almost 12% – of the travel time on the section.

Diagram 4. Changes in the travel time distribution for the Bałtówka – Annopol section and reference sections



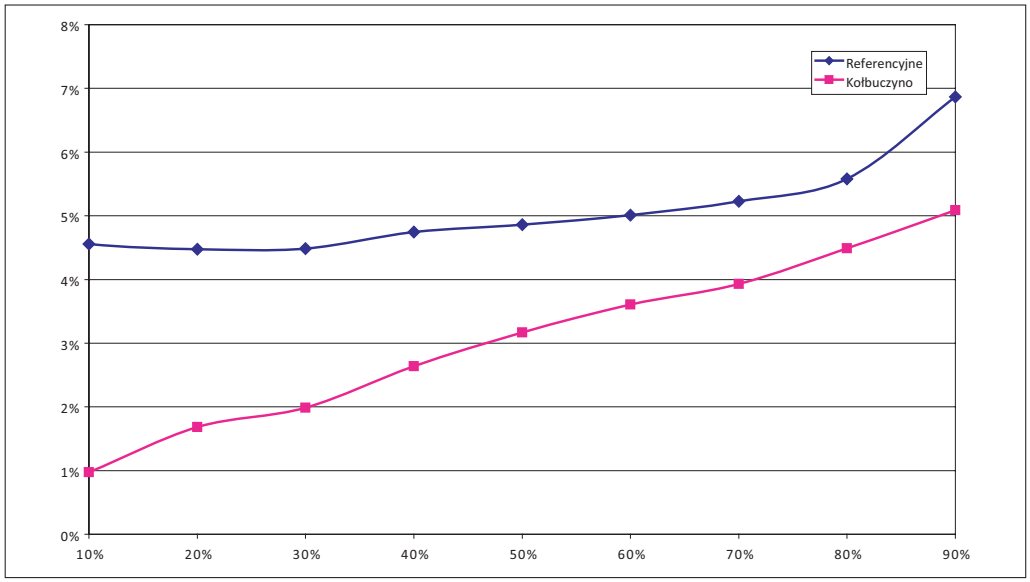
In the case of the Magnuszew-Ryczywół section (ref. diagram 5) for the majority (55%) of vehicles, the travel time remained at the same level as in the previous year. Large changes concern only the highest deciles (a drop in the ninth decile by 6%). It was also a considerably better result than in the case of counterfactual sections, for which the real travel time for 90% of vehicles increased between 2008 and 2009 by 6%.

Diagram 5. Changes in the travel time distribution for the Magnuszew – Ryczywół section and reference sections



A different situation can be observed in the case of the Kłobuczyno-Egiertowo section (ref. diagram 6). The diagram evidently shows a slight increase in the travel time for the majority of vehicles. However, in the case of counterfactual sections, a similar level of impediment to traffic capacity is observed. 90% of vehicles left the Kłobuczyno-Egiertowo section after a time 5% longer than a year earlier, yet for reference sections the average travel time increased by 7%.

Diagram 6. Changes in the travel time distribution for the Kłobuczyno – Egiertowo section and reference sections



Evaluation conclusions from the conducted research

It should be stressed that a positive gross effect was observed on two out of the three sections concerned – i.e. on the sections Bałtówka – Annopol and Magnuszew – Ryczywół. The travel time on these two sections decreased by 6% and 1.5% respectively.

In the case of the Bałtówka – Annopol section, success was visible right away, and further analyses clearly confirmed it. The net effect is as much as 8.1%, due to a slight drop in the travel time on the reference sections (by 2.1%). Hence, it should be expected that without rebuilding, the travel time on the road examined would also have increased by this value.

Shorter travel times involved, in statistical terms, all drivers, but most of all the slowest ones (first two deciles moved by only approximately 0.2 minute, whereas from the fourth to the eight deciles, the change amounted to 0.4 minutes), which reflects well on the reconciliation of traffic capacity improvement with traffic safety issues. A large drop in the ninth decile allows for a thesis to be advanced about the limitation of the longest journeys that could result from traffic jams, blocking the road by persons turning left or accidents.

Those theses are confirmed by the results of the field visit. The expansion of a junction with Voivodeship road No. 755 in Bałtówka was surely of great importance for traffic flow. The junction could previously decrease traffic flow (ref. photo. 4). It is

not without significance that this section includes long straights, which allow the overtaking of low-speed vehicles and lorries (ref. photo. 5).

Photograph 4. The expansion of a junction in Bałtówka could contribute to an increase in traffic flow

Photograph 5. The Bałtówka – Annopol section includes straights which allow overtaking.

Photograph 6. …although there are sharp turns and a built-up area.

In the case of the Magnuszew – Ryczywół section – despite the low net effect – the gross effect of shorter travel time is very similar to the previous section and equals almost 7%. It results from a significant extension of the travel time on reference sections, exceeding 5%.

An analysis of the travel time distribution shows that in the case of the fastest users, the travel time remained virtually unchanged, it even it increased minimally. Slightly greater changes are visible only in the case of the eighth and the ninth deciles, which – as in the previous case – indicates a drop in the number of situations in which a road becomes jammed.

Comparison of this data with changes on reference sections allows us to notice that a slowdown in traffic on these sections concerned both “faster” and “slower” road traffic participants equally. Hence, without the rebuilding of the Magnuszew – Ryczywół section, the travel time of the fastest participants in road traffic would probably have also increased.

The results of the field visit show that the investment was conducted with concern for road traffic safety. Even at junctions with very small roads, it was decided to section off short left-turn lanes (ref. photo. 7), which helps to prevent vehicles running into other vehicles awaiting an opportunity to turn (“pile-ups”). At the same time, it obviously increases traffic flow, which is also positively influenced by a large number of straight sections.

In built-up areas, the priority was to improve safety (ref. photo. 8), which was reflected in the construction of islands calming the traffic and making it impossible to use turn-lanes for overtaking. Such activities could reduce the gross effect of the changes in traffic flow, however they are completely justified. Some controversy might arise from the construction of a roundabout at the junction with the Magnuszew – Przewóz road (ref. photo. 9).

Photograph 7. On the Magnuszew – Ryczywół section, even at minor junctions, short turn-lanes were sectioned off, which prevents the blocking of traffic and reduces the number of collisions and accidents

Photograph 8. Within built-up areas – especially in the Magnuszew village, being the HQ of the commune office – many islands and pedestrian refuges were placed to calm the traffic

Photograph 9. At the three-entrance junction with the road to Przewóz Stary, a roundabout was also constructed, which might give rise to some controversy (such a solution was applied at the junction with DW 736).

An extreme situation occurs in the case of the Skrajna Kłobuczyno – Egiertowo section, where a negative gross effect was observed, which means the real travel time increased on average by 3.5%. Moreover, this concerned mainly higher deciles, meaning persons who already travelled the slowest. This is an additional negative sign because if a slowdown concerned low deciles, it could be explained by actions aimed at road traffic calming. This section is unique due to the fact that it is the only investment section where an increase in the deviation of the standard travel time was recorded, a characteristic of sections on which no rebuilding was carried out – this rate grew for 12 out of 14 potential reference sections, for which data were obtained.

However, after thorough analysis, it needs to be stated that the eight decile, with an average speed of 68 km/h, and the ninth decile – 66 km/h, do not prove a frequent source of blocking due to traffic jams, but rather slow driving at the pace of lorries, or momentary blocking of traffic flow by turning vehicles. The field visit confirms this thesis since the road is winding and there is no possibility to overtake. But first of all – in contrast with the previously analysed investments – at many junctions there are no left-turn lanes, which reduces both traffic flow and safety (ref. photo 10 and 11). It should be pointed out that this may result from the location in a hilly, wooded area, which would considerably hinder the construction of such lanes – yet perhaps two kinds of benefits would compensate for an increase in costs.

However, the rebuilding performed cannot be assessed negatively since on reference sections the travel time lengthened even more (by as much as 6.6%), which consequently means a slight, but still positive net effect (3.1%). Therefore, this indicates that without the rebuilding, the slow down in travel could be even greater.

Photograph 10. The Kłobuczyno – Egiertowo road is very winding. There are no convenient places for overtaking, whereas traffic flow and safety are significantly reduced due to the lack of left-turn lanes (although some junctions are located directly behind turns). This is the case at a junction with the road to Brodnica Górna.

Photograph 11. ...as well as to Szymbark and Stężyca.

Conclusions on the application of the method tested

The pilot research conducted concerned in fact two innovative solutions:

- the use of data collected on a large scale on real travel times on road sections in order to evaluate infrastructural investments;
- the measurement of the net effect of traffic flow on road sections.

As far as the data source applied is concerned, its serious limitation is the limited access to historical data, which translates into the necessity to analyse data from

only two months, which in turn resulted in a limited availability of data on some sections.

However, it should be stressed that this could be improved in the future, if "photographs" of travel times on major road sections were taken based on representative data for a whole year – yet such "photographs" need to be constructed specifically to evaluate transport policy because they are unnecessary for companies servicing satellite navigation systems for conducting their basic activity.

The data obtained would allow for the performance of analysis on two levels:

- the entire transport policy through an analysis of matrices of the shortest travel times between agglomerations;
- individual projects.

In the first case, the data acquired can be analysed:

- in a simple manner – through analysing the variability in the rate of the average travel time between agglomerations, however the inclusion of parameters of its distributions is essential (e.g. standard deviation – the lower, the better – or deciles for median 95% of the sample collected, after extreme cases have been eliminated) so that it is possible to interpret the reasons for changes, which should head into two directions – traffic streamlining (the elimination of the slowest journeys) and greater respect for traffic regulations (the elimination of the fastest journeys);
- as the data source for an algorithm of calculating the rate of inter-branch transportation accessibility of Poland, prepared by the Institute of Geography of the Polish Academy of Sciences – the Authors of which have currently assumed constant values for the average speed for specified categories of roads, depending on population density and landform features[5)], which makes it impossible to e.g. assess the effects of conversion of a one-lane national road into a so called, 2+1 arrangement (two traffic lanes in one direction, and one lane in the other direction), to add turn-lanes, or to extend key junctions, that is actions having a potentially high relation of cost to benefit.

In the case of single investments or programmes, two types of analyses are also possible:

- in a simple manner – as above;
- with the use of the net effect, which can be calculated in the case of smaller modernisation investments due to relatively large number of comparable items (for example road sections); yet according to the method

[5] Preparation of the methodology for calculating the rate of inter-branch transportation accessibility of Poland and its estimation (Opracowanie metodologii liczenia wskaźnika międzygałęziowej dostępności transportowej terytorium Polski oraz jego oszacowanie), joint publication edited by T. Komornicki, Institute of Geography and Spatial Organization, Polish Academy of Sciences, Warsaw 2008, p. 19–20.

applied in this paper, it is advisable to carry out field visits and to describe a counterfactual state with the use of several reference sections); obviously the situation will be different in the case of, e.g. the construction of motorways, where the gross effect will be so strong compared to changes in the counterfactual state that the estimation of the net effect might be meaningless. Additionally, it will be difficult to find a reference section.

The most crucial threat related to the application of the data source concerned is a changeable group of users applying the satellite navigation system, and in particular its structural changes (e.g. it being joined by persons driving occasionally, hence more carefully, which might be erroneously interpreted as a greater respect for the road traffic regulations by the whole of the population).

An advantage of the research on the net effect is surely the "corrected" occurrence of a range of external factors, such as structural changes in a group of satellite navigation system users, which will take place on reference and investment sections to a similar extent. However, it should be noted that in the cases examined, there were no divergences between conclusions resulting from the research on the gross changes in the distribution of travel times and the net effect. Therefore, a detailed analysis of the gross effect may lead to valuable conclusions.

It should be remembered that an increase in traffic flow ought to be each time considered as one of two basic factors in measuring the efficiency of programmes and projects implemented. The second factor should be road traffic safety. Admittedly, some actions (e.g. the construction of left-turn lanes mentioned several times) contribute to the improvement of both indicators. However, in many other cases there is no such coincidence. Therefore, one should follow both indicators at the same time to obtain their optimal relationship.

Summary of conclusions and recommendations

The research conducted presents a highly-useful method in EU fund management, which was positively verified in the pilot research, and so far has not been applied in evaluation research.

Data from satellite navigation systems allows for easy and cheap monitoring of road traffic flow. It makes it possible to determine the real transportation accessibility of Poland and the efficiency of interventions undertaken – on the level of projects and programmes – in particular, in the case of the rebuilding of national roads based on the philosophy of clearing "bottlenecks" by means of methods having a good relation of cost to benefit.

This monitoring should be based not only on the value of the average speed but also on the analysis of the travel time distribution, which will allow the frequency of congestion ("traffic jams") and changes in drivers' behaviour to be determined.

With good access to historical data and quite typical national road sections, it appeared to be possible to **measure the net effect**, useful both in traffic flow analysis and – as was proven in the previous research – in assessing the impact of the investment on safety.

Due to problems with the access to historical data, it is advisable to monitor the situation and archive the information selected on a current basis for evaluation purposes.

Applying systems thinking to evaluation of Structural Funds

Richard Hummelbrunner

1. Using systems concepts in evaluation[1)]

Although the fields of systems and evaluation share many experiences, concepts and goals, they know relatively little about each other. What each understands about the other is often crude and partial. Despite drawing on some of the same philosophical, sociological and scientific developments, the two fields have operated virtually independently since their inceptions. In recent years, however, some systems practitioners have begun applying systems thinking to evaluation work. And today there is growing interest among evaluators in what the systems field can offer them.

Systems inquiry and evaluation tend to emphasize different understandings of both the task of inquiry and the situation under study. Even when it poses the same questions as an evaluation, an inquiry using systems concepts is likely to interpret the answer from different perspectives. These include seeing the complicated as simple (but not simplistic), being critical of boundaries that define the frame of inquiry, and the notion that better insights are more likely to promote valuable action than more or better data.

Evaluations influenced by systems concepts are likely to generate rich yet simple descriptions of complex interconnected situations. In systems work, richness implies that the whole can only be understood as a product of its parts and the dynamic relationship between those parts. This does not imply that systems approaches to evaluation have to include every component of that situation plus its context and environment. In fact, the implication is the opposite. Including everything in a systems inquiry does not necessarily provide any deeper insights about the parts, nor does it necessarily offer more insights into the whole. Instead, a systems based approach to evaluation is concerned with what can be reasonably left out of the inquiry – but is also deeply and openly aware of the consequences.

The richness of a systems inquiry is not about detail but about value. And the value is contained in the relevance of the inquiry to those affected by it. This is why systems influenced approaches to evaluation are based on multiple perspectives that build stakeholdings in the situations being addressed. Stakeholders are not passive players or mere informants – they are actively engaged in the entire process. Multiple perspectives frequently reveal several (and often divergent) purposes. The

[1] Source: Expert anthology "Systems concepts in evaluation" (Williams and Imam, 2006, pp. 3–5)

inquiry itself as well as the use of multiple perspectives can create new options for changing a situation.

1.1. Three core systems concepts for evaluators

Three core concepts have emerged in the historical development of the systems field over the past fifty years (Hummelbrunner, R & Reynolds, M 2008, pp. 7–9):

- **Inter-relationships**

This is the most familiar systems concept, partly because it is also the oldest: how things are connected, by what, to what and with what consequence, stems from the earliest thinking about systems. During the 1960s and 1970s the focus was very much on inter-relationships, and methods were developed that explored these in depth (e.g. system dynamics). It is also the concept most strongly embedded in the popular imagination. One can observe and perceive systems within systems, systems overlapping other systems and systems tangled up in other systems. Thus, one should avoid focusing on just one (definition of a) system without examining its relationship with other systems.

- **Perspectives**

By the mid-1970s it was clear that the inter-relationships were not neutral, but that the importance of particular inter-relationships depended on the ascribed purposes. Thus, methods were developed that helped explore the implications of different perspectives that could be taken of the same situation (e.g. soft systems methodology). A systemic approach is more than exploring interconnections. What makes an inquiry "systemic" is how to explore and interpret them. People participate in programs for many different reasons. These motivations and the behaviours that flow from them may have little or nothing to do with the formal goals or objectives, yet they will affect how the program performs and what the results are. Hence, understanding perspectives helps to comprehend the functioning of a program and to explain – or even foresee – behaviours.

