ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 223, 2008

*Justyna Wiktorowicz**

EVALUATION OF STRUCTURAL FUNDS – METHODOLOGICAL ASPECTS

Abstract. The instrument of evaluation is being more and more valued, and the need to deploy it is perceivable in almost every field of activity. Evaluation should impose rational actions, therefore it has the same importance for the public sector as the market mechanism for the private sector. The visibly growing popularity of evaluation research results from modern tendencies of organization management (among others *Total Quality Management, New Public Management*), and in the context of the European Union from the increasing role of structural funds.

This paper has a survey character and deals with problems of evaluation in the wide perspective. The essential aspect of study is the methodology conducting evaluation of structural fund. In this paper the peculiarities of using different methods during individual evaluation phases have taken into account and preliminarily assessed their usefulness. Author has done the review of literature, legislation and Polish and UE documentation, as well as evaluation reports with the aim of indicating tendency in the application of quantitative data in evaluation.

Key words: evaluation, structural funds, methodology of the project evaluation.

1. INTRODUCTION

The instrument of evaluation is being more and more valued, and the need to deploy it is perceivable in almost every field of activity. Evaluation is – simply put – an assessment, and a well-managed organization has to be assessed. Until recently, such need has been merely perceived in relation to companies, but according to the recommendations contained in the Lisbon Strategy, every organization should be capable of learning. In accordance with the idea of *New Public Management* that is based on the central concepts of *value for money* and *Activity Based Management*, evaluation began to be perceived in the mid-80's an one of the elements of reforming the public sector (Olejniczak 2004, p. 79). The evaluation of programs financed from the European Union budget has begun to play an especially important role.

Evaluation should impose rational actions, therefore it has the same importance for the public sector as the market mechanism for the private sector. The

*Ph.D., University of Lodz.

visibly growing popularity of evaluation research results from modern tendencies of organization management, and in the context of the European Union from the increasing role of structural funds. Taking the dynamic development of management theory into account, as well as the development of European Union's structural policy (and in the case of Poland – it's fresh accession to the Union), one may assume that the interest in evaluation will be systematically increasing and is going to play a part of a characteristic "mark of management quality", also with regard to administration.

The common duty of evaluating all programs financed from the European Regional Development Fund, the European Social Fund and the Cohesion Fund has been specified in Commission Regulation No. 1083/2006, regulating, among other things, the evaluation of Union funds during the programming period of 2007–2013 (European Commission 2006). In accordance with the aforementioned documents, evaluation – along with intensified auditing and monitoring procedures – has become one of the main tools for increasing the transparency of decisions made by Union institutions, guaranteeing rationality and quality, as well as for respecting the value for the European taxpayer's money. These measures are also indirectly aimed at increasing the credibility of European institutions in the eyes of citizens.

In accordance with the European Commission's requirements, subject to evaluation are socio-economic projects and programs. During the programming period of 2004–2006, these were: the National Development Plan (NDP), operational programs and provincial contracts. In the years 2007-2013, in turn, subject to evaluation will be all operational programs deployed within the National Strategic Reference Framework (NSRF) for the years 2007–2013. The National Assessment Unit (NAU) created in the Ministry of Regional Development is responsible for NDP, whereas the evaluation of individual operational programs is carried out by evaluation teams situated in the institutions that manage these programs (e.g. PART the Polish Agency for Enterprise Development).

This paper deals with problems of evaluation in the wide perspective and has a survey character. The essential aspect of study is the methodology conducting evaluation of structural fund. In this paper the peculiarities of using different methods during individual evaluation phases have taken into account and preliminarily assessed their usefulness. Author has done the review of literature, legislation and Polish and UE documentation, as well as evaluation reports with the aim of indicating tendency in the application of quantitative data in evaluation.

2. EVALUATION – BASIC DEFINITIONS AND CLASSIFICATIONS

At present, evaluation is a key element of modern management practice in the public sector. It has no precise definition, however. In its broadest sense, it is understood as using socio-economic research methods for the systematic verification of the effectiveness and efficiency of public programs (Olejniczak 2004, p. 4). L. Koprowicz presents a broad depiction of evaluation that stresses its utilitarian character; he defines evaluation as a systematic examination of values or properties of a concrete program, measure or object, from the viewpoint of chosen criteria, in order to improve, develop or better understand them (Koprowicz 1997). The Scientific Evaluation Association, in turn, states the following: "To evaluate a policy means to examine whether the legal, administrative and financial means deployed by the program made it possible to reach the envisioned effects of a given policy and the goals attributed to it" (Conseil Scientifique de l'Evaluation 1996). The United Nations, however, perceive evaluation as a process aimed at defining as systematically and objectively as possible the accuracy, efficacy and effects of a given measure in relation to its goals (UN). M. Q. Patton, for many years president of the American Evaluation Association, created one of the most accurate definitions, describing evaluation as a process of systematic information gathering on the effects, properties and results of programs, staff and products used by specialists to reduce uncertainty regarding the program and improve its effectiveness, that also facilitate decision making based on what these programs, products and the staff do and what they relate to (Patton 2001). This definition not only systematizes the distinctive properties of evaluation (information gathering, a vast array of problems being the subject of evaluation, aim and the use of research), and not only shows that it is a universal technique, not attributed to any specific economy sector or organization type, but also stresses that it is only effective if its results are used by decision-makers in a constructive way. This means that the basic quality criterion of evaluation research is its usability.

