Evaluating the effects of regional interventions.

A look beyond current Structural Funds' practice

Warsaw, November 2011

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.

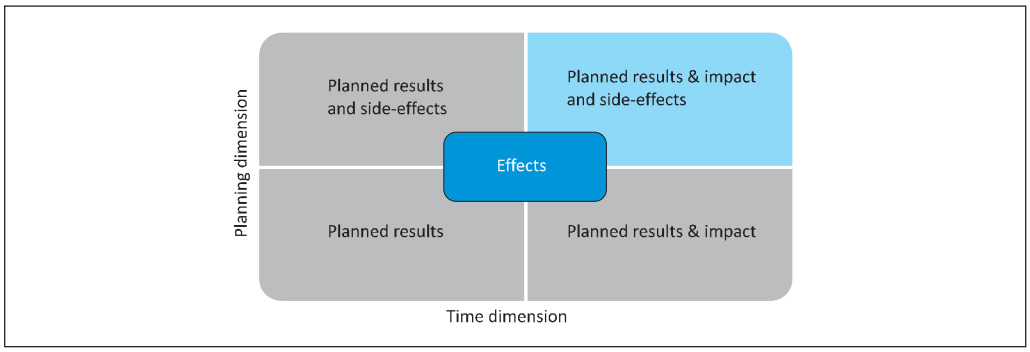
6 Mark *et al.* (2000, p.75) refer to it as "inquiry modes", arguing that this is a link connecting overall objectives of the study with very technical methods used in the analysis of a given intervention. On the other hand,Stufflebeam& Shinkfield (2007) adopt a broader interpretation, and use the term "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.

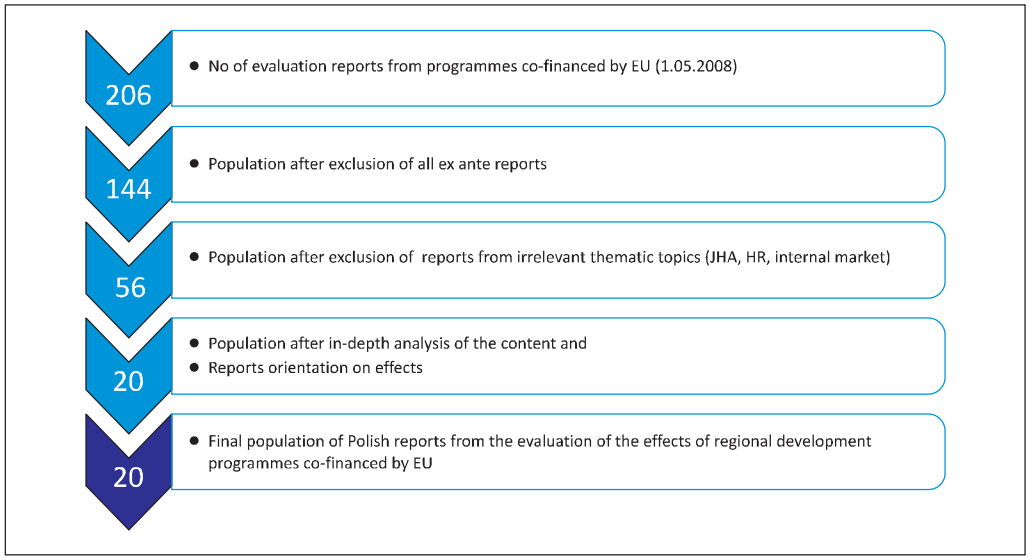
In the first stage the subject of the research was selected – these were 100 evaluation reports. Based on the findings from previous analyses, four EU countries with extensive experience evaluating public programmes were chosen (Bachtler *et al*., 2000; Furubo *et al*., 2002; Toulemonde, 2000). These 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