- **Boundaries**

During the mid-1980s it became clear that perspectives were not neutral either. They determined what was relevant and what was not, what was "in" the system and what lay outside it. Whoever defined the dominant perspective controlled the system's boundary. Therefore, the importance of studying boundaries and critiquing boundary decisions became the third key element of a systems approach. There is often a lot of energy around boundaries – they are the sites where values get played out and disagreements are highlighted. A lot of power issues are bound up in boundaries – whoever's perspective dominates decides the boundaries. Boundaries do not just define difference, but are the sites where "differences make a differ-

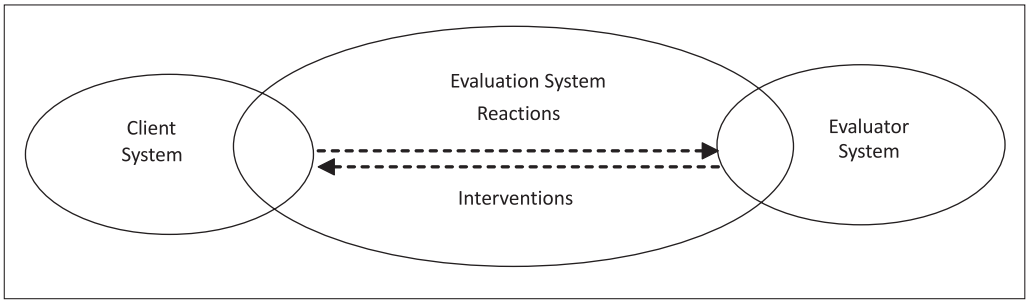
ence". Thus, systems approaches take a deliberate and often debated approach to boundary identification and boundary choice.

These three concepts are essential both for understanding systems based interventions and for distinguishing them from other approaches for dealing with complex situations. They underpin the models, metaphors, methodologies and methods used by the systems field. They provide the key to unlocking the potential benefits of systems approaches to evaluation. The important thing to remember is that all systems methods, no matter when they were developed, are used with these three basic system concepts in mind.

1.2. Evaluation as a system

Evaluation itself can be regarded as a system. Basically, it is a specific case of observation which takes place in a joint system established between two main partner systems:

- *Client System*: Consists of the funders, the operators (i.e. managers of the program/project to be evaluated) and the concerned public (other stakeholders such as beneficiaries, implementation partners, additional intended users of evaluation results).
- *Evaluator System*: The experts which are commissioned to undertake the evaluation.



The Evaluation System is usually established by contract and is limited in time. It has a joint focus based on the evaluation purpose and a structure to serve it. Elements of this structure are nodes of communication (e.g. steering group, meetings, workshops) and their respective linkages as defined in the design of the evaluation process (e.g. work packages, activities).

The Evaluation System incorporates elements of the two constituting partner systems, and the figure above can be used to establish the respective boundaries. For instance, which elements of the client system take part in specific nodes of communication (e.g. steering group, workshops)? Who from the client system participates in – or even carries out – evaluation activities (e.g. self-assessments, surveys)? An important question to be clarified in any evaluation assignment is whether (and to what extent) the funders of the program/project are to be included in the evalu-

ation. From a systems perspective, they are always part of the Evaluation System, whether they like it or not.

All three systems are part of each other's environment. Whatever happens in the Evaluation System can affect the constituting systems, and vice versa. And any evaluation is an intervention in the Client System, bringing forth reactions in the latter, which again has an influence on the Evaluation and Evaluator Systems and so forth. It is a circular process, by which all three systems mutually influence each other.

1.3. Evaluation as an intervention

Intervening means to apply external influence upon a system with the aim of inducing change. But social systems are self-determined and can only change themselves. This cannot be done by an intervener, no matter what resources, power, etc. are applied (at least not in the long run and in a sustainable manner). Due to these preconditions, interventions in social systems cannot be directive, aimed at producing intended effects in a linear way. They always bear certain risks and their outcomes are uncertain.

"Systemic" interventions are targeted forms of communication between (social) systems, which respect the autonomy of these very systems. Although they are conceived externally, they should be designed in the terms of the system which the intervention is aimed at. Therefore, those intervening should have a profound understanding of the target (client) system's internal structure and act accordingly. This requires, for example, to use similar language, respect its rules and behaviour patterns, build on existing concepts and values or address topics which are conceived as being relevant by the target system.

Systemic evaluation is an intervention in a Client System with the essential aim of modifying the client systems' internal state in order to improve the chances for producing the desired effects in a sustainable manner. This means to increase the client's capacity to understand the situation at hand, solve occurring problems and change in a way that contributes to achieve appropriate solutions. At a more operational level, evaluators should take these complex relations and linkages into account when planning or implementing their interventions during the evaluation process (e.g. interview sessions, focus groups, surveys, presentation of findings).

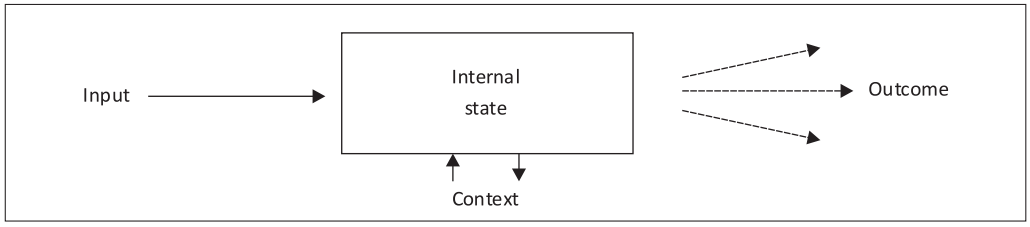
Interventions, and hence evaluations, should be based on prior hypotheses and conceived as a circular process. This means that before intervening, information about the client system should be collected, and hypotheses about the situation as well as the intended effect(s) of the intervention should be formulated. After the intervention, information should once again be gathered about its effects, and based on this information prior hypotheses reformulated.

2. Systemic evaluations in the field of regional development

2.1. Territories as open living systems

Territories are open and living systems, they are in continual exchange with their environment and their elements/subsystems can change over time. The various subsystems (e.g. political, administrative, economic, social) interact within a given territory and the actions of one system might lead to adaptation processes of another system (and vice versa). Programs or projects are implemented in such a context of interacting social systems and, due to their non-linear behaviour, rarely act one-way, but might also trigger processes which can neither be foreseen nor reduced to the original plans or intentions.

The structure of the ”target” system (e.g. territory) is decisive for the success of an intervention (e.g. development program). Because social systems are “non-trivial”, they can react differently at different times to the same input (= intervention), depending on their internal state. Their behaviour (outcome) is not linear, inexplicable from inputs or internal states, but results from the interaction of both:



The resulting behaviour of open living systems has important implications for evaluations:

- Relationships between cause and effect are neither linear nor transparent. Every action can be both cause and effect; therefore “linear” cause-effect links are replaced by “circular” interaction patterns, which consist of feedback loops and regulate the behaviour of a system. Since it is difficult to trace all the linkages or effects of an action, its impact on a given system can never be thoroughly analysed nor understood.
- Changes are essentially self-organised: open systems develop their own internal mechanisms of regulation and stabilisation (autopoiesis) and cannot be controlled externally – at least not in a direct mechanistic sense. Because changes in the system’s environment are ambiguous, they can be disturbing and trigger corrective as well as defensive action – or can be the source for further development, leading to modifications in relations as well as inspiring new (inter)actions. And adaptation processes in open social systems follow internal mechanisms rather than external influence (e.g. recommendations of evaluators).

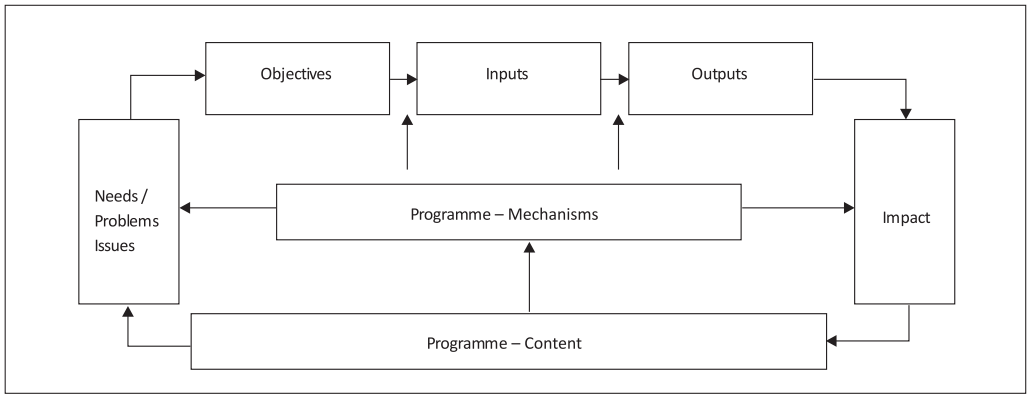
2.2. Systemic evaluation framework for regional policy

Viewing evaluations as a system also has important implications for defining their content, notably when determining an evaluation´s coverage, scope and level. From a systemic perspective two principles should be kept in mind in this respect:

- The program/project to be evaluated (evaluand) should be structured as a system, i.e. outlining the essential elements and their relations;
- The unit of observation should be the evaluand and its environment, i.e. what lies beyond the boundaries of the evaluand when seen as a system.

In the case of regional development programs, their basic elements consist of objectives, inputs and (expected) effects, i.e. outputs and impacts. The main relations are the mechanisms which link these elements; for example, planning and decision-making mechanisms that determine how inputs (e.g. financial, human, physical resources) are applied in order to achieve objectives. Or implementing mechanisms (activities, management arrangements) which are foreseen to transform inputs into outputs. In addition, programs take place in an operational context, which influences implementation in multiple ways.

Thus, local/regional development programs can be structured as systems by using the following extended version of a logic (or change) model:



It is important to note that the basic elements are linked to mechanisms and context in recursive logic. Therefore the achievement of effects (i.e. outputs, impacts) is not seen in an isolated manner, but takes the actual functioning of the program or relevant context conditions into account ("what works for whom and in which circumstance"). Impacts are conceived as the result of specific mechanisms acting in a specific context, linked by feedback loops. Because impacts modify the context, this has potential effects on program mechanisms, which in turn can affect the transformation of inputs into outputs and impacts, and so forth[2]. For instance, training activities or the establishment of facilities in the early stage of a program

[2] These recursive links are the main difference when compared with the "Realist Evaluation" approach, which only foresees linear relations between context and mechanisms (See article by Pawson, R in this volume).

changes initial context conditions and influences the way in which activities can be carried out at a later stage.

By systematically distinguishing between internal (=mechanisms) and external (= context) factors, evaluations can indicate where and how the actors involved in implementation should modify their mechanisms in order to improve the achievement of effects. Even if a program's theory of action has not been structured beforehand with such a systemic framework, it can be reconstructed during an evaluation, based on an assessment of those mechanisms and context factors which can be observed as being influential.

Open, living systems remain in balance because of continuous renewal of their elements – their only constant is change. Therefore, differences from original states are inherent to assure their stability. And adjustments in short-term targets or plans are often necessary for the achievement of long-term objectives. Evaluations of programs/projects that take place in such systems (e.g. territories) need to be aware of these characteristics.

This requires above all a different attitude towards deviations. Analysing differences in output (as well as results and impacts) from original plans can help to assess the appropriateness of a program in view of the given operating environment (e.g. framework conditions, needs of target groups, interests of implementing partners). But it can also provide valuable indications about the internal dynamics and self-organising forces that are at work, thus improving the understanding of a program and its operating environment.

For evaluations to work this way they should not be limited to observing intended routes, but look at the entire range of processes triggered and outcome produced, irrespective of whether they are in line with original intentions. Exceptions, discontinuities, unexpected results and side effects are valuable sources of information. They can provide useful clues – e.g. for relevant internal/external changes, newly emerging challenges, innovative or "informal" ways of handling situations – which can help to improve implementation.

If this is not taken into account, evaluations risk counteracting these internal mechanisms, which might lead to false conclusions and even counterproductive recommendations. For example, insisting on the implementation of original plans despite relevant changes in the operating environment might have counter-intentional effects and ultimately lead to failure.

2.3. Dealing appropriately with Stakeholder differences

Evaluations often reveal a rather diverse picture of a program/project, particularly when viewed through the eyes of various stakeholders. However, unbalanced or simplistic attempts to reduce this complex picture will not only harm the credibility of an evaluation, but also bring forth resistance from those who do not feel

themselves properly represented. Systemic evaluations should go beyond merely illustrating different stakeholder opinions (e.g. by visualising diversity through rating and ranking) and effectively work with them.

Systems thinking can be of great help to avoid undue simplifications and provides useful tools for dealing practically with differences. It can contribute to looking beneath the surface of observable phenomena and to identifying underlying patterns and causes. By revealing core dynamics it can help to provide simple – not simplistic – insights into the functioning of complex systems. And systems approaches can also contribute to improve the use of evaluations. Because generating new insights and improving joint understanding of issues across a range of stakeholders is an important pre-requisite for sustainable learning effects.

Each stakeholder (group) constructs an internal mental map for orientation, which serves as a frame of reference for action or choosing among alternatives. Thus, the existence of diverse and often conflicting views of different individuals or groups should be seen as the rule – and not as an exception from the ideal of one single truth or logic. And these differences cannot be solved by giving preference to one particular view or by synthesising them through an objective judgement (e.g. by an evaluator). Because ultimately everyone is right – but only within the boundaries of the respective mental maps!

Under such circumstance, consensus should not be taken for granted and is rarely achievable in evaluations. Yet, learning can also consist in improving the understanding of – and changing the attitude towards – stakeholder differences. Differences can be regarded as part of a complex reality, where the picture of the whole can only emerge by viewing it from multiple angles, and by using or adding different explanations in a joint dialogue aimed at reconstructing reality. And this can most likely be achieved if the various actors confront their mental maps, become aware of their own limitations and more receptive for the views of others. In short, learn to see things also with the eyes of others.

Moreover, differences among stakeholders should be treated as a resource instead of an obstacle. They can improve mutual understanding if mental maps are made explicit and visible for others. And they can contribute to a more complete picture of reality by linking individual mental maps and working towards the emergence of collective mental maps.

3. Application of selected systemic tools with Structural Funds

This section contains some selected systemic tools which have been used by the author – or his colleagues from ÖAR Regionalberatung – in evaluations of regional development and Structural Programs in Austria.

3.1. Circular Dialogue

Dialogue techniques are based on three key assumptions:

- Different observation positions make a difference: Exposing oneself to different perspectives helps to overcome mental barriers or unilateral thinking and to find solutions and answers that are acceptable for all.
- A system can only be changed from within: A dialogue never aims at direct influence or persuasion, but rather at referencing experiences which enable the partners to change their "mental maps".
- Language counts: The purposive use of specific language patterns helps to understand the partner's mental map "from within" and at the same time to overcome its limitations.

Whereas in discussions or debates standpoints are confronted and – if possible or desired – harmonised with each other, a dialogue is a means to foster collective intelligence. A well-structured, meaningful dialogue is a gain for all individual participants and for the group's state of mind. Contrary to a debate it does not make sense to try and "win" a dialogue.

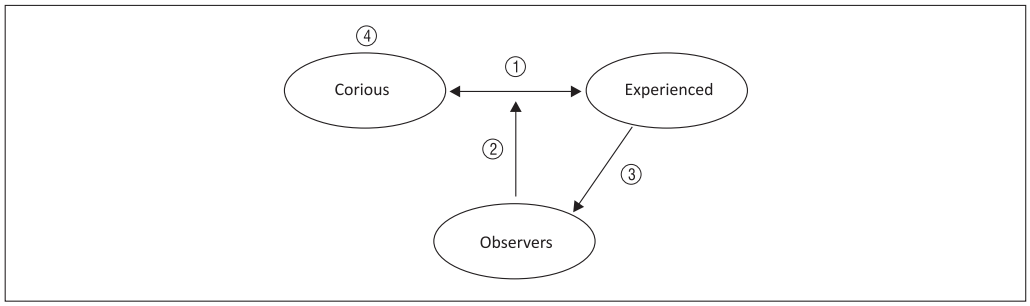
In Circular Dialogues, participants have the opportunity to perceive a given theme from at least three perspectives. Guided by facilitators, participants are asked to communicate in a strictly structured manner, mutually interviewing and observing each other without direct discussions. The participants represent different roles and are invited to contribute from various perspectives.