One should not confuse evaluation with monitoring and audit (which unfortunately happens quite often, making research difficult for evaluators). Yet, monitoring aims at improving management in the public sector through intervention (it solely monitors the program deployment progress). The audit, on the other hand, solely focuses on the financial aspects of programs and their legality.

To close the above discussion, the definition used by the European Commission will be used as a point of reference for further considerations. In accordance with that definition, the evaluation of a policy, program or project is perceived as defining the value of a policy, program or project in relation to previously defined criteria, based on appropriate information (Kierzkowski 2002, p. 12).

In most cases, the evaluation practice distinguishes between ex-ante, mid-term and ex-post evaluation. This classification corresponds to the previously discussed classification depending on the time limit in relation to program deployment and is regulated by European Council Regulation of June 21, 1999 (Polish Economic Association 2003, p. 8).

Ex-ante evaluation takes place at the beginning of a project cycle, before its deployment. Its task is to guarantee that the program's premises are internally coherent and match the actual social needs. In accordance with Art. 48, par. 2, Council Regulation No. 1083/2006, ex-ante evaluation is aimed at optimizing the allocation of budget resources within operational programs and improving the programming quality. It focuses on the analysis of strengths and weaknesses of the program, diagnoses the quality of the assumed aim-achieving strategies and the aims themselves, as well as the program's feasibility. *Ex ante* evaluation is also the basis for monitoring measures and future evaluations through specifying measurable indicators of achieving the aims assumed within the program. Thus, in the long run, it supports the decision-making process related to the program. Apart from the undeniable effect towards budget means optimization, ex-ante evaluation additionally facilitates the program's promotion and the propagation of information about this program.

In Poland, at the turn of the year, an ex-ante evaluation of all programs for the programming period of 2007–2013 was performed. This evaluation was performed by external evaluators. The experiences from this evaluation were the subject matter of an open discussion during last year's Second Evaluation Conference in November, organized by the Ministry of Regional Development (MRD) and PARP.

Mid-term evaluation (evaluation in the middle of the program's deployment) analyses its first effects, the quality of its financial management, the quality of the program's means of implementation and the monitoring that is being conducted. It also examines whether any changes in the program's socio-economic environment has taken place since performing the *ex ante* evaluation. The *midterm* evaluation is based on information gathered through the monitoring system, i.e. on indicators specified during the *ex ante* evaluation. It also refers to research results of similar programs deployed in the past, if such programs exist. It is formative in its nature - i.e. depending on its results corrections may be performed with regard to the deployment of the program that is subject to the examination. Conclusions from the *midterm* evaluation of programs deployed by means of the European Unions' structural funds are also one of the important factors allowing the European Commission to make decisions related to the allocation of the so-called performance reserve.[1] Since this evaluation is performed

[1] In accordance with Council Regulation No. 1260/99, 4% of the means from structural funds granted a given member state at the beginning of the programming period is not actually programmed, but is withheld as the so-called performance reserve. This reserve is used to finance

during the program's duration, it mostly focuses on expenses and results, not on the influence of the measures being deployed.

In addition, during the intervention, especially in the face of certain problems with the program's operation, an **on-going evaluation** is performed. The purpose of the on-going evaluation is to offer support during the program deployment process and current verification of the rationality of measures undertaken; it may also be an element of a future ex-post evaluation. It may be particularly helpful for clarifying reasons for occurring blockades and delays. It allows to analyze and preliminarily assess the products and effects of the investment in question, as well as evaluate the monitoring and deployment system. Thanks to this kind of evaluation it is usually possible to show to which degree preliminary assumptions could be put into practice from the operational viewpoint and assess their remaining accuracy. When comparing the present condition with the initial situation, it is also possible to assess how much the context of the intervention has changes and to which degree the premises remain valid. The on-going evaluation makes it also possible to verify if the given program is congruent with other measures undertaken simultaneously by the public sector. In effect, it is feasible to work out management system corrections and suggest possible program shifts.

To sum up, on-going and mid-term evaluations are meant to improve the quality of deploying measures by the public sector, facilitate learning processes and strengthen the partnership and joint ownership, as well as responsibility.