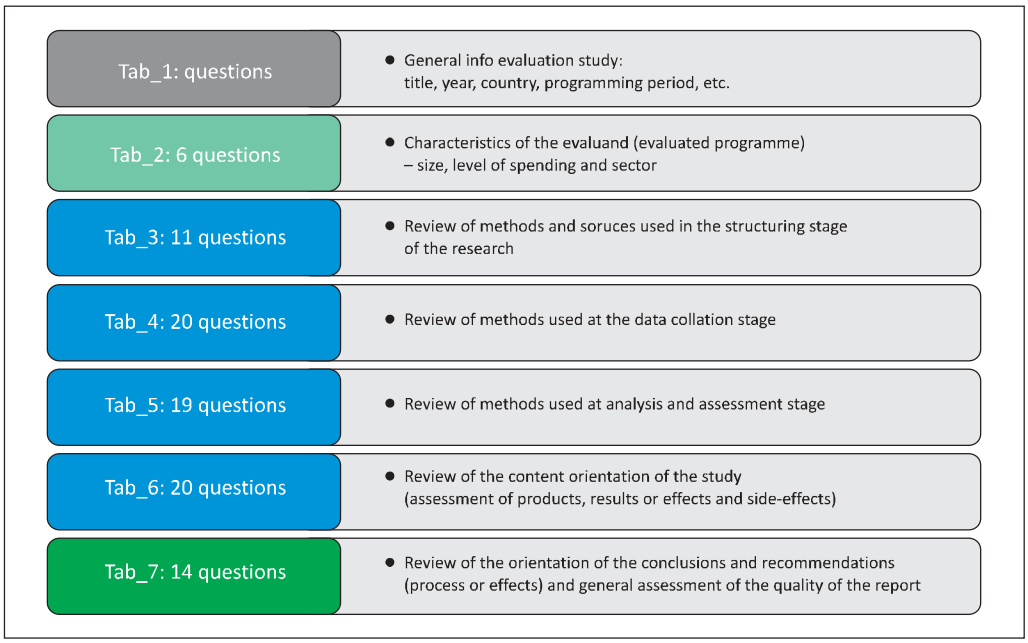


Source: prepared by the author.

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.

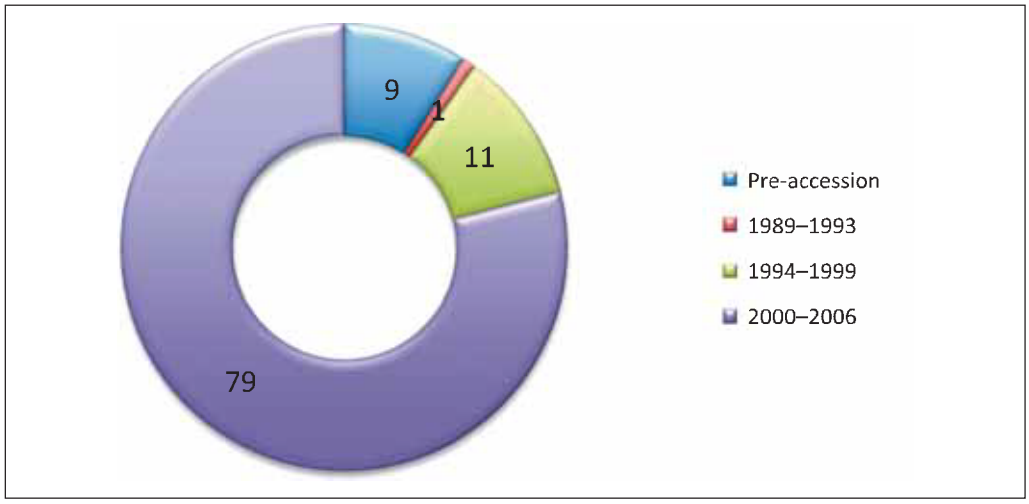
3. Results of the analysis

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.

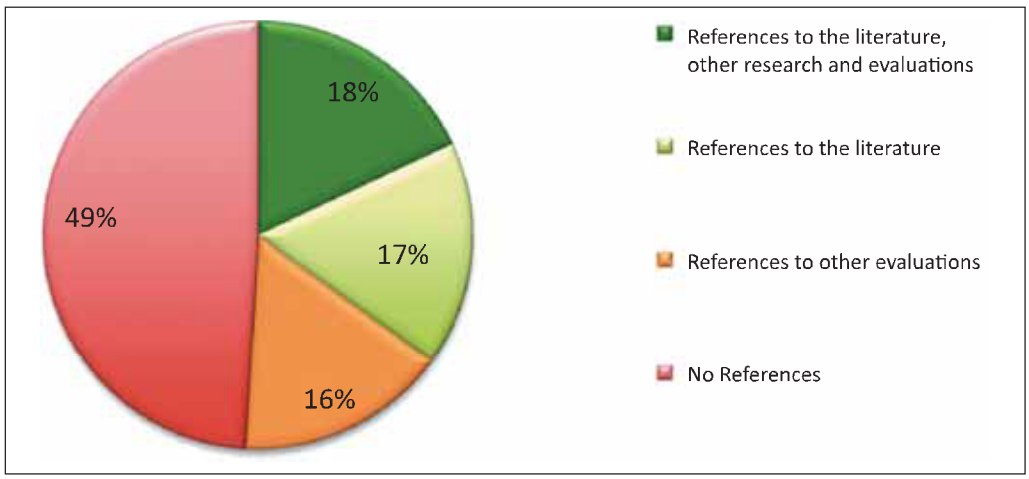
Only a small share of the studies (18%) were founded on broader methodological and conceptual groundwork, whether theoretical (textbooks and manuals) or practical (other research studies). The bulk of such casesare from Sweden (7) and – quite encouragingly – from Poland (also 7).