For example, a Circular Dialogue on a subject in a diverse audience with a varying degree of familiarity on the subject could be built around the following roles (see figure below):

- Curious: They have little experience on the subject, but want to find out more
- Experienced: They have some experience on the subject and are willing to share this knowledge
- Observers: They are asked to observe and comment, if considered useful this role can be split up further (e.g. the sceptics, the convinced).

The Dialogue session would take place in the following sequence:

1. It starts with a question-answer session between the Curious and the Experienced.
2. Then the Observers comment on this sequence from their respective perspective.
3. Next the Experienced respond to the comments of the Observers.
4. At the end, the Curious conclude what they have learned from the entire dialogue session. And they might be invited to start another round…



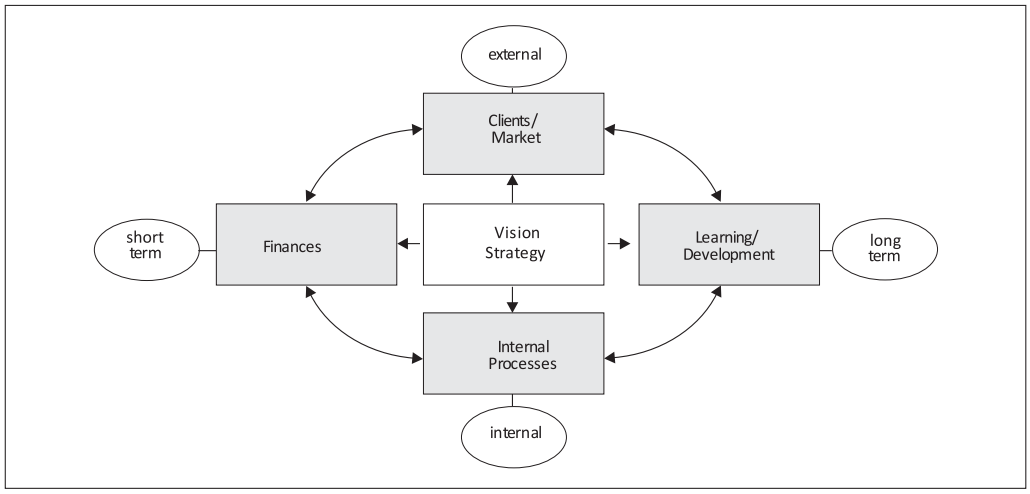
There are many variations to this basic sequence: roles can be assigned according to specific functions (as in the example above), or they are defined in order to represent stakeholders. The facilitator can either remain passive (e.g. a time manager) or have a more active role (e.g. intervening to clarify questions or answers, re-focus the dialogue). Circular Dialogue is a simple yet powerful tool to constructively work with stakeholder differences. By organizing participants as a "learning system", the resources of different viewpoints and roles are made effective and new understanding can emerge. In evaluations it can be used to organize stakeholders for interpreting data sets or discussing preliminary findings. This can either take place in smaller settings or in large groups sessions.

3.2. Balanced Scorecard (BSC)

The BSC was developed at the beginning of the 1990s by a team at Harvard University led by Kaplan and Norton to counteract the strong emphasis by American management systems on the financial management aspect in all matters of planning and reporting. To this end, they looked for a more "balanced" set of both financial and non-financial measurement parameters. It turned out that just a small number of essential measurement parameters was enough to meet complex demands for steering an enterprise.

The purpose of the BSC is to translate strategy into action. To this end a strategy is differentiated in terms of four distinct perspectives (see figure below): financial management, the customer/market perspective, internal (business) processes, learning and (employee) development. To assess performance, a set of indicators is defined for each perspective and assembled in "scorecards". However, none of these indicators sets stand alone; they are connected to form "feedback loops". This conscious and deliberate linkage should ensure that no indicators are defined which are without relevance to the strategy and/or the other perspectives, and that potential conflicts between goals are easier to perceive.

When defining the indicators, care should be taken to achieve a balance by addressing monetary and non-monetary aspects, using variables that measure internal and external perspectives, and identifying "lagging" (backward) as well as "leading" (forward) indicators, which can provide early warning signals of change or problems.



In Austria, the BSC is for instance used to monitor the performance of decentralised components of the National Rural Development programs (regional programs established under the LEADER priority axis). A set of 15 indicators was established by a working group of representatives from the national and regional level. These indicators are grouped in four dimensions according to the BSC model, modified for use in rural development: results and impacts, implementation process, learning and development, resources. The Local Action Groups assess their performance on an annual basis, normally by consulting a range of concerned actors in their region. The assessments of the individual programs are then aggregated and analyzed at the national level. Finally, this comparative analysis is discussed by a quality assurance working group established at the national level.

The main purpose of this monitoring is to assure and improve quality by comparing and reflecting the performance of individual programs. The national level expects to gain insight into the implementation status of the programs and to identify areas where additional external support would be appropriate. It is not used to rank the programs, nor to sanction or reward them according to their performance.

3.3. Process Monitoring of Impacts

Monitoring Systems for Structural Fund Programs are essentially input driven and focused on inputs and outputs. Their problems and limitations are widely acknowledged and the current Regulations foresee a shift in focus of Monitoring and Evaluation towards impact and strategic goals. The “Working Paper on Indicators” recommends a more impact-led approach and emphasises result indicators as a core management instrument. But the sole use of indicators has only limited value for capturing results, because the information on their achievement arrives rather late (often too late for corrective action) and it is frequently difficult to provide evidence for the links between effects and program activities.

This is the rationale for applying a different approach to monitoring which observes the processes that are expected to lead to results or impacts – and not just indicators as their final measure. This is the core of the "Process Monitoring of Impacts" approach (Hummelbrunner *et al*. 2005). It builds on the basic assumption that inputs as well as outputs have to be used (by someone, in a specific manner) to produce desired effects. Thus, focus is placed on those uses of inputs or outputs (by project owners, beneficiaries, etc.), which are considered crucial for the achievement of effects and can be influenced by program actors.

To this end, assumptions are made about these uses and described as processes (activities, behaviour or communication patterns). They are inserted in impact diagrams (or logic charts) as decisive connections between inputs and desired effects, and they must be observed during implementation to check whether these links remain valid and actually take place.

However, the impact diagram and the assumptions contained therein are "mental maps" and should not be confused with reality. Thus, care should be taken to capture the entire range of effects (including those unintended or unexpected) which can be observed in a defined area. And to regard deviations from intended routes not a priori as negative phenomena, but deal with them in a more differentiated manner. Differences between plan and implementation as well as exceptions or unexpected effects are important sources of information for learning and improving implementation, as they can help to identify weaknesses, point at possible alternatives or lead to new solutions.

This monitoring approach responds to the information needs of impact-led management and produces information needed to understand impact-creating processes. Moreover, it is not necessary to wait until a chosen indicator target is met for assessing the achievement of results. Instead, understanding and observing the underlying processes provide early indication whether a project / program is on the right track or risks missing desired results.

A core challenge is to limit the administrative burden for the collection of data and information by integrating as much as possible into existing work routines. Therefore, project applications or reports should be used as information sources and evaluators for analyzing the data.

Since its original development in 2005, the approach has been tested both at the project and at program level. In recent *ex-ante* evaluations it was used to clarify the intervention logic and to assess the likeliness of achieving expected results and impacts[3]. Last but not least, the approach is currently applied as part of an on-going evaluation of two Austrian OPs for the Objective "Regional Competitiveness".

[3] The procedure resembles the "prospective impact evaluation" approach recommended in the recent Barca Report and notably shares the intention to 'make explicit the expected results and the linkages between means and ends" (Barca, F. 2009, p- 179-180).

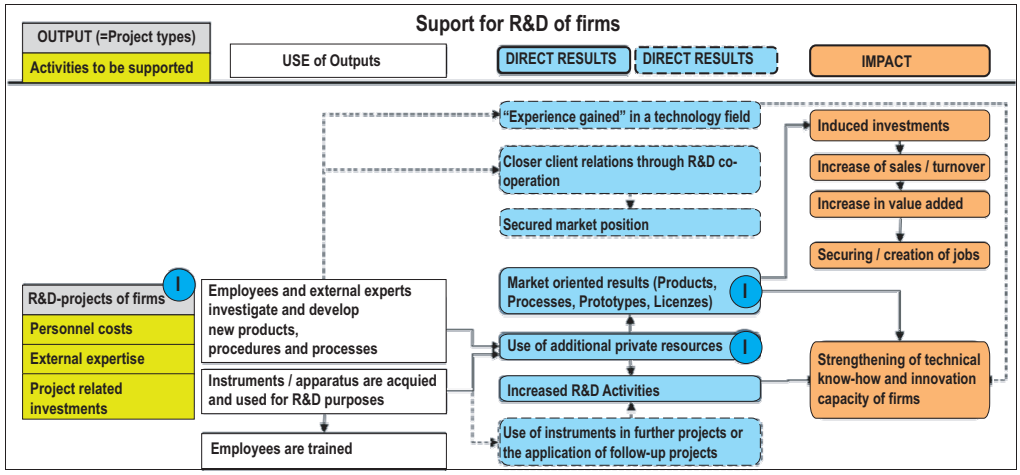
These cases will also be a test for using the approach with large quantities of data (several hundred approved projects).

The following page illustrates the application of this monitoring approach. The Figure on top shows part of the impact diagram for the Area of Intervention “R&D activities of firms” (taken from an OP for the Objective “Regional Competitiveness”). The structure of the impact diagrams follows the time sequence of intended effects: outputs are placed to the left, whereas results and impacts are located on the right hand side of the diagram. A column containing the main process assumptions (labelled “use of outputs”) is placed in between and illustrates the essential linkages between these effects. In this case, written assumptions were only made for the use of outputs, whereas the connections between expected results and impacts were merely indicated through arrows. The impact diagram was drawn up during the programming phase, in collaboration with the involved funding authorities and based on the descriptions contained in the draft OP.

This has helped to clarify the intervention logic and identify missing links. As a consequence, several “indirect results” were inserted in the diagram which were not included in the OP description but were derived from other information sources (e.g. experience, implicit goals of the authorities, findings of evaluations). Last but not least, the indicators foreseen in the OP were inserted in the diagram (mentioned in the table below), clearly showing which parts of the intervention logic can actually be captured by quantitative indicators alone.

The diagram also served to identify a set of questions for observing whether the underlying process assumptions actually take place (see table below). To integrate this observation in existing routine practices, the questions in this case were grouped so they are in line with the assessment and evaluation procedures of the responsible Implementing Body (project reports or final meetings with project owners, follow-up survey).

Process Monitoring of Impacts of an R&D support scheme (extract)



Observation questions based on process assumptions:

Reports / Final Meeting with project promoters	Follow-up Surveys
– Has the R&D project led to a market oriented result? If yes: ❍ New Product, New Service, New Process – Are investments planned for their introduction? – Are prototypes planned / have they been realised? – Will patents/licences be applied for? Has this already been initiated/achieved? – Have employees been trained to work with instruments / apparatus acquired with financial support? If yes, how many? – Is there a plan to use these instruments / apparatus in future R&D projects? – Is a follow-up R&D project foreseen as a result of the current project? Will public support be applied for? – Has the R&D project been carried out in collaboration with clients? Will this co-operation improve client relationships?	– How were the results of the R&D project implemented in economic terms? – Did the development of new products /services/ processes lead to investments? What was their volume? Has public support been applied for or has it been granted? What was the increase in turnover obtained through sales of new products / processes? – Did the implementation of new products / processes lead to the creation / securing of jobs? How many? – If a market-oriented result has not been obtained, has experience been gained in the respective technology field as a consequence of the supported R&D? Did this strengthen the technical know-how and innovation capacity of the firm? – Did the collaboration with clients during the R&D project actually improve client relationships? In which respect?

Indicators:	
Output Indicators	**Result indicators**
– Number of firms' R&D projects	– Market-oriented result: New Product, New Service, New Process – Private resources invested for R&D project

3.4. Causal Loop Diagrams

Causal Loop Diagrams are based on the concept of "feedback", a language for representing "circular" relations that was originally developed in cybernetics. System elements are linked by two types of feedback mechanisms:

(-) Negative feedback: interaction works as a limiting factor and leads to a compensation process aimed at closing the gap between a desired and actual state (stabilisation);

(+) Positive feedback: interaction leads to an increase of the previous state in the same direction (growth or decline).

There are two basic types of Causal Loop Diagrams:

- Influence diagrams: They are used to explore the strengths and weaknesses of factors from different viewpoints or to identify factors that might need particular attention in order to bring about change. They identify factors of influence (including actors), differentiate between strong and weak influence and identify potential obstacles.
- Multiple cause diagrams: They illustrate the various causes of a certain event or situation and thus explore why something has happened or why a situation is as it is. They are often used for finding out why something went wrong or why a problem keeps recurring and can be useful in identifying solutions to problems.

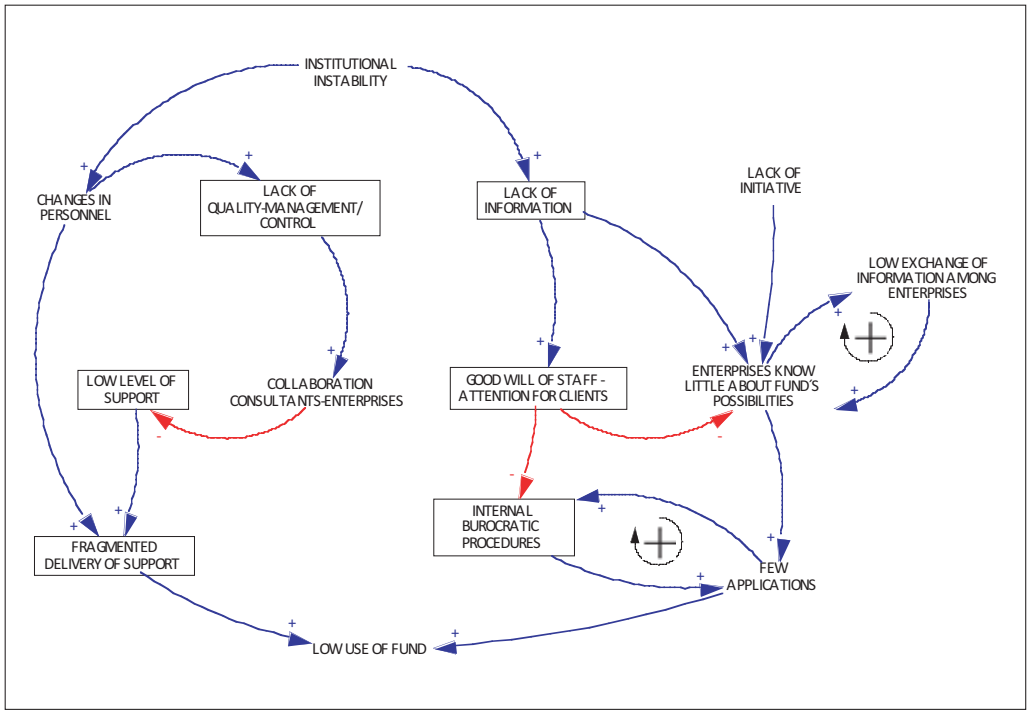
Causal Loop Diagrams are a relatively easy tool for visualizing complex relationships, they can be generated on paper, pin boards or computers. Since they facilitate communication on complex issues and can be modified rather easily, they are very appropriate for group work.