The **ex-post evaluation**, on the other hand, sums up and assesses the program as a whole, especially from the viewpoint of its effects, after its deployment. The program is analyzed first and foremost with regard to effectiveness, efficacy and quality of using resources assigned for deployment. The evaluation is also usually supposed to formulate conclusions that go beyond the examined program and can be related to other programs, regions etc. *Ex post* evaluation it therefore rather summarizing in its nature, but due to the fact that information that is indispensable for assessing the actual effects of the program is often unavailable for many years after the program had ended, ex-post evaluation has in many cases a restricted ability to deliver a full assessment of the program's effects. Since many EU programs are replaced by successive programs, questions that are characteristic for a shaping evaluation may be justified even at the ex-post assessment stage. In practice, this form of evaluation is carried out at the latest within 3 years after the programming stage has ended, in accordance with the European Commission's recommendations.

---

operational programs or their priorities which, based on a set of indicators previously agreed upon between the Commission and a given member state or region, show the best results with regard to the degree of achieving the assumed aims, management quality and deployment advances.

Due to the specificity of each aforementioned evaluation type, different program assessment criteria may be verified at different stages of a program's evaluation and deployment. At the stage of ex-ante, mid-term or on-going evaluation it is not possible to credibly specify the intervention's usefulness or durability, whereas from the viewpoint of ex-post evaluation it is difficult to determine the intervention's accuracy.

Another criterion of classifying evaluation types is the question of people performing the evaluation being dependent on the people deploying the program. If they are independent from each other (i.e. there is no actual or potential conflict of interest), we deal with an external evaluation. If they are from the same institution (e.g. a specialized entity), they may be performing an internal evaluation. It is also possible that people directly involved in managing and employing a given program are performing the evaluation. In such a case we deal with self-evaluation. Due to the above distinction we talk about a different evaluation culture. The evaluation culture differs in individual EU member states. In Italy, for example, this distinction is regional in nature – the north of the country counts on external evaluation, whereas in southern Italy usually internal evaluation takes place. Sometimes a mixed model is used (Raimondo 2006). In the Czech Republic and in Holland, on the other hand – similarly to Poland – external evaluation predominates (Kokeš 2006). It should be noted that none of these solutions is ideal. Internal or self-evaluation is undoubtedly connected with perfectly identifying the issue at hand, can suffer from a lack of objectivism, however. In the case of external evaluation, on the other hand, we deal with a fresh look at the program being assessed, in most cases, however, the program is insufficiently examined and understood.

When we talk about evaluation culture in a broader sense, the status in EU-15 countries can be summarized as follows (Yuill, Bachtler 2006):

– Germany, Netherlands, Sweden, the UK have an evaluation culture that pre-dates the Structural Fund,

– Ireland, Italy, Austria and Finland have an evaluation method that developed with the Structural Funds,

– France, Spain, Portugal and Greece did not have a deeply rooted tradition of policy evaluation.

Approaches to Structural Funds evaluation in EU-15 countries could be characterized as follows:

– Belgium, Denmark, Ireland ⇒ beyond regulatory requirements,

– Austria, Finland, Germany, the Netherlands, Sweden, the UK ⇒ some additional measures,

– Greece, Italy, Luxembourg, Portugal and Spain ⇒ generally restricted to regulatory requirements.

3. EVALUATION RESEARCH TOOLS

An evaluation examination should be performed in four stages: structuring evaluation, obtaining data, analysis and assessment that accounts for future recommendations. Each of these stages requires different analysis methods to be used. Due to the fact that quantitative methods on which I would like to focus in this paper are especially useful at the observation and analysis stage, I am going to put the emphasis on the methodology of these evaluation stages (the remaining aspects will be the subject matter of another work).

At the **structuring evaluation** stage, the following must be specified: subject, aim, entity, method, assessment criteria, as well as time and resources. An important element is also the establishment of indicators for measuring the desired program effects. When defining the evaluation object, the logical structure of the assessed investment must be specified, among other things. The following methods are most often used at this stage: concept or issue mapping, stakeholder consultation, evaluability assessment, logic models, formative/developmental evaluation. In Poland, the methods most often used are logic models, especially logical framework, concept mapping of impact and SWOT.

During the **obtaining data** stage, on the other hand, the subjective and objective examination scope is defined (e.g. which persons should be interviewed and on what topics), on the basis of which data and facts are gathered that are connected with the assessed program. For this, both primary data (e.g. from survey research performed by the evaluation team) and secondary data (e.g. from monitoring) are used. Therefore, the following are most often used at this stage: social survey, beneficiary surveys, individual (stakeholder) interviews, priority evaluation, focus groups, case studies, local evaluation, participatory approaches and methods, use of secondary source data, administrative data and observational techniques. Usually, more than one data collection method is used during the research. Such behavior makes it possible to supplement information gathered by means of one method with other pieces of information. This is also beneficial from the viewpoint of verifying and intensifying the collected data. The so-called triangulation, i.e. the diversity of techniques related to analysis and information collection, makes it possible to recognize and understand the examined object and creates the opportunity to perform references and comparisons. Triangulation may be used not only with regard to methods of data collection (diversity of the methods used), but also to information sources (data collection from different respondent groups). This allows it to create a broader research material for evaluation and conclusion, which in turn makes it possible to create a possibly objective analysis that takes into account the viewpoint of many different groups interested in the examination subject. As the practice shows, for evaluation purposes in Poland, the following are most frequently used: participating observation, secondary data analysis, individual interviews, focus groups, survey

research, case studies. The list of these methods clearly indicates connections between evaluation methodology and social research.