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.

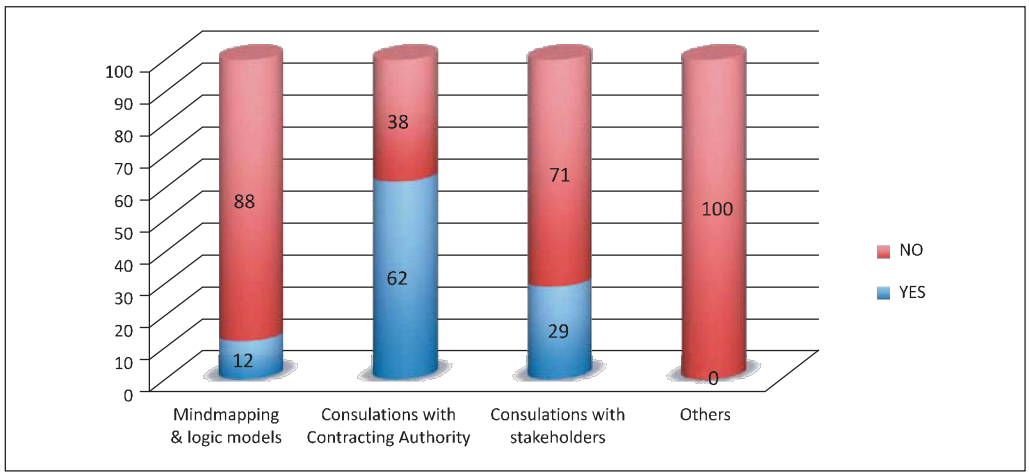
Figure 5. Sources used in the development of methodology



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.

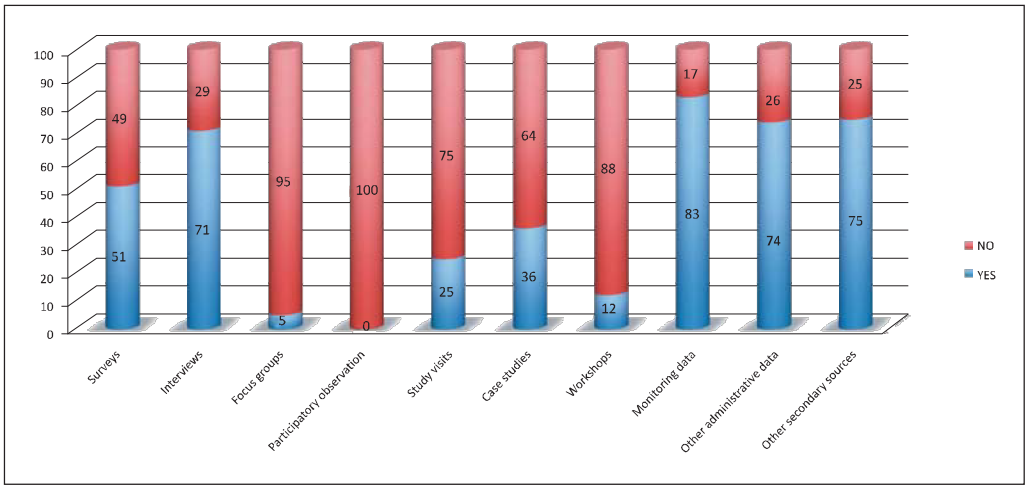
In nearly 40 situations, case studies were used to collect and then analyze the data. This was usually connected with on-site visits.[12)]As a rulethe subjects of case studies were projects; although there were a few situations when the beneficiaries' partnerships were surveyed (e.g. institutions or 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