They can also be used to obtain a more comprehensive picture of reality by connecting individual views. To this end, stakeholders are first asked to provide explanations for a given situation from their individual points of view. As a next step, these different explanations for the same phenomena can be exchanged and explored in more depth. Stakeholders are asked to justify their specific boundary choices (why have they chosen certain elements or linkages), but also to question the boundary choices of others. And finally a reflection takes place as to which of these explanations are compatible with each other (i.e. complimentary or mutually reinforcing) and which of them are antagonistic. This might lead to further question their rationale in terms of power relations, back-up evidence or value judgements.

Linking different explanations can lead to multiple descriptions of the same phenomenon, being able to see and value emerging "both-and" patterns instead of "either-or" relations can create new insights and open the way for new solutions.

For example, the Causal Loop Diagram below was produced during the evaluation of a Technical Assistance Fund for private enterprises. Low use of the Fund was evident

from data previously collected and identified as one of the core issues. Then explanatory factors for this situation were collected from a series of group interviews with stakeholders (e.g. beneficiary businesses, consultants providing technical assistance, business institutions). And finally these factors were arranged by the Fund operators (who incidentally were present during the interview sessions), and were linked to form reinforcing (+) and balancing (–) feedback loops. This provided Fund operators with a much richer picture of reality and valuable insight from their client stakeholders. And it identified the leverage points, i.e. factors that can be directly influenced by the Fund operators and can have considerable influence on other elements. These factors were framed in the Causal Loop Diagram and actions were designed to change the situation in the desired direction.



In addition to modelling diverse stakeholder opinions, such diagrams can also be used to identify structures underneath observable phenomena ("symptoms"). For instance, when unintended effects of an intervention are linked to the original theory of action (modelled as feedback processes), their generative mechanisms can be revealed and indications are given on how they could be curbed or even avoided.

4. Conclusions and outlook on the application of systems thinking

The conceptual framework proposed in section 2 and the examples provided in section 3 demonstrate the possibility and practical utility of applying systems thinking

with evaluations of Structural Funds – already at this stage and within the current framework. But some requirements should be met and precautions need to be taken.

4.1. Requirements for systemic evaluations

Systems thinking is particularly suited for complex realities, as in the case of cohesion policy implemented through multi-objective and multi-actor programmes. It can be helpful in situations where evaluations aim at producing learning effects beyond the level of individuals, e.g. at the level of projects or programs, and in determining which go beyond proposing short-term solutions for particular problems and aim to contribute to future successful action in complex situations.

Systemic tools are important for achieving such "higher level" learning processes, as they enable the evaluator to make use of – and work with – stakeholder differences in a constructive manner, in order to increase mutual understanding, achieve consensus or bring forth joint solutions. They are well suited to visualize complex realities (e.g. Causal Loop Diagrams, Social Network Analysis[4]) and to better understand differences (e.g. dialogue techniques). They allow to identify relevant patterns emerging from inter-connected indicator sets (Balanced Scorecard) or to focus the entire management cycle on the achievement of impacts (Process Monitoring of Impacts).

Applying systemic thinking and tools also requires a specific stance from evaluators. They essentially have the role of external observers, who should not pretend to be objective, but provide additional points of view and specific skills for managing the process. But he/she is more than a facilitator and should intervene actively based on systemic principles, collect information and feed it back in varied (often surprising) ways to trigger reactions within the evaluated system, in order to find solutions or develop new patterns of interaction. An evaluator can (and should!) also express his/her own opinions, but in an open, non-directive manner that allow for the stakeholders to choose and decide (i.e. present options or alternatives instead of straightforward recommendations).

To apply systems thinking, evaluators need to be good observers, capable of discovering how social systems work, i.e. their behaviour patterns and regulatory mechanisms. A good way to do this is by observing their reaction on past or present interventions – including one's own during the evaluation! Systemic evaluators use specific skills for organising data or debates, but need not necessarily be experts in systemic tools (nor study extensive literature on their use). However, they should be aware of their possibilities and know where to find suitable expertise. Perhaps most importantly, they should have a general understanding of systems thinking and be capable of working in line with the stance outlined above.

---

[4] See the following article by Ploszaj A. in this volume

Another important requirement for systemic evaluations is flexibility in implementation. They are designed as iterative processes, consisting of successive reflective loops. Thus, it is important for clients that they can cope with recursive designs, where only a basic outline can be defined at the start (including the available budget, a timeframe with milestones for the delivery of services or interim findings), but where the process should remain sufficiently open to respond to new findings, requirements or issues. And evaluators must maintain an overview of resources and time requirements and be able to steer the assignment within this general framework despite changing demands.

But even though Structural Fund programs are complex realities, per se, does not mean that a systems approach is appropriate for evaluating each aspect or situation. This will essentially depend on the nature of an intervention, and for the purpose of structuring and clarifying, Stacey's three-part typology can be applied (Stacey, R. 1992):

- Simple: repeatable, can be specified and standardized, with known causal strands that can be clearly observed;
- Complicated: several activities/actors which are part of a larger package, often located at different levels and operating contexts, multiple yet knowable causal strands;
- Complex: multiple, yet dynamic activities/actors which are adaptive and responsive to emergent needs or opportunities, recursive causality that is unpredictable in advance.

The appearance (or rather the combination) of these characteristics in a given situation will determine the utility of a systems approach. Whereas it will hardly be needed for evaluating simple situations, it can be useful in complicated situations (perhaps complementary with logic models) and indispensable in making sense of complex situations. Thus, systems approaches are not a panacea for evaluating complex realities, but should only be applied when suitable. Situational responsiveness – matching the designs of an evaluation to the purpose as well as the needs and constraints of a specific case – is particularly relevant for impact evaluations (Rogers, P. 2009). This involves an appropriate choice and mix of methods, which contrasts the advocacy recently displayed by DG REGIO to "rigorous impact evaluation", and the use of control groups or RCTs as "gold standard" for impact evaluations.

4.2. Application of systems thinking for program governance

But the relevance of systems thinking for Structural Funds goes beyond evaluation and encompasses the entire program cycle – notably the design, implementation and monitoring of development strategies.

Most of the strategic objectives of Cohesion Policy or the Lisbon Agenda (e.g. creating jobs, increasing economic competitiveness) are far beyond the immediate

reach of the public decision makers, i.e. beyond traditional "government". They require the interaction of economic, social, cultural and physical resources within a given territory. Seen from a systemic perspective, success depends on the quality of inter-relationships between key actors having access to – or being responsible for – these resources. Due to the relative autonomy of these actors, strategies will only work if they understand and accept at least the main objectives and procedures of the strategy. Thus, the coordination of actions cannot be based on traditional command and control but requires softer and more differentiated forms of intervention ("governance"), which helps creating joint understanding and acceptance.

However, in Structural Funds a traditional "command and control" governance model has been applied so far. The deficiencies of this model are widely acknowledged and a departure has been announced (see for instance the Commission's "White Paper on European Governance"). But although the Regulations for the program period 2007–2013 contain some elements in this new direction, the strategic follow-up foreseen by the Guidelines is based on the old mechanistic model (Strategic Reports based on a set of quantified indicators). But monitoring complex strategies requires more than indicators and needs to depart from the assumption of a "linear" progression of effects (outputs – results – impacts), which happens irrespective of the actors involved or the framework conditions. And reporting on socio-economic effects without clarifying the links to cohesion policy cannot inform on the validity of the strategies pursued. Moreover, this might reinforce tendencies to claim observable effects, regardless of whether cohesion policy has actually contributed to their achievement. Last but not least, the current approach builds excessively on written reports and lacks interaction or communication among program actors – in short: governance.

The basis for an improved strategic governance of Cohesion Policy is a clear understanding of impact creating processes, i.e. the "theory of action". A strategy needs to identify the crucial processes capable of linking immediate public action (e.g. funding of projects) with the ultimate objectives, starting by the desirable (i.e. to be funded) actions of public and private actors, their expected links to the ultimate objectives, and the (external) framework conditions required or conducive for achieving these effects.

This can – and should – first be done at the planning stage, but given the complex nature of these strategies plans can only be a hypothesis and need to be reviewed and adapted in light of experience gained. Therefore a shift in attention from planning to reflected management of implementation is needed, spending far less time and resources for preparing programs and much more on monitoring and evaluation during implementation.

A specific challenge for this review is the generation and exchange of relevant information, as much of it is decentralized, located with a wide range of actors, which include implementing partners or project owners. And often this information is

tacit knowledge, i.e. based on experience and observation but not formalized or documented. To tap this knowledge, the sources (individuals, organizations) must be addressed in an adequate manner to make their knowledge explicit, which is best achieved through dialogue and interaction. Communication plays a crucial role for strategy implementation, because this multi-actor, multi-level process requires mutual understanding in a systemic sense, i.e. "mental maps", perspectives and boundary choices of involved actors. But creating a common language and joint understanding takes time, cannot be proclaimed or created artificially, but will only emerge gradually during an ongoing communication process, which requires appropriate formats for joint reflection and learning.

Therefore, communication cannot be limited to strategy documents or reports, but has to be established as an ongoing process of reflection parallel to the implementation process. Such a strategic governance process has been implemented in Austria since 2007 for the NSRF (strat.at) and nine Objective 2 OPs. It is steered by the existing network of Managing Authorities, which defines an annual work plan based on thematic priorities. A series of encounters are organized each year including more informal formats that allow for an open exchange of experience, critical reflection and creative generation of new ideas. This governance process should also incorporate the findings of evaluations done within the various OPs and will contribute to the elaboration of Strategic Reports required by the European Commission.

But a more effective governance process needs to be established at the EU level as well. The recent Barca Report (Barca F. 2000) proposes a reform of governance for cohesion policy in order to achieve greater coherence of the recommended place-based approach. The ten "pillars" of this proposal take into account many of the aspects mentioned above. They include a new strategic framework for cohesion policy, a new contractual relationship and an enhanced strategic dialogue between the Commission and Member States, a move towards prospective impact evaluation as well as implementation and reporting aimed at results.

References:

Barca, F. 2009, 'An Agenda for a reformed Cohesion Policy', A place-based approach to meeting EU challenges and expectations, European Commission, DG REGIO, Brussels

Williams, B & Imam, I (eds.) 2006, '*Systems Concepts in Evaluation*', An expert anthology, AEA Monograph series, Edge Press.

Hummelbrunner, R & Reynolds, M 2008, *Systems Concepts in Evaluation*, Workbook for a Pre-Conference Professional Development Session, European Evaluation Society.

Hummelbrunner, R, Huber, W & Arbter, R 2005, '*Process Monitoring of Impacts – Towards a new approach to monitor the implementation of Structural Fund Programs*', Bundeskanzleramt Abteilung IV/4, Wien.

Rogers, P. 2009 '*Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation*', in 'Designing impact evaluations', Working Paper 4, International Initiative for Impact Evaluation.

Stacey, R. 1992, '*Managing the Unknowable*', San Francisco, Jossey-Bass.

NETWORKS IN EVALUATION

Adam Płoszaj

Introduction

The conception of a network is one of today's leitmotifs. On one hand, this is most likely the result of a specific trend dominant in science and media reports. On the other hand, it is difficult to negate the existence of such network phenomena as the Internet or social networks, both virtual (e.g. Facebook) and those existing in physical space. According to some thinkers such networks become so ubiquitous that they begin to provide the basis for explaining the modern world. This gave origin to the notion of the Network Society, popularized by Manuel Castells in his famous work "The Rise of the Network Society" (1996) . As a consequence, the increasing presence of networks in the theory and practice of evaluation comes as no surprise. However, despite the growing interest in them, networks in evaluation applications are still perceived as something new and promising. Broadly defined network analysis is still rarely used in evaluation practice. There are also very few studies discussing practical applications of networks in evaluation. This relative lack of practice-oriented studies of network analysis in evaluation is the main reason for this chapter.

In this chapter, I discuss elements of network analysis most important for the evaluation of development policy programmes and projects. The chapter begins with a discussion of the main concepts in network analysis, accompanied with a presentation of the sources in which particular issues are discussed in detail. The main part of the chapter presents cases of network analysis applied in evaluation research. They have been taken from literature on the subject and, predominantly, from my own professional experience, as I have performed some network analyses in the last few years as part of evaluation studies conducted in Poland (and scientific research of a similar nature). Moreover, network visualisations, presented alongside the discussed cases, are important for this study. They alone can provide inspiration for readers, who one day may confront the difficult task of visualising a complex network. The chapter concludes with a short summary of the strengths and weaknesses of applying network analysis to evaluation.

Networks in evaluation – the main concepts

Origins of the concept

Only recently has network analysis been used in evaluation, which is exemplified by the fact that it was only in 1998 that it first appeared as a methodological ap-

proach in the programme of a conference organized by the American Evaluation Association (Durland, Fredericks, 2005, p. 7). However, network analysis as such has a long tradition, rich literature and a number of impressive applications (see Freeman, 2004). A major part of network analysis research is developed within social sciences in the form of the Social Networks Analysis (SNA). However, it is not possible to imagine today's network analysis without the contribution of exact sciences, mostly mathematics and graph theory (see e.g. Wasserman, Faust, 2007, 10-17). One leading network researcher, Steven Strogatz, vividly describes network analysis as: "concerned with the relationships between individuals, the patterns of interactions. The precise nature of individuals is downplayed, or even suppressed, in hopes of uncovering deeper laws. A network theorist will look at any system of interlinked components and see an abstract pattern of dots connected by lines. It's that pattern that matters, the architecture of relationships, not the identities of the dots themselves. Viewed from these lofty heights, many networks, seemingly unrelated, begin to look the same" (Strogatz, 2003, p. 231–232; see also Bendyk, 2004 p. 257).

A basic, yet time consuming, introduction to social networks analysis is provided in an eight-hundred-page work by Wasserman and Faust (2007). The on-line manual by Hanneman and Riddle, "Introduction to social network methods" (www.faculty.ucr.edu/~hanneman/nettext), can also be recommended. Philip Ball (2004) in a very straightforward manner makes the reader acquainted with the latest results in broadly-defined network research. An interesting position related to network analysis in evaluation is the special issue of the "New Directions for Evaluation" entitled *Social Network Analysis in Program Evaluation*, edited by Durland and Fredericks (2005). An article similar in nature is "Network analysis: methods and application in evaluation" by Dominik Batorski (2008). An interesting overview of the integrated approach to network evaluation, called "Network Evaluation from the Everyday Life Perspective" (NEELP) is presented by Finnish researcher Liisa Hopelli (2009).

Background of the network concept

Every network consists of nodes and the relationships between them. Nodes may be people, organizations, their organizational units, events, projects, etc. Nodes have their attributes or characteristics (for people they may include age, education level, sex, etc.). The relationships may take the form of an exchange of information, cooperation, participation in the same projects, friendship, but also mutual competition, among other things. Flows pass between the nodes along the relationships lines, so to speak. These flows may include flows of funds, information, employees, etc.. Node, link and flow (see e.g. Barney 2008) are the basic concepts in network analysis. It should be noted, however, that in the development of this research approach a number of various specialized concepts have been coined which are not going to be discussed in detail in this chapter, as this would make it too long and because there is ample introductory literature available on the subject (see above).

Moreover, in evaluation practice the most effective analysis tools seem to be the most basic ones. One should bear in mind that evaluation has a very utilitarian aim, the results of which should be clear and easy for practical implementation. Accordingly, a particularly interesting feature of network analysis is the opportunity for creating visualisations; for example, presenting connections between employees in an organization in such a way that at first glance and without previous preparation one can understand the structure of the network: who is in its centre, who has many connections and who remains on the margin with just a few relationships with other nodes, what groups and cliques emerge, etc. The importance of simple methods is also emphasised by researchers dealing with networks analyses for enterprises. Based on many years of their research experience, Cross and Parker (2004) show the power and practical importance that a very simple graph may have.

Network data – collection, analysis and presentation

The most commonly used data collection method in network analysis evaluation practice is the questionnaire. A wide range of different tools are in use, among which the most important are: face-to-face questionnaires, telephone interviewing and more frequently as Computer Assisted Telephone Interviewing (CATI), Computer Assisted Web Interviewing (CAWI) and mail questionnaires. Network data are also collected by using other tools, such as: face-to-face interviews, observations, archival records and databases (for more details, see examples below as well as Wasserman, Faust, 2007).