Participating observation (official or anonymous) delivers information on the actual actions undertaken by the participants of the process, allows to recognize the context of the actions, makes it possible to ask questions and verify the answers during research. This method, however, is of limited usefulness, mainly due to the large subjectivism of the assessment, difficult access to all phenomena, as well as being time-consuming. It is mainly used during *ex post* evaluation.

Secondary data analysis can take the form of an analysis of existing statistical data (e.g. Main Statistic Office data, data from monitoring), of archival documents (e.g. court registers, governmental and quasi-governmental documents, data from registers kept by the public registrar's offices) and/or content analysis (literature overview). In my opinion, the greatest disadvantage of such data is the fact that usually they do not fully reflect the researcher's needs (no data at the chosen aggregation level, no comparable data in certain sections, time series too short, and in certain cases unclear method of data collection or variable definition, which reduces the reliability of such information) and are usually not easily accessible (especially in the short time the evaluation is normally planned for). These objections do not discredit the purposefulness of using secondary sources, since they deliver information on formals aspects of measures, assumptions and proofs of their deployment. They allow to look at reality from the viewpoint of anticipated goals and values and "proofs" of their deployment. Let us hope that the increasing demand of evaluators for such data is followed by adjusting them to – often quite specific – evaluation needs. This, however, will require a closer cooperation of the Main Statistic Office with entities responsible for evaluations. The creation of a list of indicators by the European Commission that are to be obligatorily used for evaluating programs in the years 2008-2013 may be the right step in this direction. The next step should be adjusting the research of the Main Statistic Office to these requirements.

The next method group – individual interviews – is directed at people involved in designing programming documents, responsible for the deployment of the program with its direct and indirect beneficiaries and people who, despite fulfilling appropriate conditions, were not involved in the program's deployment. Besides individual interviews, focus group interviews are very popular with evaluation research. During the conversation, people involved in the program are provoked by the presenter to present their opinion on the program, including value judgments, which activates creativity among the participants of the debate. The aim is to thoroughly examine detailed issues through finding out how the debaters perceive them. The interviews allow to obtain information on the participants' experiences, their interpretations, feelings, motives etc.

A method related to individual interviews is survey research. In this case, we also ask about the participants' experiences and feelings but – due to the form of

the research – we usually reach a greater number of recipients. The research instrument is the survey that contains (or at least should contain) clearly formulated questions related to a given intervention. This method allows to reach a significantly greater number of entities connected with the intervention, which – especially in the case of an ex-post assessment that is connected with the necessity to reach a large number of entrepreneurs, employees, unemployed etc. – is quite important. As the practice shows, the problem is that the return of these questionnaires is not satisfactory, which may discredit the credibility of such research. An important aspect is also the question of choosing a representative sample of the whole population (that oftentimes is not well-defined or is devoid of the proper register that could serve as a drawing appraisal study). A difficulty of using this method effectively lies in the frequent lack of interest of recipients to fill out the questionnaire. Despite these restrictions, evaluators willingly use this method due to its relatively low cost. This goes especially for mid-term and ex-post evaluation.

A case study actually uses many classic social research methods simultaneously (interviews, focus group interviews, participating observation etc.). It is most commonly used when evaluating a specific issue, especially with ex post evaluation. Projects that are representative for the whole program or one of its components are chosen for the examination, also including, however, model projects or projects that failed to meet the expectations.

Table 2

Techniques of obtaining data in the evaluating process

Method	Ex-ante				Ex-post and mid-term		
	Design	Obtaining data	Analyzing data	Judgments	Obtaining data	Analyzing data	Judgments
Social surveys		x			x		
Beneficiary surveys					x		
Individual (stakeholder) interviews		x			x		
Priority evaluation				x			
Focus groups		x	x		x	x	
Case studies		x	x		x	x	
Local evaluation					x	x	
Participatory approaches & methods	x			x			x *
Use of secondary source data		x			x		
Use of administrative data		x			x		
Observational techniques						x *	x *

* only in-depth analysis

Source: own analysis based on: Resources for the Evaluation of Socio-Economic Development (www.evalsed.com).

The aforementioned methods (interviews, surveys, case studies) allow to view the reality of a given intervention through the eyes of the people involved (on different levels). In connection with the analysis of existing data and participating observation, they allow to recognize the problem comprehensively and create a database for a deep (depending on the needs) analysis and intervention assessment. For comparison, methods used during the observation stage have been synthetically presented in Table 2.