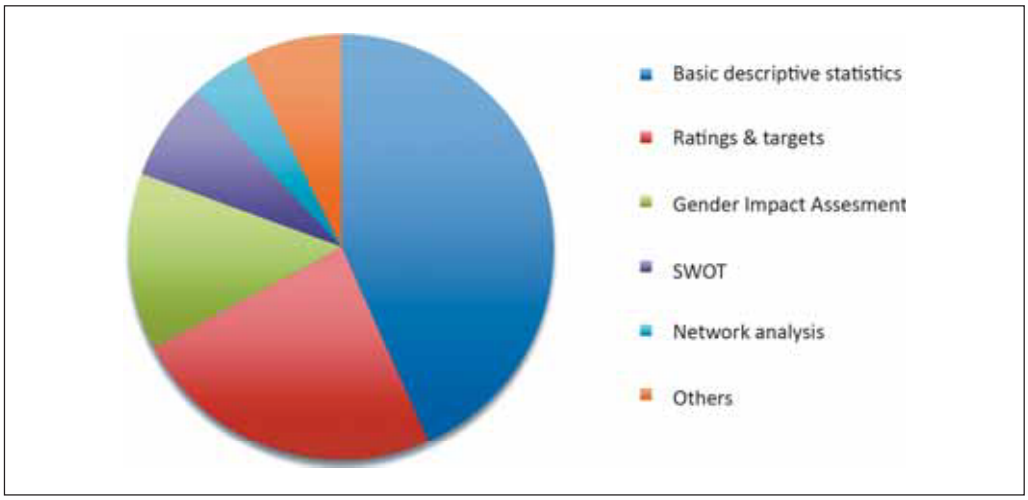
Programme	Method/Tool	Description
Poland: *ex post* evaluation of the PHARE-CBC 1994–99 Poland-Germany Programme *ex post* evaluation of Poland-Germany Programme	School questionnaire Review of local press	In order to assess the impact of the PHARE-CBC programme on the awareness of the residents of the 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.

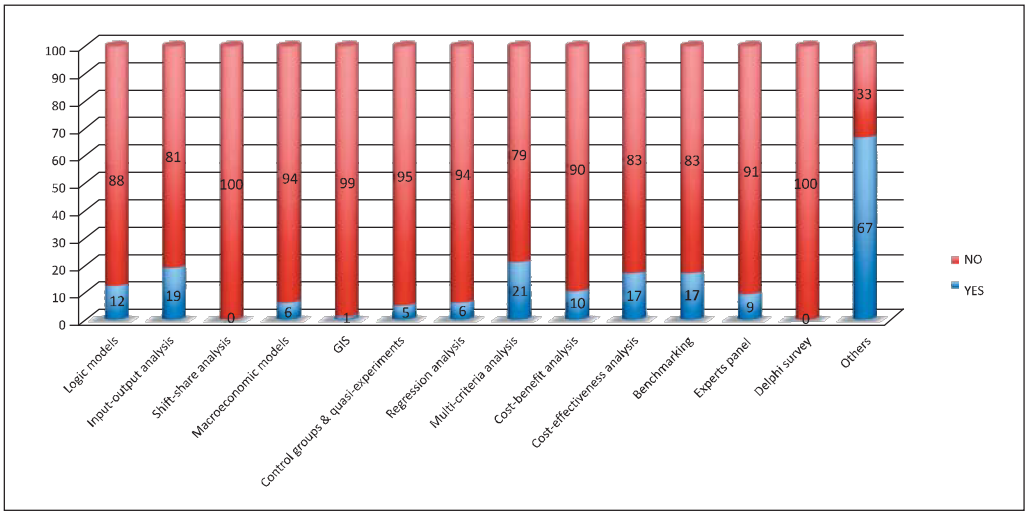
Figure 8.Analysis and judgement methods labelled as "others"



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.