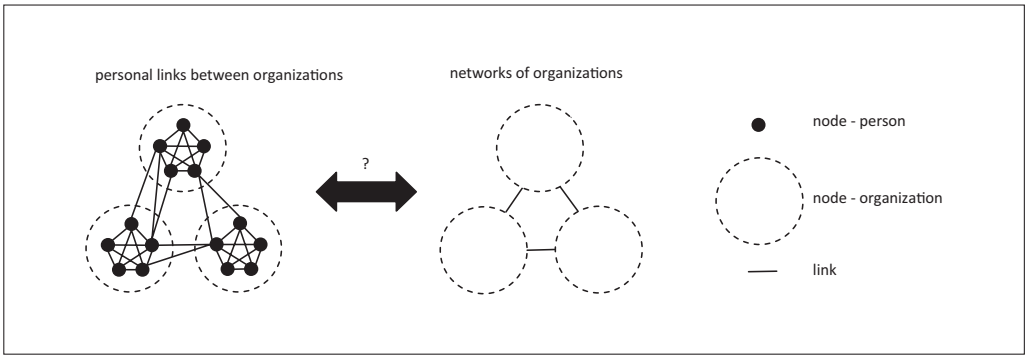
Network data needs special treatment in data processing. Fortunately, there is a wide range of network analysis and visualisation software available, both proprietary and freeware: Commetrix, EgoNet, InFlow, NetDraw, NetMiner, Pajek, and UCINET, just to name a few.

Network of people vs. network of organizations

From the point of view of evaluation practice it is important to pay attention to the multidimensional character of networks existing in organizations and between organizations. First of all, we can discuss networks of people, networks of organizations (see Fig. 1) or a network of organizational units (departments, divisions, institutes, etc.). As organizations consist of the people creating them, a network of organizations will be, *de facto*, a network of people affiliated with them. Transition from the level of relationships between people in various organizations to the level of relationships between organizations is not always simple and straightforward. It may be particularly troublesome in the case of research based on interviews (questionnaire-based or individual in-depth interviews). In principle, the research should include all the people from all institutions. In practice, however, this may be very difficult – due to organizational factors, but mostly to time and financial constraints. Consequently, analyses usually include only the most important actors from the con-

cerned organizations. They are either asked to describe the relationships between their organization and other organizations, or to indicate their individual relationships with members of other organizations, later assumed as a basis for conclusions on the relations between the organizations. This approach, however, raises some concerns. Firstly, selecting one representative of an organization, for example, as the source of information on its relationships with other organizations, one has to accept that the person does not have to be aware of all the relationships existing between the analyzed organizations. Moreover, the answers may be influenced by a given person's greater of lesser inclination to cooperate, as well as personal experiences in this respect. However, the discussed method of collecting data on organizational networks is quite commonly used (see e.g. Galaskiewicz 1985; Fredericks, 2005), mostly due to the organizational and cost aspects of the research.

Fig. 1. Organizational network on the level of organizations and on the level of interpersonal relationships



Source: prepared by the author.

The functioning of an organizational network (or other networks) may be analyzed from various perspectives. There are usually two complementary perspectives on organizational network analysis: the perspective of particular networking organizations and the perspective of the whole network (Provan, Fish, Sydow, 2007). Those perspectives are sometimes also referred to as the micro- and macro- level approach (Wasserman, Galaskiewicz, 1994), or the egocentric network perspective and the whole network perspective (Kilduff, Tsai 2003). Depending on the object and aim of a given evaluation one, or both, of the perspectives should be applied as appropriate, (see Provan, Milward, 2001). The rationale depends mostly on the kind of effects under consideration, i.e. whether we concentrate on the effects for particular organizations in the network or the real-life effects brought by the network as a whole. It is worth noting that the efficiency of an action on one of these planes does not necessarily translate into efficiency on another plane. For example, a lack of expected effects of state intervention on the whole network level does not preclude the possibility of positive effects for particular networking organizations. Moreover, particular actors may have individual, perhaps conflicting goals:

> *The main problem is that (…) there are a lot of actors involved, each with their own differing perceptions, goals, and strategies. It is not a clear case of which goal the evaluator should take as a starting point for evaluation, especially not if more than one public authority is involved. But even if there is only one public actor involved, the question still remains of whether the goals of this actor have to be the sole evaluation criteria. If one looks at interactions in policy making and management from a multi-actor perspective, it seems logical to look at evaluation from a multi-actor perspective too. It is not very likely, however, that the different actors will have a collectively formulated goal at the beginning of policy interactions that can serve as a keystone for evaluation (Klijn 2005, p. 273).*

One-mode and two-mode networks

When thinking about a network we usually imagine a set of interconnected nodes of the same type, such as people, organizations, computers, etc.. In the case of such a network, called one-mode networks, we are dealing with connections between just one category of nodes; that is, connections between people and people, or organizations and organizations. An example of such a network is presented in the table and graph below (Fig. 2). The nodes may, for example, be cooperating scientists. In such a case A, B and C mutually cooperate with one another, while C additionally cooperates with D (who cooperates only with C).

Another type of network is a two-mode network, also known as affiliation network. In such case the nodes belong to two different categories, and their relationships are usually affiliation relationships. They can, for example, consist of participation in the same projects or membership in the same organizations. In relation to the example of cooperating scientists – the relation here may consist of participation in the same projects (Fig. 3). Let's assume that A, B, C and D participated in projects X, Y and Z. A participated in X and Y; B participated in X; C participated in all the three projects; while D participated only in project Z.

The large potential of network analysis lays in the fact that a two-mode network may be quite simply (using the appropriate function in SNA software) transformed into a one-mode network. Accordingly, from the network in Figure 3 we may obtain the network presented in Figure 2. Consequently, having the information that, for example, persons A, B, C and D participated in projects X, Y and Z (Fig. 3), and assuming that participation in the project requires cooperation, we may construct a cooperation network between the analyzed individuals. This technical procedure involving transformation of an affiliation network into a one-mode network is of significant practical importance, as the data on network characteristics are often much easier to obtain than the data on relationships existing directly between the elements of interest. Among other things, this method allows for studying the re-

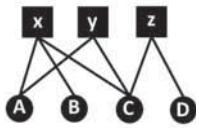
lationships between scientific institutions. Having a database of projects and their participating institutions (i.e. affiliation network) we can easily learn which institutions mutually cooperate and how often (i.e. in how many joint projects), which of the institutions cooperate with a significant number of other institutions and which with just one or two other institutions .

Fig. 2. One-mode network

Nodes	A	B	C	D
A		1	1	
B	1		1	
C	1	1		1
D			1	

Fig. 3. Two-mode network

Nodes	x	y	z
A	1	1	
B	1		
C	1	1	1
D			1



Source: prepared by the author. Source: prepared by the author.

Evaluation of networks and evaluation using network analysis methods

Network evaluation does not have to use network analysis methods. On the contrary, it is entirely possible to use methods other than network analysis in analysing a network, and such approach is quite common (see e.g. Rank, Williams, 1999; Ahrweiler, de Jong, Windrum, 2002; Szałaj, Ledzion, 2008). This does not mean, however, that network analysis may be replaced with other methods (e.g. standard quantitative research). Nor does this mean that network research not using network analysis lacks methodological foundations. Such an approach will simply supply somewhat different knowledge about the object of the study, and will not allow for characterizing the relationships in a given network in a systematic manner. That is why in such cases the application of network analysis turns out to be particularly beneficial.

It is important that the subject of evaluation using network analysis methods does not necessarily have to be a network. It can be any phenomenon conceptualized as a network. For example, relationships in a system of strategic goals or operational programmes (see, for example, Davies, 2005, 2007).

Evaluating networks – real-life examples

Network analysis can have a number of powerful applications in evaluation practice. In this chapter we will discuss some of the most important and impressive. The discussion begins with the cases of the programmes implementations networks, both at the level of organizations and organizational units. Then, network analysis is presented as a tool for the evaluation of cooperation in partnership projects. Subsequently, the case of identifying knowledge resources with network analysis is introduced. Another topic is the evaluation of formal institutional networks. The final case describes using Geographic Information Systems for network analysis.

Evaluation of programmes implementation networks

The implementation of programmes is usually a complicated process involving many institutions and individuals. Effective programme implementation requires proper operation of the whole system, efficient information flow and effective cooperation between the actors. A consecutive sequence, which is standard for implementation tasks, necessitates their performance in due time, so as not to create bottlenecks in the implementation system. A problem may also result from lower efficiency of particular elements in the implementation system and the resulting "bottlenecks", slowing down the whole process. Moreover, the effectiveness is also influenced by the nature of the cooperation and information flow between the actors concerned. Network analysis allows for a systematic study of such issues.

We can analyze the implementation (or management) network on many levels. For example, on the macro-level we can analyze all the institutions in a given country dealing with the cohesion policy implementation; in the case of Poland in the current 2007–2013 programming period, this is over 145 organizations. One can also analyze the implementation of particular operational programmes or their parts. In the micro-level one can study a network of organizational units within a given organization or even cooperation between individuals involved in a given implementation/management process, etc. Moreover, one can study various aspects of the functioning of networks, such as information flow, cooperation on joint problem solving, mutual learning and exchange of good practice, as well as diffusion of innovations (see e.g. Rogers 2003).

Evaluation of programmes implementation networks – level of organizations

An interesting example of programme implementation network analysis is provided by a study conducted by Swianiewicz *et al*. (2008). Even though it was not aimed as an evaluation, because of its character, we can treat it as an element of an evaluation study. The subject of the study was the informal network of institutions involved in the implementation of the Integrated Regional Development Programme (IRDP) in two Polish regions: Małopolskie and Dolnośląskie (Lower Silesia). In the case of Małopolskie there were 20 institutions covered, and in the case of Dolnośląskie – 52. The information on connections between them came from representatives of the institutions and was collected during interviews. The questions asked to particular actors pertained to the frequency of contact in general, and of individual contact (i.e. other than official) between the employees of a given institution with the employees of other institutions dealing with implementation of IRDP in the region. Thus, the data collected in this way pertained to the declared relationships between institutions from the perspective of representatives of the analysed institutions. The results show that the networks in both regions are similar. In both regions the central place in the implementation networks was occupied by the bodies formally

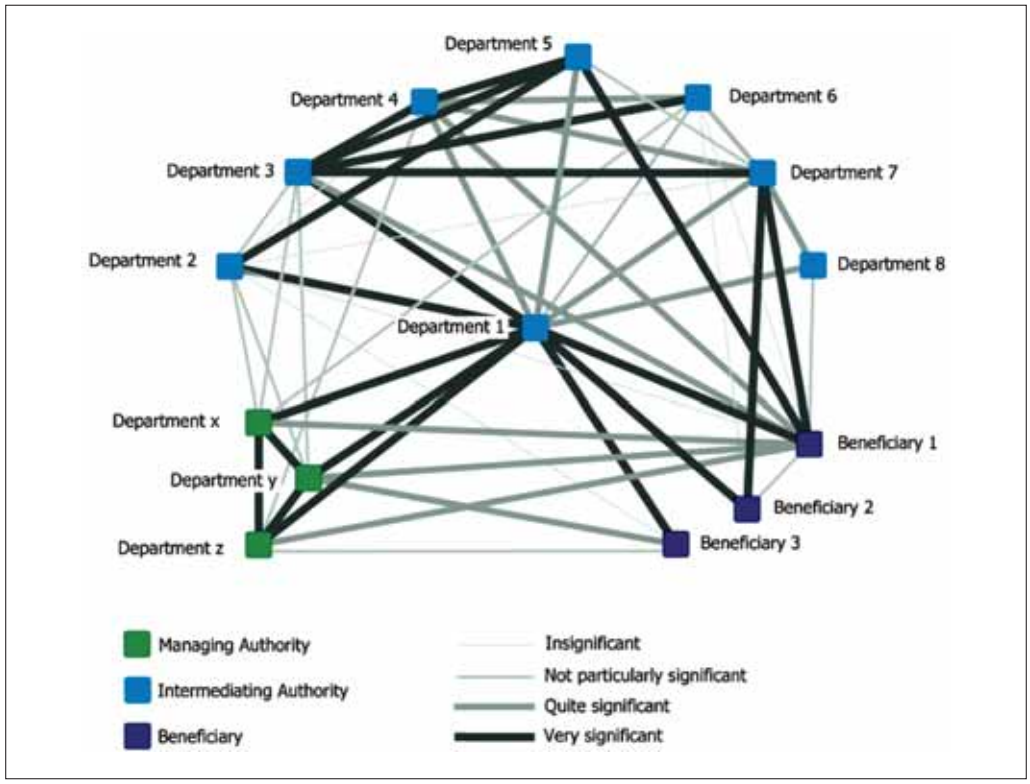
responsible for IRDP implementation: The Marshall Office and the Voivodeship (Province) Office. However, the authors of the study pointed out a couple of important differences. First, they pointed out the fact that the respondents from Lower Silesia were more willing to talk about their contacts than the respondents from Małopolska, which translates into a less dense network in Małopolska than in Lower Silesia. Moreover, the differences in the role of the Province Capital in the IRDP implementation network are also clearly visible. The Wrocław City Hall has a central place in the Lower Silesian institutional network, while the Kraków City Hall has a rather peripheral position in its province. In Małopolska non-governmental organizations (mostly NSZZ, the Solidarity trade union) were relatively important, while in Lower Silesia they were far less important. Moreover, the authors also mention some other features specific for the IRDP implementation networks in the studied regions. How can the aforementioned analyses be translated into the practice of evaluation? One potentially very productive application is the use of implementation networks' characteristics as variables explaining the course and effects of IRDP implementation in particular regions. The relationships between the institutions dealing with implementation may have considerable impact on, for example, the level and pace of implementation, as well as the effectiveness of investments. If during evaluation it turns out that the character of the links in the implementation network significantly influences the course of intervention, this may constitute a basis for making recommendations concerning the optimum structure of cooperation between the implementing bodies.

Evaluation of programmes implementation networks – level of organizational units

The functioning of the programme implementation networks may also be analyzed on a level lower than that of particular organizations; that is, the level of particular organizational units in the analyzed institutions dealing with implementation. Such an approach was chosen in the study entitled "Ewaluacja pierwszego etapu wdrażania Działania 5.1 Programu Operacyjnego Kapitał Ludzki" ["Evaluation of the first stage of implementation of Measure 5.1 of the Operational Programme – Human Capital"]. Communication and cooperation were subject to analysis between 14 organizational units in a couple of organizations: The Managing Authority, the Intermediate Body and three beneficiaries. The data for analysis were collected from interviews with representatives of particular organizational units. Quite unexpectedly it turned out that cooperation in tasks connected with implementation is more complex than would seem from the organizational chart representing the Measure's implementation structure (see Fig. 4). This applied to the dense cooperation network between the organizational units in the Intermediate Body, numerous units of the Intermediate Body and the Managing Authority, and most of all to the direct contacts between the Beneficiaries (see Fig. 4. beneficiaries 1 and 3) and the Managing Authority. Network analysis showed the actual scope of cooperation,

which in this case turned out to be satisfactory, except for one case. Namely, the analysis revealed faults in cooperation between one of the beneficiaries and one of the organizational units in the Intermediate Body. In this case, network analysis allowed for the problem to be spotted precisely, and the resulting recommendations to include proposed suitable corrective measures.