Having collected the appropriate data, the evaluation team uses appropriate techniques in order to interpret the data, compare it and estimate the program's effects (analysis stage). The Sourcebook of evaluation published on the website www.evalsed.com suggests to use the following methods to analyze the pieces of information: input/output analysis, econometric models, regression analysis, experimental and quasi-experimental approaches, Delphi survey and SWOT analysis. The subject literature also contains other suggestions, however. The manual of structural fund evaluation (Polish Economic Association 2003; Polish Economic Association 2005) gives the following classification of analysis methods: statistical methods (descriptive statistic methods, statistical conclusions), analytical techniques (cost-benefit analysis, cost–effectiveness analysis, SWOT), analytic models (input-output models, micro- and macroeconomic models, statistical models) and methods of analyzing quality data. M. Ekiert, on the other hand, states (see: Evaluating socio-economic programs. MEANS Collection. Vol. 3, European Commission 1999, pp. 89–91, 103–110) that the following analysis methods are most commonly used: analysis of comparative groups, shift-share analysis, macroeconomic model and factorial analysis.

Due to the specifics of the publications, typical quantitative methods, i.e. input/output analysis, econometric models, regression analysis, experimental and quasi-experimental approaches, Delphi survey and SWOT analysis will not be discussed extensively, I will only show their advantages and limitations from the viewpoint of evaluation research.

The most interesting contribution of input-output matrices concerns impacts on sector distribution and trade. The value of this approach lies above all in the broadness of its scope and in its coherent treatment of the main economies of the European Union benefiting from Objective 1 support. Unfortunately, input-output matrices are limited to the estimation of effects on demand, rather than supply. Therefore, they do not take into account one of the most important objectives of interventions, i.e. lasting effects on productive potential. Most effects on supply, which are likely to lead to a sustainable increase in the growth rate of assisted regions and enable these regions to catch up with more developed areas, are totally overlooked (for example, the creation of new productive capacity, construction of infrastructure, productivity gains throughout

the economy, spread of technological progress). These effects cannot be estimated using this tool.

With regard to econometric models, on the other hand, the HERMIN model exists for evaluation purposes in Poland, similarly to many other EU countries, that has been originally created for the needs of Irish economy. At first, this model was used in four countries: Greece, Ireland, Portugal and Spain, and after 1997 it has been also deployed by the Czech Republic, Slovenia and Romania. Estonia, Latvia and Poland deployed it after 1999. In the years 2000-2003 it was also deployed by the German Eastern Lands, Northern Ireland and Mezzogiorno in Italy, and in 2003 by Hungary. Since 2005, it has been used in Poland on the provincial level as well (Zaleski et al. 2007). Each HERMIN model has three broad sub-components (a supply side, an absorption side and an income distribution side) which function as an integrated system of equations. A conventional Keynesian aggregate demand mechanism underpins the absorption side of the model. There is some degree of sector disaggregation with a supply-side sub-component helping to determine traded (manufacturing) output as a consequence of national price and cost competitiveness. Interest and exchange rates are exogenous to the HERMIN model, in line with the general assumption that the cohesion economies are 'small' and 'open'. (Bradley 1997, 2006). This model predicts changes quite accurately in the aforementioned area, still, it is not free from flaws. The main complaint is that it has been too mechanically transferred into Polish conditions which - in some cases – causes some lack of classificatory coherence (e.g. the classification of activity is not adapted to the Polish Classification of Activity).

Apart from the HERMIN model, the QUEST, E3ME and REMI models are used for evaluation research (http://www.evalsed.info/page.aspx?id=mth120).

QUEST is developed by the European Commission Service Directorate and is a multi-country model designed to analyze the business cycle, the long-term growth of the European Union Member States and the interactions of these states with the rest of the world, especially with the United States and Japan. This model has real interest rates and exchange rates determined endogenously, and this does allow for the possible 'crowding-out' effects of Structural Funds on the private sector to be taken into account (EC 1997).

E3ME, an energy-environment-economy model for Europe, is a multi-sector, regionalized, dynamic econometric model of the EU. It is not a Computable General Equilibrium (CGE) model, but a disaggregated time-series, cross-section econometric model that has benefited from some of the techniques used in CGEs relating to calibration on recent data. The model combines economic, energy and environmental components (CEC 1995).

Until recently, the REMI Policy Insight Model has been only applied in North America, but within the past year or so some applications have been

carried out on structural funds impacts for the European Commission. The model is econometric in origin, but the structure is the same for all market-based economies except for differences in a few key parameters such as the speed of migration response to changes in economic conditions and the response of wage rates to labor market conditions. The model parameters are estimated over a large sample of regions and are used for all implementations of the model (F. Treyz, G. Treyz 2003).

Summarizing, the goals of the Structural Funds are defined at the macroeconomic level. The output from a macro-econometric model is therefore generally consistent with the requirements, meaning that this is practically the only tool that can be used to formally ascertain whether an European policy has achieved its aim. The work involved in constructing such models from scratch usually means that existing models, which may not be ideally suited to the purpose of evaluation, are adapted instead. There is a serious weakness in them. This applies, among other things, to constructions of production functions.