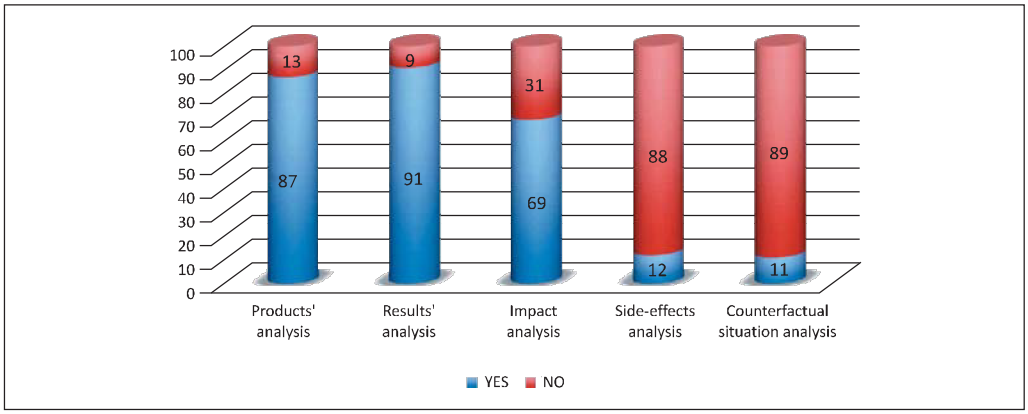
Source: prepared by the author on the basis of the project's database and evaluation reports

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

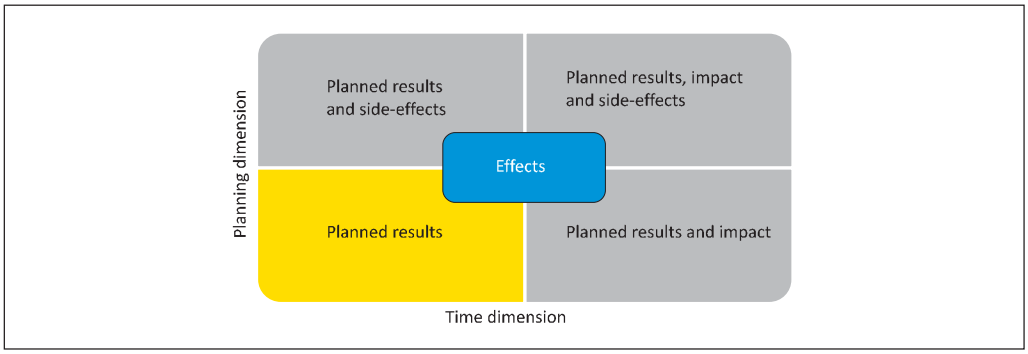


Source: prepared by the author on the basis of the project's database.

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.

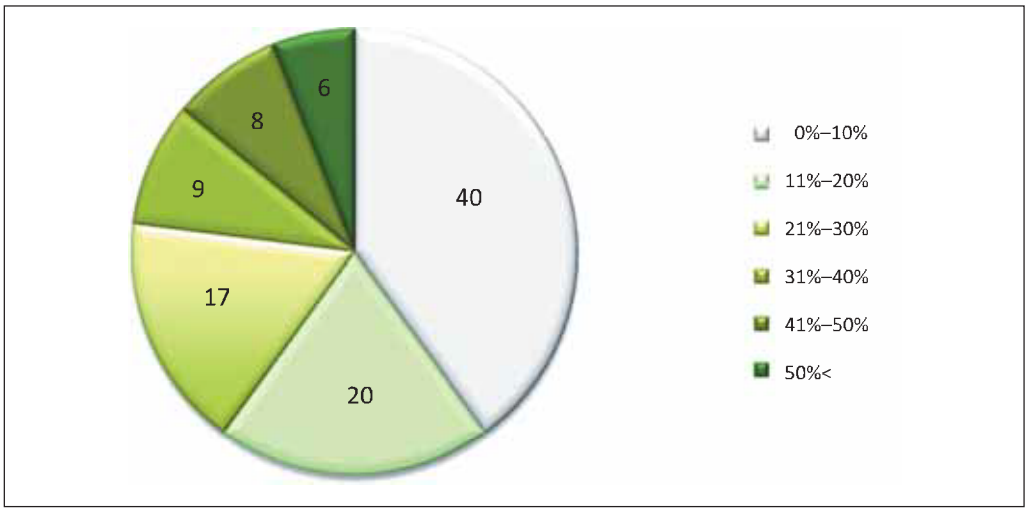
Figure 11. Two dimensions of intervention effects



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.