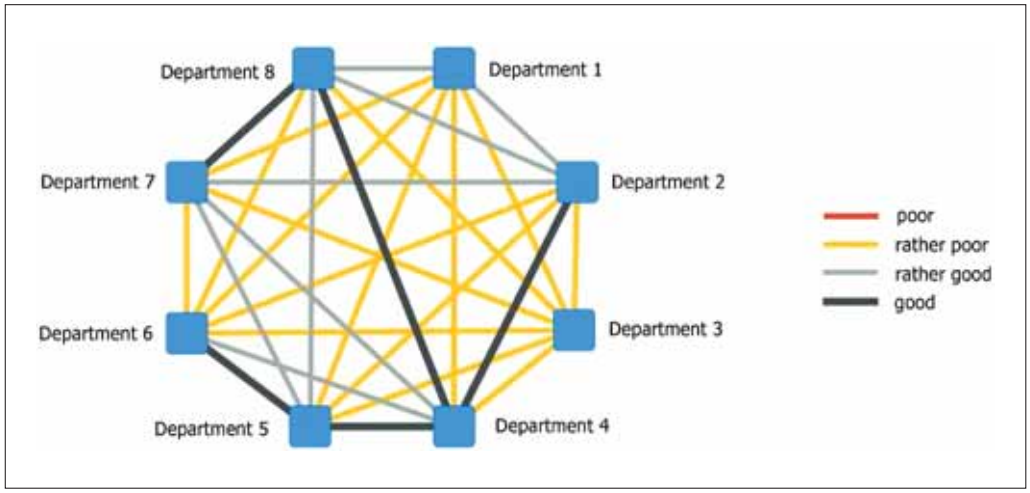
Fig. 4. Cooperation network – significance of relations



Source: prepared by the author.

A similar analysis on the level of organizational unit held for one body implementing the Operational Programme – Human Capital in one Polish region determined quite the opposite picture. In this case it turned out that the information flow between particular entities is generally poor (see Fig. 5). The network analysis showed, rather unsurprisingly for people working in the organization (even though the exact scale of the phenomenon might have been unknown to them), that no methods for quick and effective communication or information exchange had been worked out. In this case, problems with information flow quite significantly affected the assessment of cooperation, which had often been negative. The network analysis conducted at a relatively early stage of implementation allowed for taking proper corrective measures in order to minimize the situation's negative impact on the implementation of the programme.

Fig. 5. Cooperation network in the Voivodeship Labour Office – information flow



Source: prepared by the author.

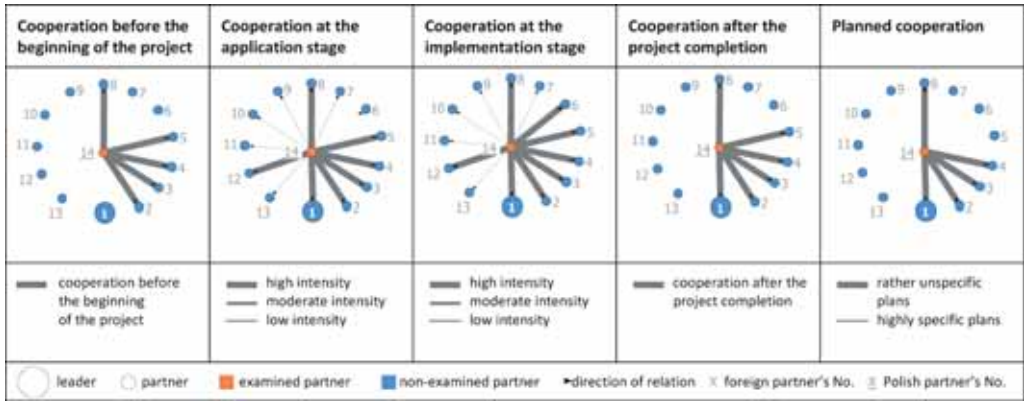
Evaluation of cooperation in partnership projects

A broad area of network analysis application is the evaluation of so-called soft projects, which are aimed at constructing partnerships between various institutions. By collecting information on the relationships between particular institutions in particular points of the project (programme) implementation, one can precisely characterize not only the state of the network, but also its dynamics. As a result, using network analysis one can quite see precisely whether the partnerships are lasting. In this case network analysis acts as a tool for results-oriented evaluation.

An example of such a study is provided by research on "Ewaluacja sieci partnerów w projektach INTERREG IIIB CADSES" ["Evaluation of partnership networks in INTERREG IIIB CADSES"] (Ego s.c. 2008). A characteristic feature of territorial cooperation programmes is the fact that they are usually implemented by a number of institutions from various regions and countries. It is supposed to lead to lasting cooperation and an exchange of experience and good practice. However, the actual cooperation may take various courses. Information on cooperation between partners in the projects was collected from representatives of Polish institutions (partners in CADSES projects) in a telephone interview. They were asked about the intensity of the cooperation between the analyzed institution and all the institutions involved in a given project at various stages of its implementation (formulating the conception and implementation of the project) and before it (whether they had cooperated earlier), as well as after its completion (planned and ongoing cooperation). This approach allowed for studying the relationships between project partners (and their influence on its implementation) as well as the assessment of how lasting the cooperation turned out to be after the project's completion. The analysis allowed for differentiating between projects in respect to the greater or lesser intensity of

cooperation that they involved at particular stages and showed the extent to which the implementation of the programme contributed to establishing lasting cooperation between the projects' partners. Most importantly, the study demonstrated that participation in a project does not necessarily translate into mutual lasting cooperation (see Fig. 6).

Fig. 6. Evolution of cooperation within the partnership project



Source: prepared by the author.

Network analysis as a tool for identifying knowledge resources

Network analysis may be successfully applied in mapping competences, experience and knowledge of organization(s). Such an approach may be useful when, for example, analyzing cluster initiatives, evaluating regional research and development potential, or assessing the competence of the evaluated organizations' employees. This aspect seems to be very important, as knowledge is seen as a major capital in the so-called knowledge-based economy (see e.g. OECD, 1996).

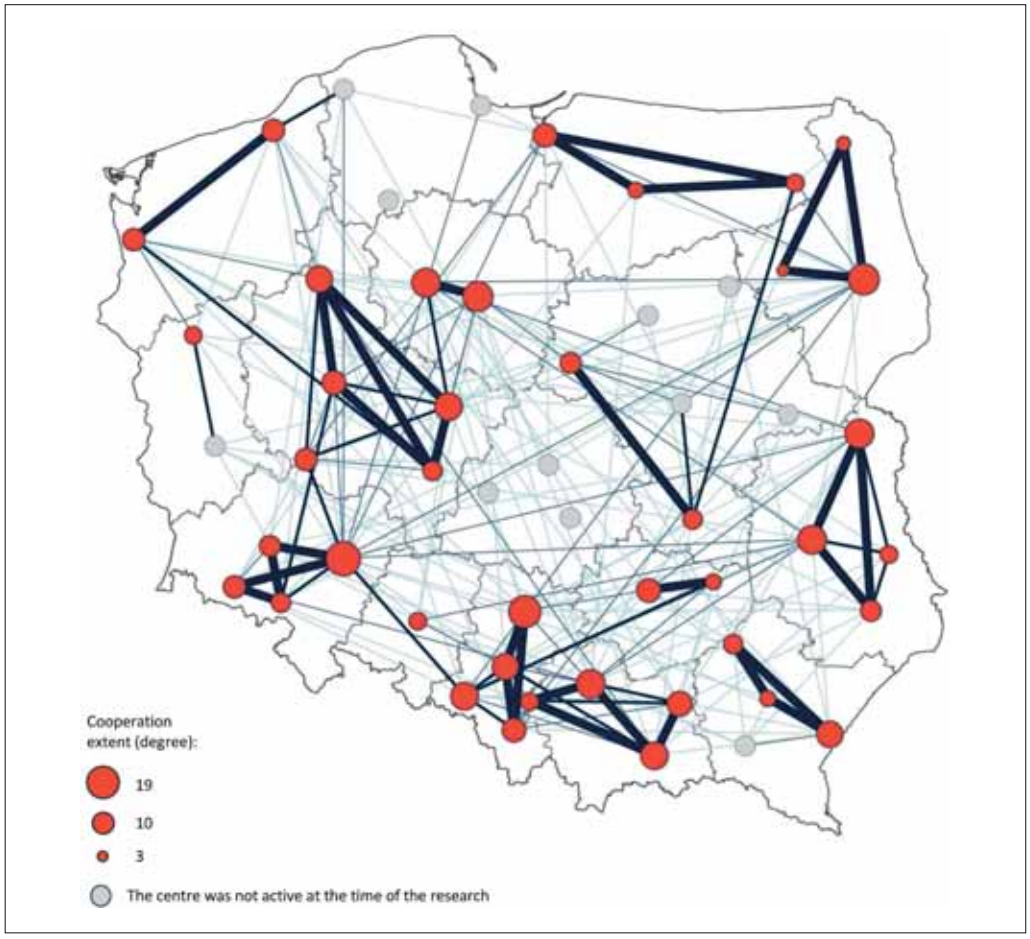
A study of this type was conducted in 2004 at the Idaho National Laboratory in the U.S.A. (Birk, 2005). The analysis was necessitated by the planned integration of existing separate research projects connected with hydrogen fuels into one large "Hydrogen Initiative Program". The programme was supposed to become the basis for the laboratory's new business strategy. In order to efficiently launch the new initiative, the assessment of the current laboratory's competences was made. This was particularly important due to the large number of researchers working in various buildings and locations. All the staff (38 people) to be employed in the new initiative was asked to indicate the people that they considered to be experts in each of 47 specified categories of knowledge connected with hydrogen fuels. The subjects could indicate both people from within and outside of the laboratory. The results of the questionnaire were analyzed using network analysis methods and resulted in the creation of 47 graphs presenting experts in particular areas. The results of the analysis were consulted with a focus group comprising some of the laboratory's researchers, who provided a detailed and reliable interpretation of particular

graphs. The results of the study singled out a group of researchers most frequently indicated as experts, i.e. key for the success of the initiative. They also revealed that in some areas the laboratory had more than one, while in others just one expert. Such information may have extreme importance for programme management. If, for example, one expert is expected to retire soon, measures aimed at recruiting or training a person to take his place when he retires should be taken beforehand.

Evaluation of formal institutional networks

Network analysis finds its somewhat obvious application in the evaluation of projects with names as well as characteristics indicting their network character, consisting of the creation of a system of networking institutions. An example of such an initiative may be the European Union project of Innovation Relay Centers, currently transformed into the Enterprise Europe Network. An interesting project of this type is the Regional ESF Centres network initiative, aimed at providing comprehensive support for beneficiaries and potential beneficiaries of the European Social Fund. The network consists of a few dozen centres located throughout the country. The creation of the network was accompanied with the assumption that particular centres will cooperate by exchanging so-called good practice, i.e. cooperating in finding beneficiaries for the partnership projects. Evaluation of this project was aimed at checking whether such cooperation actually takes place. The basic level of analysis in this case consisted of individual regional centres (there were 40 such centres at the time of the study); however, in order to make the results more credible, the study included the whole professional staff (241 people) of the centres in question. The study was held using telephone questionnaires. The results showed, among other things, that the cooperation between centres is quite frequent, although usually pertains to the centres located in the same region and relatively rarely involves centres from various regions (see Fig. 7). Therefore, the simple recommendation followed to expand the network by mechanisms stimulating interregional cooperation.

Fig. 7. Cooperation within the Regional ESF Centres network



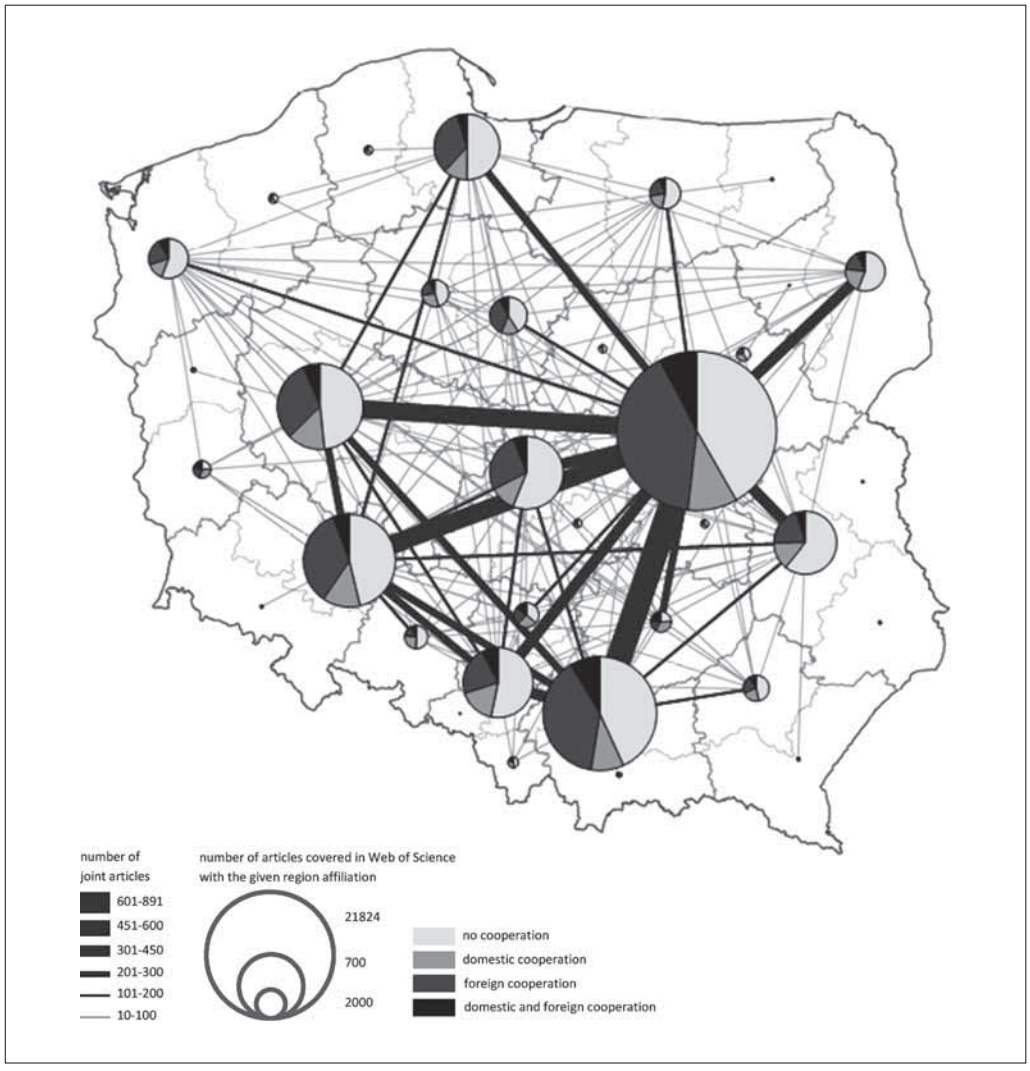
Source: prepared by the author.

Networks evaluation and Geographic Information Systems (GIS)

Many networks have a significant spatial aspect. This is exemplified by the previously discussed Regional ESF Centres network. The use of GIS tools in analysis and visualisation of the network is potentially very important. Most significantly, it provides an attractive presentation of spatial network data. Moreover, for a potential user the map presentation may be more familiar and readable than an abstract graph (not to mention a table or data matrix). An example of visualisation of cooperation network using typical elements of thematic map is presented below (Fig. 8) and shows the cooperation of research centres in Poland measured by joint publications. Moreover, it seems that GIS may provide good inspiration for visualising networks in an effective manner. Cartographers have for ages worked out various concepts and principles allowing for clear presentation of numerous pieces of information, even the most complicated ones. In practice, however, network visualisation frequently has to deal with the problem of making the graph readable. Consequently, it seems

that the GIS experiences may also be successfully applied to the visualisation of networks with no spatial characteristics, per se.

Fig. 8. Network of scientific cooperation



Source: Olechnicka, Płoszaj 2008.

Conclusions

The examples discussed above show how network analysis may be applied to the evaluation of state intervention. As already mentioned, it is usually used in combination with other research methods, which conforms to a more general postulate of triangulation of research methods in order to make the analyses more credible. In the case of regional development programmes, network analysis turns out to be useful in all situations where an important element of the programme (project)

implementation or its effects, are formal or informal relations between institutions or people.

Nevertheless, one should keep in mind that network analysis, as with every method, has its own strengths as well as limitations. To sum up the lessons learned from the examples discussed above, we will discuss the three main benefits of network analysis in evaluation and then three major challenges.

Network analysis' main benefit is the fact that it allows for capturing even very complicated and multifaceted relationships between numerous elements in an accurate and quantified manner. This increases the objectiveness of the analyses, and consequently raises the objective results of the studies and recommendations. It is also important that network analysis already has a solid theoretical basis, elaborated and well-grounded research methods and procedures (see the beginning of the chapter).