Regression analysis is used in the evaluation to understand the statistical dependence of one variable on other variables. When it is successfully executed (with a statistically valid adjustment), regression analysis can produce a quantitative estimate of net effects. One should remember that relations between the different explained and explanatory variables are often circular (X explains Y and Y explains X). In this case, it will be better to use correlation coefficients.

Experimental approaches are often described as the 'gold standard' of evaluation, but using this approach for the evaluation of a social program has both proponents and opponents. On the one hand, it is the only way in which the causal relationships which are assumed to be the basis of a particular intervention can be scientifically proven. On the other side, there is a danger that the results, even if statistically significant, will fail to contribute in a real way to the understanding of the effectiveness of the intervention under different circumstances. Experimental and quasi-experimental approaches are frequently supported by statistical modeling techniques such as probity analysis, survival analysis and hierarchical regression analysis. In the classic depiction, methods of statistical inference such as variance analysis, t-test or their nonparametric equivalents are used.

Delphi is primarily used to facilitate the formation of a group judgment. This evaluation method tends to be used when significant expertise exists on the subject, but it may also be used to specify relations of causes and potential effects in the case of innovative interventions. The method is recommended in an ex ante evaluation context if the evaluation concerns public intervention of a technical nature. Thus, it was very often used in the framework of energy policies, for example, for prospective studies on the impact of changes in taxation. In the case of Structural Funds evaluation, for example, the Delphi

inquiry has been recommended for obtaining macro-economic estimations when the phenomena involved are complex (for example, to quantify the impact of a major infrastructure project). A major problem is the tendency for experts to over-simplify particular issues and treat them as isolated events. This is particularly the case in forecasting.

All aforementioned methods may be classified as statistical methods or analytical methods. One should bear in mind, however, that all used methods have limitations. The advantage of statistical methods is undoubtedly the possibility of assessing the statistical credibility of the results, which allows the evaluator to draw conclusions and makes it possible to sum up the evaluation results in a clear, transparent and reliable way. On the other hand, however, not all program effects can be analyzed using it. The resulting data and results should be always approached with humility, one should examine the quality of the data, their credibility and reliability, and also bear in mind the assumptions of the analysis. If these assumptions are disproved, the obtained results should be treated with great caution or – preferably – alternative tools should be used. In the case of analytical models one should bear in mind their economic assumptions, since even small changes in the basic structure of these models can lead to different conclusions (it is therefore recommended to perform a solution sensibility analysis or apply an alternative approach in order to confirm the obtained conclusions). Analytical techniques, on the other hand, even such as SWOT, suffer from a large amount of subjectivism and, what is more, do not allow to explain concrete results or consequences. Their use also requires the results to be confirmed by means of other methods.

Table 3

Techniques of the analysis in the evaluation

Method	Ex-ante				Ex-post and mid-term		
	Design	Obtaining data	Analyzing data	Judgments	Obtaining data	Analyzing data	Judgments
Input/output analysis			x			x	
Econometric models			x			x	
Regression analysis						x *	
Experimental and quasi-experimental approaches					x *	x *	
Delphi survey		x *	x *				
SWOT	x						x

* only in-depth analysis

Source: own analysis based on: Resources for the Evaluation of Socio-Economic Development (www.evalsed.com)

As already mentioned, some authors also name other analytical techniques as useful at the analysis stage, e.g. cost and advantage analysis. These methods, however, are more useful at the assessment stage.

At the last evaluation stage (*judgments*), the program's effects in relation to certain criteria (evaluation questions) are assessed and synthetic conclusions and recommendations are formulated, i.e. the actual quality of the program in question is examined. Tools to form evaluative judgments are: cost-benefit analysis, benchmarking, cost effectiveness analysis, economic impact assessment, gender impact assessment, environmental impact assessment, strategic environmental assessment, multi-criteria analysis and expert panels. For evaluation purposes in Poland, the following are most frequently used: expert panels, benchmarking, cost-benefit analysis and cost effectiveness analysis. In addition, SWOT, econometric models and input-output analysis are often used. As can be seen, some of these methods are perceived as useful by some authors already at the analysis, or even the observation stage. This naturally depends on the specificity of the given evaluation. For instance, in the case of an ex-ante assessment, particularly more advanced techniques are usually used, e.g. multi-criteria analysis or expert panels, whereas at the ex-post or mid-term evaluation stage, cost and advantage analysis or benchmarking are also widely employed.

4. CONCLUSION

Evaluation research plays an ever-growing role. It constitutes an inherent element of managing structural funds in EU countries, obviously including Poland as well. This influences the development of both evaluation methodology and the market of evaluators as such. A shortcoming of the present Polish evaluation market is the still very faint theoretical basis for evaluation measures being undertaken. This is the result of the discontinuity of evaluation measures, and is also due to the fact that the Polish evaluation market is dominated by consulting companies. There is also a deficiency of scientific publications in this area. This leads to a situation where EU regulations and the European Commission's methodological notebooks are considered as theories. The very weak involvement of research centers in the development of evaluations also seems quite disturbing. It is first and foremost the consulting companies that react to the growing needs of the market. From higher education institutions comes a very weak response. Unfortunately, this leads to some shortcomings in the presented evaluation reports, especially from the methodological viewpoint.

Due to the intensity of programming works the problem of performing document work simultaneously with the evaluation of these documents arises. This occurred especially with ex-ante evaluation. Unfortunately, it led to a situation where evaluators assessed a version of the document that had already been

modified before the end of the evaluation. Obviously, this makes it impossible to propagate organizations capable of learning and questions the plausibility of performing the evaluation as such.

The solidification of the positive participation evaluation model, i.e. the co-operation of evaluators with the managing institution, may be recognized in favor of Polish evaluation procedures. This creates a positive atmosphere of cooperating for a common goal, which is to obtain a high quality programming documentation. An ever-improving cooperative atmosphere between experts and civil servants is created. What is more, the evaluators' technique is improving despite the lack of major interest from scientific circles. Previously, document analysis dominated the evaluation techniques, whereas now various types of interviews, participating workshops, brainstorming sessions, simulations of socio-economic processes, reality falsifications and adequacy matrixes are becoming more and more popular. The public administration's potential also develops, proper authorities create suitable organizational entities which staff can order and obtain evaluation analyses.

Undoubtedly the biggest problem of all evaluations is the indicator system. This results from the generally low measurability level of various socio-economic phenomena, but also from the insufficient quality of statistical data that often make it impossible to compare the effects in space or time. Due to the short period of our membership and, following from this, the short time of the functioning of interventions being evaluated, there is no sufficiently long time series that would allow, for example, reliable econometric modeling or predictions. Issues with assigning the net effect, the program's actual added value also remain – it is difficult to assess, for example, to which degree the increased employment in a region is the effect of a given program, and to which degree it the result of an independent event (Szlachta 2006).

It should be clearly stressed that evaluation is not the monitoring of an intervention, but its assessment and, following from this, it is not enough to simply present the facts or subject them to a quantitative and qualitative analysis, which – unfortunately – happens continuously in evaluation reports. The main purpose of evaluation is the program's assessment, and therefore the basic product of evaluation should be recommendations that are legal, connected with the research topic and also suitable for the intervention in question. Experience shows, unfortunately, that there are evaluation teams suggesting program structures that would violate previously accepted agreements that are often a compromise of many parties. They do not consider the consequences of their suggestions, especially at the deployment stage. This constitutes a problem for the customer ordering the evaluation and is often an argument in favor of performing internal evaluations. Evaluation practice shows, however, that such solution only works well in the case of ex-ante and on-going evaluation. Mid-term and ex-post evaluation should be performed by external evaluation teams.

As already mentioned, evaluation leads to establishing a program's effects. This would not be possible without using the proper research tools. During the assessment, we distinguish desirable from side effects. The latter may be either positive (leverage, impulsion & multiplier, synergy, additionality) or negative (deadweight, displacement, substitution, double-counting). The main purpose of the evaluation should be the assessment of the intervention's net effect, i.e. an effect that has been purified from co-influences of other factors (e.g. economic prosperity). It is recommended to determine the net effect with the following formula:

net effect = gross effect – deadweight – substitution – displacement.

Unfortunately, the estimation of each aforementioned component is not easy, which makes the assessment of the program's net effect (its added value) extremely difficult. An inherent element of the assessment are indicators that constitute a quantitative measure of changes occurring as an effect of the measures deployed. The measurement of indicators, however, is merely a necessary condition, but it is not sufficient to perform the intervention's assessment. Since evaluation research puts an ever-growing emphasis on estimating the intervention's net effect, it should be assumed that the methodology is going to develop exactly in this direction. In my opinion, an interesting solution may be the use of logistic regression that may serve for instance as a means to estimating the probability of finding a job due to the intervention (taking many variables into consideration may lead to assessing the co-influence of many factors, including – of course – the program in question, and thus show the net effect of a given intervention as compared to other determinants).

To sum up, one should bear in mind that in accordance with EU regulations, all measures financed from structural funds must be subjected to evaluation that should constitute an element of operational and strategic management, having the same importance for the budget sector as the market mechanism for the private sector. This should positively influence the development of the evaluation market in Poland, while the obligation to publish the results should translate into greater attention to evaluation quality. In addition, through using the key indicator system, comparative evaluation analyses regarding similar interventions in the whole European Union and in individual member states will become possible.