It is also invaluable that "network research can integrate qualitative, quantitative and graphical data, allowing more thorough and in-depth analysis" (Kilduff, Tsai 2003; p. 19). Consequently, "the network approach enables the analyst to retain the richness of the data rather than having to sacrifice richness for statistical power" (Kilduff, Tsai 2003; p. 25). Network analysis thus constitutes a perfect environment for using mixed-method methodology, particularly suitable for research on inherently complex state intervention programmes and projects.

Moreover, various methods of visualising networks (using suitable software, both specialized network visualisation software and standard graphics applications and GIS software) allow for interesting and revealing presentations of the research results, which can considerably increase their usefulness for final users. A visualisation of network connections may also be very useful in the analytical stage of research, providing a method for selecting institutions for deeper analysis (one can expect that the institution most connected with others will also be the most "aware" of what is going on in the whole network).

A very important difficulty in network research results from the fact that the analyses usually have to include the whole studied population. Network analysis is very susceptible to lacking data and it is practically impossible to extrapolate the results from a sample to the whole population. The necessity for the analyses to take into account all elements constituting a given network obviously results in many problems to be dealt with, especially in the case of questionnaire-based surveys, where obtaining a 100% response rate is practically impossible. Therefore, in gathering network data, interviews usually provide a better tool than questionnaires. Another method consists of using data confirming the existence of formal connections (for example, taken from official documents, such as membership in associations). The information stored in various databases is also frequently used (e.g., databases of institutions cooperating in research projects or networks of joint authorship of research publications; see, for example, Olechnicka, Płoszaj, 2008; Olechnicka,

Płoszaj, 2010). The necessity to take into account the whole network (no sampling possible) results in the fact that in some cases reliable network analysis may require significantly higher costs than standard quantitative study. Here, funds are not only a concern, but also the time needed to gather and compile the empirical material.

Another limitation results from the fact that network data cannot be analysed using standard statistical methods. A limiting factor here will be the competence for conducting such research. Also, it should be noted that network analysis requires the ability to use specialized software.

Network research is also complicated by its problematic ethical aspects. Network studies, especially those involving people, cannot usually be anonymous, which is not the case with aggregate data from questionnaires, for example. Network analysis, practically by its very nature, involves determining the relationships between specific, and not abstract, actors. While some methods of dealing with such problems have been worked out (see e.g. Penuel *et al*., 2005), when planning network research one should still pay sufficient attention to assuring the subjects' privacy protection.

Despite the discussed challenges, network analysis seems to be a promising tool (or approach) for evaluation. The scope of its use is quite wide and not yet deeply exploited. There is still a lot of space for new practical applications as well as for the development of theoretical approaches of networks in evaluation.

Bibliography

Ahrweiler P., de Jong S., Windrum P. (2002), Evaluating Innovation Networks, [in:] Pyka A., Küppers G. (eds), Innovation Networks: Theory and Practice, Edward Elgar, Cheltenham.

Ball P. (2004), Critical Mass: How One Thing Leads to Another, Farrar, Straus and Giroux, New York.

Batorski D. (2009), Network analysis: methods and application in evaluation, [w:] Haber A., Szałaj M., Evaluation in the Making. Contexts and Methods, PARP, Warszawa.

Bendyk E. (2004), Antymatrix. Człowiek w labiryncie sieci, Wydawnictwo W. A. B., Warszawa.

Besussi E. (2006), Mapping European Research Networks, UCL Working Papers Series (103/2006).

Brik S. (2005), Application of Network Analysis in Evaluationg Knowledge Capacity, [in:] Durland M. M., Fredericks K. A., Social Network Analysis in Program Evaluation, New Directions for Evaluation, 107/2005.

Castells M. (1996), The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I. Blackwell, Cambridge, MA; Oxford, UK.

CASE-Doradcy Sp. z o.o. (2009), Bariery i problemy w realizacji projektów finansowanych z EFS w województwie wielkopolskim, Warszawa.

CASE-Doradcy Sp. z o.o. and EGO s.c. (2009), Ewaluacja pierwszego etapu wdrażania Działania 5.1 Programu Operacyjnego Kapitał Ludzki, Warszawa.

Cross R., Parker A. (2004), The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations, Harvard Business Press, Boston.

Davies R. (2005), Moving from Logical to Network Frameworks: A modular matrix approach to representing and evaluationg complex programs.

Davies R. (2007), Social Network Analysis as an Evaluation Tool: Experiences with International Development Aid Programmes.

Durland M. M., Fredericks K. A. (2005), An Introduction to Social Network Analysis, [in:] Durland M. M., Fredericks K. A., Social Network Analysis in Program Evaluation, New Directions for Evaluation, 107/2005.

EGO s.c. (2008), Ocena udziału polskich partnerów w projektach realizowanych w ramach programu INTERREG IIIB CADSES, Warszawa.

EGO s.c. (2009), Analiza spójności sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego, Warszawa.

Freeman L. C. (2004), The Development of Social Network Analysis: A Study in the Sociology of Science, Empirical Press, Vancouver,

Fredericks K. A. (2005), Network Analysis of a Demonstration Program for the Developmentally Disabled, [w:] Durland M. M., Fredericks K. A., Social Network Analysis in Program Evaluation, New Directions for Evaluation, 107/2005.

Galaskiewicz J. (1985), Social Organization of an Urban Grants Economy, Academic Press, New York.

Hopelli L. (2009), Network Evaluation from the Everyday Life Perspective, Evaluation (15/2009).

Jonak Ł. (2007), Analiza sieci społecznych i dynamika sieciowa, [in:] Nowak A., Winkowska-Nowak K., Rychwalska A., Modelowanie matematyczne i symulacje komputerowe w naukach społecznych – podręcznik dla studentów I roku, Wydawnictwo SWPS Academica, Warszawa.

Kilduff M., Tsai W. (2003), Social Networks and Organizations, SAGE, London.

Klijn E. H. (2005), Networks and Inter-organizational Management: Challenging, Steering, Evaluation, and the Role of Public Actors in Public Management, [in:] Ferlie E., Lynn L. E., Pollitt C. The Oxford Handbook of Public Management, Oxford University Press, New Yourk.

Martin J. (1978), The Wired Society, Prentice-Hall, Englewood Cliffs.

Olechnicka A., Płoszaj A. (2008), Polska nauka w sieci? Przestrzeń nauki i innowacyjności. Raport z badań, Warszawa.

Olechnicka A., Płoszaj A. (2010), Spatial Aspects of Collaborative Networks in Science – Lessons from Poland. Regional and Local Studies. Special Issue 2010.

OECD (1996), The Knowledge-Based Economy, OECD, Paris.

Penuel W. R., Sussex W., Korbak Ch. (2005), Mapping the distribution of Expertise and Resources in School: Investigating the Potential of Using Social Network Analysis in Evaluation, paper presented at the Joint Conference of the Canadian Evaluation Society and the American Evaluation Association, Toronto, 2005.

Provan K. G., Fish A., Sydow J. (2007), Interorganizational networks at the network level: A review of the empirical literature on whole networks. Journal of Management (33).

Provan K. G., Milward B. (2001), Do Networks Really Work? A Framework for Evaluationg Public Sector Organizational Networks, Public Administration Review (61/4).

Rank D., Williams D., (1999), Partial benefit/cost in the evaluation of the Canadian Networks of Centres of Excellence, Ewaluation and Program Planning, 22/1999.

Roediger-Schluga T., Barber M. (2006), The structure of R&D collaboration networks in the European Framework Programmes, Proceedings of the 11th ISS Conference "Innovation, competition and growth: Schumpeterian perspectives".

Rogers E. M. (2003), Diffusion of innovations (5th ed.). New York: Free Press.

Strogatz H. (2003), Sync: The Emerging Science of Spontaneous Order, Theia, New York.

Swianiewicz P., Herbst J., Lackowska M., Mielczarek A. (2008), Szafarze darów europejskich. Kapitał społeczny a realizacja polityki regionalnej w polskich województwach, Wydawnictwo Naukowe Scholar, Warszawa.

Szałaj M., Ledzion B. (2008), Ewaluacja bieżąca sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego, [in:] Olejniczak K., Kozak M., Ledzoin B. (2008),

Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik Akademicki. Wydawnictwa Akademickie i Profesjonalne, Warszawa.

van Dijk, J. (1991), De Netwerkmaatschappij, Sociale aspecten van nieuwe media, Samsom, Alphen aan den Rijn.

Wasserman S., Faust K. (2007), Social network analysis: methods and applications, Cambridge University Press, New York.

Wasserman S., Galaskiewicz J. (1994), Advances in Social Network Analysis: Research from the Social and Behavioral Sciences, Sage Publications, Newbury Park.

Conclusions

Karol Olejniczak, Marek Kozak, Stanisław Bienias

"Hercule Poirot's methods are his own. Order and method, and 'the little grey cells'."
Agatha Christie, "The Big Four. A Hercule Poirot Mystery"

The world around us is becoming increasingly complex, turbulent and unpredictable. Humanity is undergoing structural changes relating to a number of socio-economic processes shaping our surroundings. Many of these processes are contradictory.

Globalization made production, commerce and economic growth in the modern world highly interconnected and interdependent. Some authors even argue that *"the world is flat"*, pointing out the converging processes that takes place around the globe (Friedman, 2007). Others argue that "*the world is spiky"* because economic capital (including knowledge and innovation), wealth, growth and development processes have been highly spatially concentrated, in just a few regions and metropolises of the world (Florida, 2005).

Also, the world economy has been moving from its industrial to post-industrial era. The last 40 years have been marked by a number of phenomena, such as economic restructuring and changing economic drivers. Unlike in the industrial era, it is not infrastructure that drives progress. Contemporary factors of development are knowledge, innovation, human and social capital (including institutional settings) and culture (OECD, 2005). In some countries, processes of economic change, globalization and digitalization were accompanied by a transformation from centrally planned economies to market democracies. For at least some time this imposed another layer of complication and uncertainty (EBRD, 2011).

The transformation of the economy leads to unavoidable adjustments in the social sphere. Among the most noticeable aspects is the creation of an information society as a condition for the functioning of a knowledge – based economy. The digitalization of the economy and networks of social communication became a fact of everyday life, where it is not technical competence to use electronic devices but one's ability to transform data into information and then use it for improving a professional and private life is becoming an issue. Digital exclusion is considered a very special and dangerous form of exclusion, as it can additionally strengthen unwanted results of traditional exclusions (see: Norris 2001).

All of this makes every public intervention more and more difficult. In trying to address these complex, interconnected issues, our programmes have also become complex – multi-objective, multi-annual, cutting across traditional sectors and involving extensive networks of stakeholders.

In order to cope with this kaleidoscope of changes and processes we need a better understanding of the context in which we operate, the mechanisms that shape our environment as well as the consequences of our actions. We also need to constantly stay alert, critically examine our situation, and be ready to adapt. Evaluation is just one feedback mechanism that can help us in doing so. It can provide our institutions with evidence-based knowledge, adjusting both the goals and the toolbox of our development policies. We believe, however, that in order to make evaluation matter, we must address three challenges that emerge for the current Structural Funds practice. These are:

- Being useful to the public and decision-makers
- Being able to explore complexity
- Providing credible evidence

Below, we discuss each challenge in more detail.

Challenge 1. Being useful to the public and decision-makers

Evaluation, despite ambitions of some researchers, is not a basic research science. It is an applied research, which means that it is a consumer-driven exercise. The challenge is to address questions that truly matter for decision-makers and the general public. Moreover these answers need to be provided just in time in order to feed the policy-making process.

Practice shows that stakeholders are interested in the answers to questions of what programmes work, for whom and in what circumstances (Davies et al., 2009). Looking closely at the nature of these questions, two issues emerge for our current evaluation practice.

First, we need to radically switch our attention from narrow, process-oriented studies to a holistic exploration of the effects of our programmes. We should identify and assess real effects, not only those that are planned. This does not mean that we completely ignore the topics of implementation and funds absorption. We just put them in the right place – as the means to reach a goal, not the goal itself. In research practice, it will result in treating the topic of implementation as just one of the explanatory factors of a programme's success or failure in obtaining change.

Secondly, the public expectation of revealing causal mechanisms that led to the observable change put us in the role of detectives. We are supposed to trace factors, reveal hidden determinants that trigger or hamper results and provide evidence that would allow concluding, beyond a reasonable doubt, the cause of change. As some leading American scholars humorously point out, modern evaluation is more

like a CSI job (Crime Scene Investigation) than academic laboratory work[1]. We share this view, as our opening quote of the famous Belgium detective indicates.

The quest for revealing causality requires a good analytical apparatus from us – both approaches and methods that can address this problem. This volume offers some suggestions. However, readers should also be aware of various intense debates taking place around the world. Some developments focus on such issues as:

- a growing concern about the usefulness of Randomised Control Trials for complex programmes (e.g. Cook et al., 2010; Kusek, Rist, 2004; Mark et al., 2011).
- alternative approaches to causal investigation in a social, multifaceted environment (e.g. General Elimination Methods, Modified Success Case Method) (Cook et al., 2010, Appendix A),
- issues of attribution, contribution and substitution analyses in complex interventions (Bamberger et al., 2011, Chapter 16).

This last issue – complexity – leads us to the second challenge.

Challenge 2. Being able to explore complexity

We need a proper toolbox to accurately grasp a multi-faceted and dynamic socio-economic reality. No method or approach provides a holistic solution. Thus, we need to combine different tools of inquiry. Triangulation of data is already a well-established standard in evaluation. We have also been developing methodological cross-examinations. However, we think that we can move forward by triangulating approaches. It seems that a mixed-method is a promising way to do evaluations. It combines ways of comparison and generalization offered by quantitative approaches with inside of qualitative approaches. In our publication we did not devote a separate chapter for this topic, but the philosophy of mixed-method inquiry is present in most of the research cases we discuss throughout the book. Readers who are interested in this topic can explore recent social research literature (Creswell, Clark, 2010). The world of science moves even further beyond traditional borders of disciplines. During the last years there has been a growing literature and practice of truly interdisciplinary research, which are studies that combine different academic fields – from natural science to humanities (Repko, 2011). We think that evaluation could benefit from this new development, especially taking into account the fact that it has already been dealing with public programmes that focus on real-life problems regardless of academic or sectors divisions.

The postulate of looking at the world "through many eyes" clearly implicates methodological diversification and flexibility. In our opinion, we should avoid one-method dominance across the systems of evaluation, because it would narrow down the

[1] We refer to Michael Scriven's speech during the Claremont Graduate University seminar and latest debate interventions by Melvin Mark at the AEA Conference 2011.

perspective and risk omitting important investigation trails. A warning example comes from the United States. Long-term, sophisticated yet highly inflexible RCT (Randomised Control Trials) designs in the field of education totally overlooked the revolutionary, social changes of 1960s in the young population (Chelimsky, 2011). The RCT one-fit-all approach was re-introduced during the Bush administration and resulted in a fierce discussion on the rationality and real utility of this strategy.[2)] We are deeply convinced that exploring more carefully the lines of argumentation used by all sides of this international debate can be highly beneficial for our evaluation practice. The latest conclusion of American evaluators is that the only "golden standard" is that methods should be appropriate to the topic and subject of research (Donaldson et al., 2008). In practice this means weighing strengths and weaknesses of each method, combining them and minimizing their limitations.

When presenting a method or approach, every author in this book points out the strength and limitation of each approach and method. We hope that by doing so we put forward a clear message that we need to embrace different methods, while being aware of their strengths and limitations.

Challenge 3. Providing strong evidence

Various research studies indicate that the utility of evaluation is heavily determined by the quality of evidence provided (Ferry, Olejniczak, 2008; Shulha, Cousins, 1997; Weiss, Bucuvalas, 1980). To put it simple – credible evidence makes evaluation useful. This, however, triggers such questions as: "what counts as strong evidence? When should decision-makers trust evidence? What makes a study trustworthy?".