REFERENCES

Bradley J., *Aggregate and Regional Impact: The Cases of Greece, Spain, Ireland and Portugal*, Office of the Official Publications of the European Communities, Brussels, 1997

Commission of the European Communities, *E3ME: An Energy-Environment-Economy Model for Europe*, No 16715, Official Publications, Luxembourg 1995

Conseil Scientifique de l'évaluation, *Petite guide de l'évaluation des politiques publiques.* Mars Paris: CSE, 1996

Council Regulation (EC) No 1083/2006 of 11 July 2006 laying down general provisions on the European Regional Development Found, the European Social Found and the Cohesion Found and repealing Regulation (EC) No 1260/1999

Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down general provisions on the Structural Found

Ekiert K., *Ewaluacja w administracji publicznej. Funkcje, standardy i warunki stosowania,* Ministerstwo Rozwoju Regionalnego, Warszawa 2004

Ewaluacja Funduszy Strukturalnych, Ministerstwo Gospodarki, Pracy i Polityki Społecznej, Warszawa 2003

Ewaluacja – kwestie ogólne, materiały szkoleniowe, Warszawa 2005

European Commission, *QUEST II. A Multi-Country Business Cycle and Growth Model*, Economic Papers, No. 123, October 1997

Evaluating EU Expenditure Programmes: A Guide: Ex post and intermediate evaluation including glossary of evaluation terms, http://www.europa.eu.int/comm/budget/evaluation/ key-documents_en.htm

Ex-ante evaluation: A Practical Guide for Preparing Proposals for Expenditure Programmes, http://www.europa.eu.int/comm/budget/evaluation/keydocuments_en.htm

Kierkowski T., *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, PARP, Warszawa 2002

Kokeš M., *Ex-ante evaluation of NSRF 2007-2013 – Czech experience*, II Konferencja Ewaluacyjna, Warszawa, 22.11.2006

Koprowicz L. (ed.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warszawa 1997

Olejniczak K., *Ewaluacja programów pomocowych Unii Europejskiej – wprowadzenie*, [w:] G. Gorzelak, J. Bachtler, M. Kasprzyk (red.), *Współpraca transgraniczna Unii Europejskiej. Doświadczenia polsko-niemieckie*, Wydawnictwo Naukowe Scholar, Warszawa 2004

Patton M. Q., *Utilisation - focused evaluation,* Sage, Beverly Hills, CA 1986

Patton M. Q., *Qualitative Research & Evaluation Methods*, Wyd. 3, SAGE Publications, Newbury Park, CA 2002

Raimondo L., *The Italian experience in conducting ex-ante evaluation for 2007–2013 period in the light of lessons learned in 2000-2006 period*, II Konferencja Ewaluacyjna, Warszawa, 22.11.2006

Rossi P., H. Freeman, *Evaluation. A systematic approach*, Sage, Newbury Park, CA 1993

Szlachta J., *Polskie doświadczenia w zakresie ewaluacji,* II Konferencja Ewaluacyjna, Warszawa, 22.11.2006

The Mid Term Evaluation of Structural Fund Interventions; http://www.europa.eu.int/comm/regional_ policy/sources/docoffic/working/sf2000f_en.htm

Treasury H. M., *Policy evaluation: a guide for managers,* Londres: Her Majesty's Stationary Office, London 1998

Treyz F., G. Treyz, *Evaluating the Regional Economic Effects of Structural Funds Programs Using the REMI Policy, Insight Model*, 5th European Conference on Evaluation of the Structural Funds, Budapest, 26–27.07.2003

Yuill D., J. Bachtler, *Evaluatingthe Cohesion Policy: The Development of Ex Ante Evaluation of the Structural Funds*, II Konferencja Ewaluacyjna, Warszawa, 22.11.2006

Zaleski J., J. Bradley, P. Tomaszewski, M. Zembaty, *Ocena wpływu funduszy Unii Europejskiej na polską gospodarkę za pomocą modelu makroekonomicznego HERMIN*, Konferencja „Wpływ funduszy UE na sytuację makroekonomiczną Polski", Warszawa, 25.01.2007

Zarządzanie cyklem projektu, Urząd Komitetu Integracji Europejskiej, Warszawa 2003

Justyna Wiktorowicz

EWALUACJA FUNDUSZY STRUKTURALNYCH – ASPEKTY METODOLOGICZNE

Ewaluacja staje się coraz bardziej doceniana, a potrzeba jej prowadzenia zauważana jest w niemal każdej dziedzinie działalności. Ewaluacja powinna wymuszać racjonalność działań, jest więc dla sektora publicznego tym, czym mechanizm rynkowy dla sektora prywatnego. Obecna rosnąca popularność badań ewaluacyjnych wynika ze współczesnych tendencji w zakresie zarządzania organizacjami (m.in. *Total Quality Management*, *New Public Management*), a w kontekście Unii Europejskiej z wzrastającej roli funduszy strukturalnych.

Niniejsza praca ma charakter przeglądowy i porusza problematykę ewaluacji w szerokim ujęciu. Kluczowy aspekt analizy stanowią zagadnienia metodologiczne związane z ewaluacją funduszy strukturalnych. Autorka dokonała przeglądu literatury przedmiotu, aktów prawnych oraz dokumentacji polskiej i unijnej, a także raportów ewaluacyjnych w celu wskazania tendencji w zakresie aplikacji metod ilościowych w ewaluacji.

Słowa kluczowe: ewaluacja, fundusze strukturalne, metodologia badań ewaluacyjnych.