In the current publication we introduced this issue by discussing the validity and reliability at a very practical level. However, readers should be aware that those apparently simple questions relate to the centre of a wider, fascinating, yet complex discussion on epistemological foundations of modern science. On one hand, studies that follow quantitative orientation have well-established standard criteria for judging their quality as well as strategies to meet these standards (Bryman, 2004, part 2; Shadish et al., 2001, p.37-42). On the other hand, research that follows a qualitative paradigm does not share one standard, mainly due to the unstandardized nature of qualitative exploration. Although some suggestions have been made in this respect both in the academic domain (Flick, 2011) and evaluation studies (Spencer et al., 2003), there is still no unified solution.

This makes our evaluation practice very challenging, especially attempts to combine qualitative with quantitative approaches. We think that this issue is of crucial importance for not only the development of evaluation practice itself, but also for

[2] The high temperature of this debate is well illustrated by the fact that some commentators called it "the war of paradigms".

the development of Evidence-Based Policies, both within and outside the domain of regional development.

The greatest detective in history, Hercules Poirot, gives us a good hint on where to start a discussion on the procedure for properly solving a mystery. First, his concept of "order and method" we interpret as conducting a systematic inquiry. Being methodical is what in fact distinguishes evaluation assessment from our everyday judgements. Second, the expression of "the little grey cells" is a postulate of being intellectually self-reflective and self-critical at every stage of the evaluation inquiry. This allows us to build a coherent story from a body of observations and opinions. The greatest part (and most fun) of Hercules' investigations were his final explanations on how he reached the conclusion. Thus, the third factor is transparency. The trustworthy answer is the one that is clearly presented with all bases of assessment, claims and conclusions explained. The lessons taken from Agatha Christie's mystery books are simply a starting point for our discussion.

The content of this book's chapters and conclusions are just a starting point both for individual exploration as well as a joint discussion. We aim at signalling emerging issues, showing the need of evaluation research and showing the richness of evaluation approaches.

We think it is a good moment to move evaluation forward because we are currently in the middle of the process of rethinking Cohesion Policy, preparing ourselves for the programming of the new perspective of Structural Funds. Proposals of the new regulations for the programming perspective 2014-2020 put much more pressure on the effectiveness and efficiency of Cohesion Policy programmes. One of the key instruments that will allow for the shift towards more result oriented structural support is evaluation. The solutions and good practice examples provided by the authors of this publication are in line with European Commission guidelines and should be carefully examined while shaping the evaluation systems for the next programming perspective.

Even if most readers of this book are involved in evaluation and programming of Cohesion Policy, the lessons learned and inspirations provided by the authors can also be used in other public policies. We should always see things in their broader context – going beyond structural funds.

References

Bamberger, M., Rugh, J. & Mabry, L. (2011) *RealWorld Evaluation. Working Under Budget, Time, Data, and Political Constraints Second Edition*. Thousands Oaks: Sage Publications.

Bryman, A. (2004) *Social Research Methods*. Oxford: Oxford University Press.

Chelimsky, E. (2011) "Final remarks for the conference closing", *Annual Conference of American Evaluation Association*, Anaheim.

Cook, T.D., Scriven, M., Coryn, C. & Evergreen, S. (2010) Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven. *American Journal of Evaluation*, 31(1), 105–117.

Creswell, J.W. & Clark, V.L. (2010) *Designing and Conducting Mixed Methods Research*. London: Sage Publications, Inc.

Davies, H., Nutley, S. & Smith, P. (2009) (ed.) *What works? Evidence-based policy and practice in public services*. The Policy Press, Bristol.

Donaldson, S.I., Christie, C.A. & Mark, M.M. (2008) (ed.) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Sage Publications, Inc, Los Angeles.

EBRD (2011) *Crisis and Transition: The People's Perspective. Transition Report 2011*, EBRD Publications, London, UK

Ferry, M. & Olejniczak, K. (2008) *The use of evaluation in the management of EU programmes in Poland,* Warsaw: Ernst & Young - Program "Sprawne Państwo".

Flick, U. (2008) *Managing Quality in Qualitative Research*. Thousand Oaks: Sage Publications.

Florida, R. (2005) The World Is Spiky. *The Atlantic Monthly*, October, 48–51.

Friedman, T.L. (2007) *The World Is Flat 3.0: A Brief History of the Twenty-first Century*. New York: Picador.

Kusek, J.Z. & Rist, R.C. (2004) *Ten Steps to a Result-Based Monitoring and Evaluation System*. Washington D.C.: World Bank.

Mark, M.M., Patton, M. & Reichardt, C. (2011) "A Conversation With Michael Patton About how Values Undergird the Assessment of Program Effects", *Annual Conference of American Evaluation Association Evaluation 2011: Values and Valuing in Evaluation*, Anaheim.

Norris, P. (2001) *Digital Divide: Civic Engagement, Information Poverty and the Internet Worldwide*, Cambridge, MA: Cambridge Univ. Press.

OECD (2005) *Local Governance and the Drivers of Growth*, Paris: OECD Publishing

Repko, A.F. (2011) *Interdisciplinary Research: Process and Theory*. Thousand Oaks: Sage Publications, Inc.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2001) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Shulha, L.M. & Cousins, B.J. (1997) Evaluation Use: Theory, Research, and Practice Since 1986. *Evaluation Practice*, 18(3), 195–208.

Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003) *Quality in Qualitative Evaluation: A framework for assessing research evidence,* London: UK Government Chief Social Research Office.

Weiss, C.H. & Bucuvalas, M.J. (1980) Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research. *American Sociological Review*, 45(2), 302–313.

GOLDEN PARAGRAPHS

Kathryn Newcomer is the director of the Trachtenberg School of Public Policy and Public Administration at the George Washington University. She teaches public and nonprofit, program evaluation, research design, and applied statistics. She routinely conducts program evaluations and training for federal government agencies and nonprofit organizations. She received the Elmer B. Staats Award for her work on Accountability in Government, presented by the National Capital Area Chapter of the American Society for Public Administration in 2008. She is an elected Fellow of the National Academy of Public Administration, and currently serves on the Comptroller General's Educators' Advisory Panel. She is an elected member of the Board of the American Evaluation Association (2011–14). She served as President of the National Association of Schools of Public Affairs and Administration (NASPAA) for 2006-2007. She has received two Fulbright awards, one for Taiwan (1993) and one for Egypt (2001–04). And, she has lectured on performance measurement and public program evaluation in Poland, Ukraine, Brazil, Egypt, Taiwan, and the UK.

Michał Wolański Ph.D. is Adjunct Professor at the Warsaw School of Economics, Department of Transport. He has both extensive academic and consultant experience. He was a visiting researcher at the University of Sydney, Institute of Transport and Logistics Studies as well as took part at a number of research projects at the Technical University of Warsaw, University of Gdansk and Leon Kozminski University in Warsaw, as well as at the home university. His advisory experience includes projects for such customers, as European Investment Bank, Polish Ministries of Regional Development and Infrastructure, Railway Regulatory Office as well as local authorities and transport operators. His Ph.D. thesis has been awarded the best Ph.D thesis in transport in 2010 by Polish Minister of Infrastructure.

John Bachtler is Professor of European Policy Studies at the University of Strathclyde (Glasgow) and Director of the European Policies Research Centre. He specialises in comparative research on the design, implementation and evaluation of regional development policies across Europe and particularly EU Cohesion policy. He has published extensively in these fields and has been an expert adviser to numerous national government departments and European institutions.

Ewan Ferlie is Professor or Public Services Management at King's College London. He has published widely as aspects of public management and on organisational change, especially in the fields of health care and higher education where he has contributed to a cross European comparative project. He is co editor of the Oxford Handbook of Public Management published by Oxford University Press.

Richard Hummelbrunner is a senior associate of ÖAR Regionalberatung Graz, Austria with more than 30 years of professional experience as a consultant/evaluator in the fields of regional and international development. During recent years he has been active in promoting the use of systems thinking in evaluation as a practitioner, trainer and author.

Martin Ferry is a Senior Research Fellow at the European Policies Research Centre (EPRC) at the University of Strathclyde. He specializes in regional economic development and policy in Europe and the governance of national regional policies and EU Cohesion policy with emphasis on the UK and Central and Eastern Europe. He has been involved in several major evaluation studies of EU territorial and spatial development policies, including cross-border cooperation programmes and the evaluation of Structural Funds programmes in different parts of the UK.

Alberto Martini is associate professor of Public Policy Evaluation at the Università del Piemonte Orientale in Northern Italy. He is also directing Progetto Valutazione, a non profit research center to disseminate the use of rigorous evaluation methods. Professor Martini has 25 years of experience in the application of quantitative methods to policy issues. After a Ph.D. in Economics from the University of Wisconsin in 1988, he worked on the evaluation of social policies by means of experimental and non-experimental methods at Mathematica Policy Research and Urban Institute. Since taking an academic position in Italy in 1998, he has conducted evaluations in a variety of fields, including enterprise support, employment policies, and education. He is directing a random assignment evaluation of an employment program for persons with mental illnesses. He authored two monographs in Italian quantitative methods for program evaluation, and wrote the section on impact evaluation of the EVALSED Guide. In 2001–02 he was President of the Italian Evaluation Association.

Tomasz Gapski is a graduate of european studies at the Jagiellonian University and PhD studies in economics at the Warsaw School of Economics. Chief Specialist at the National Evaluation Unit in Ministry of Regional Development of Poland. Since 2003 he deals with analysis and evaluation of public policies including in particular European Cohesion Policy. Lead author of guidelines for evaluation of cohesion policy in Poland, author of evaluation reports and publications, lecturer at the postgraduate studies.

Maureen A. Pirog is the Rudy Professor of Public Policy Analysis in the School of Public and Environmental Affairs at Indiana University. She holds a joint appointment with the Evans School of Public Affairs at the University of Washington. Professor Pirog is also the editor-in-chief of the Journal of Policy Analysis and Management, the ISI's top-ranked journal of public administration, a position which she has held for over seven years. Professor Pirog research interests are focused on the methods

and trends in the evaluation of government programs, policies, laws and managerial innovations as well as social programs and problems including child support enforcement, welfare-to-work, employment and training programs, adolescent parenting, and social program up-take by mixed status (legal and illegal immigrants) families.

Lanlan Xu is a Ph.D. candidate in the School of Public and Environmental Affairs at Indiana University. Her concentrations are policy analysis and public finance, with a Ph.D. minor in Sociology. She is interested in urban poverty and economic inequality. Her research is focused on the evaluation of government programs and policies such as child support enforcement, low-income housing policy, education policy and immigration policy.

Adam Płoszaj holds Master of Arts in Spatial Economy. Since 2006 he is a scientific assistant at the Centre for European Regional and Local Studies EUROREG, University of Warsaw. He has participated in numerous research and applied projects, including a number of evaluations, on regional innovation systems and strategies, local and regional development, sustainable development, information society, research policy, research and development organization, scientometrics, higher education, territorial cooperation, inter and intra organizational cooperation. He has conducted research projects in Poland, France, Ukraine and Canada.

Stanisław Bienias is the head of National Evaluation Unit, Department of Structural Policy Coordination, Ministry of Regional Development. Since 2004 Stanisław Bienias has been working in the Polish Ministry of Regional Development for the National Evaluation Unit, Since 2006 as head of this unit. He has been responsible for organising the process of evaluation on the level of the National Development Plan 2004–2006 and the National Strategic Reference Framework 2007–2013. Furthermore, he has coordinated the work of more than 50 evaluation units located in operational programmes and for building the evaluation culture in public administration. He took part in the preparation of more than 100 evaluation researches and dozens of conferences and evaluation trainings. Stanisław Bienias represents Poland in DG REGIO Evaluation Network meetings and gives lectures concerning evaluation, monitoring and indicators on postgraduate studies.

Grzegorz Gorzelak, Professor of Economics, has scientific experience in applied research on economic, social and territorial development, spatial planning and EU policy development in that field. Has coordinated a number of domestic and international research projects and is active in providing expertise to DG Regio (recently on territorial cohesion and evaluation of Cohesion policy). Has extensively published papers and books on regional development and regional polices in the new Member States and recently on the financial crisis in Central and Eastern Europe. He is the president of the Polish section of the Regional Studies Association and Editor-in-Chief of the refereed journal *Studia Regionalne i Lokalne*.

Marek W. Kozak is associate professor of physical planning at the Center of Regional and Local Studies EUROREG, University of Warsaw. He is teaching regional/cohesion policy, tourism in development and evaluation of local and regional development. Teaches European cohesion policy at the Collegium Civitas, Warsaw. Involved in international research projects (ESPON, 5, 6 and 7th framework Programme, Regional Studies Association Research Network). He is member of the Committee of Spatial Economy and Regional Planning, Polish Academy of Sciences. For most of his over 30 years long career he worked mostly in academia, except for 1991–2002 when he was employed initially as a director at the Office for European Integration and Foreign Assistance (part of Prime Minister Office) and served later as a CEO of the state organization Polish Agency for Regional Development (in charge of pre-accession integrated regional development programmes financed by the EU and other donors). Member of the advisory team to Ministry of Regional Development. Deputy Chairman of the State Council of Physical Planning. He authored well over 100 articles and authored or co-authored 6 books (plus 5 books edited).

Dominika Wojtowicz, Assistant Professor in Department of European Studies at Kozminski University. Her research interests focus on the widely understood effectiveness of EU cohesion policy European Union. She has extensive practical experience – took part in a series of research projects and evaluations of Regional Operational Programs and projects financed from EU funds. Involved in few international research projects (eg. PARITEA – coordinated by International European Movement, Ex Post Evaluation of Cohesion Policy Programmes 2000–06 Co-financed by ERDF realized for European Commission, COCOPS project – Coordinating for Cohesion in the Public Sector of the Future – implemented within 7th framework Programme). Works as an advisor to local governments in the acquisition and use of EU funds field. She is the author of a number of papers and books on determinants of effective absorption of European Union funds, their effects on local and regional development and methods of evaluation.

Veronica Gaffey is acting Director in the Policy Development Directorate of the Directorate General for Regional Policy of the European Commission. This Directorate has responsibility to design the future Cohesion Policy, to define the territorial and urban dimensions of the policy and to carry out economic analysis and evaluation. Since 2007, she has been head of the Evaluation Unit of the DG. During 2006, she was head of the Innovative Actions Unit of DG REGIO. In the Evaluation Unit in DG REGIO, her priorities have been to significantly improve the evidence available on the performance of Cohesion Policy. Together with her colleagues she launched an ambitious ex post evaluation of the 2000–2006 period which has delivered new insights into the performance of the policy which are feeding into the policy debate. The second priority is to improve the reporting by Member States on the achievement of their objectives, to support the building of evaluation capacity in

the Member States, and to encourage Managing Authorities to carry out more evaluations exploring the effects and impacts of the policy.

Karol Olejniczak is an assistant professor at EUROREG – University of Warsaw, Poland (www.euroreg.uw.edu.pl). He holds MA degrees in regional economics, MPA in European public affairs and PhD in management. He specializes in the evaluation of public interventions, mainly in the fields of regional policy, territorial cooperation and good governance. He has professional experience in evaluation research executed for Polish government and for the European Commission. He has been conducting research in United Kingdom, Ireland, The Netherlands and U.S.A. (visiting scholar at Professor Elinor Ostrom's worskhop on Policy Analysis). His current academic interest focuses on organizational learning in public sector. Since 2008 Karol is the head of the Academy of Evaluation – an elite post-graduate study programme for Polish senior civil servants.

Łukasz Widła-Domaradzki is currently a senior researcher in Ipsos Poland, with 12 years of professional experience as a statistician and analyst in social and market research. He studied sociology at the University of Warsaw and worked as an analyst first in MDM (Medical Data Management) and later in Millward Brown SMG/KRC. Up to now, in cooperation with University of Warsaw as well as government organizations, he took part in various scientific and evaluation projects involving the development of new quantitative analysis and research methods. He is currently responsible for designing statistical models for measuring the change in organizational learning in Polish ministries as well as building structural models in TERCO (European Territorial Cooperation as a Factor of Growth, Jobs and Quality of Life). Łukasz is also an experienced trainer in statistics and quantitative methodology. He co-authored several evaluation reports and chapters concerning quantitative methods in evaluation projects.