Agnieszka Haber

Process of ex-post evaluation – conceptualisation, operationalisation, implementation of evaluation survey

Introduction

Evaluation is now a very fashionable term connected mainly with managing public financial support from structural funds. Its main task is to provide justification for adopted strategic and operational decisions\(^1\). It is often perceived as a description and valuation of the effects of a given policy, intervention, programme but also as a tool for defining problems and limitations of a specific reality and searching for such strategies and solutions which would solve these problems with the use of specific instruments. Therefore, by its very nature, evaluation has all features of applied research and thereby it does not differ in principle from other research undertakings in terms of designing techniques of collecting data and methods of analysis\(^2\).

The present chapter is precisely a description of evaluation as a research process, aiming at obtaining objective, independent assessment based on reliable empirical data obtained according to the principles of selected methods, a process which requires a careful consideration, design and implementation in order to be completed.

Objective of ex-post evaluation

In relation to ex post evaluation we can say that it is a systematic and objective assessment of a completed project, programme or policy – in the context of their planning, implementation and obtained results\(^3\). Its objective is the determination of real effects and justification of intervention in a particular form. In other words, its tasks are, for example: to assess the level of implementation of planned objectives, products and results; to determine whether these intervention objectives and priorities maintained until the end their adequacy in relation to initially formulated problems; to assess the impact of an undertaking and sustainability of its achievements; to formulate a diagnosis of effects (planned and unplanned); to identify strong and weak points of an intervention; to indicate directions of development and modification of future interventions.

---


However, irrespective of its detailed scope, ex-post evaluation should provide reliable and useful information due to its **conclusive function**, or in other words accounting function (**accountability**), and to a smaller degree its **cognitive function** (**knowledge**) and its **formative function** – stimulating facilitation and organisational development (**development**) in relation to designing and implementing future interventions of similar nature⁴.

Specifying the need to conduct evaluation, the objective for which it is to be implemented and key functions it has to perform, is the first step both in the process of searching for information and answers to survey questions, as well as in conceptualisation and operationalisation of the evaluation study.

**Conceptualisation and operationalisation of ex-post evaluation**

The process of preparing the study for implementation should be divided in two phases. The first of them, **conceptualisation** (planning the evaluation), includes 1) specifying the need, value and importance of evaluation of a particular project/programme, 2) possibility of implementation of evaluation, 3) use of the results, 4) specifying potential areas of risk and 5) its political implications⁵ (in other words, the vision of success of a study included in the initial idea). The second phase, **operationalisation** (designing the evaluation), is our vision of implementation and ensuring the quality of evaluation as well as estimating the necessary resources.

**Conceptualisation – evaluation planning**

The first phase of evaluation covers accepting a project for further development and strengthening objectives of the project implementation through 1) institutional regulations (internal or external, resulting from formal requirements, e.g. obligation of institutions to conduct evaluation), 2) detailed specifying of objectives, 3) evaluation functions and 4) consequences of the assessment (e.g. decision to improve, reorient, continue or cease to implement the next programme edition; in case of finding irregularities, legal and financial consequences) and in this context, 5) identifying the evaluation recipients⁶ and 6) their main information needs. This is a phase when foundations for managing the product quality, communication and distribution of results which are decisive in the potential evaluation success, are laid. Only foundations constructed in such way may be completed with further content influencing the initial evaluation project.

The second phase is detailed specification of the evaluation object. This stage may be described as a “look at the programme”, including e.g. the following:

- type of intervention and its characteristics,
- schedule of project implementation,

---

⁴ More about this issue in the article included in this publication by K. Olejniczak, *Teoretyczne podstawy ewaluacji ex-post* [Theoretical basis of ex-post evaluation]; J. Gomiak, *Funkcje ewaluacji* [The functions of evaluation], in S. Mazur (ed.), *Evaluacja funduszy strukturalnych – perspektywa regionalna* [Evaluation of structural funds – regional perspective], p. 16–23, Uniwersytet Ekonomiczny w Krakowie [Economic University in Cracow], Kraków 2007; D. Nevo, *Konceptualizacja ewaluacji edukacyjnej* [Conceptualisation of educational evaluation], in L. Korporowicz (ed.), *Evaluacja...,* op. cit., p. 54.


⁶ See table 1.
• intervention objectives,
• area covered by an intervention (country, region, etc.),
• information/data describing the real course and method of implementation (e.g. beneficiaries’ reports, interim monitoring reports, databases etc.),
• a moment when we perform assessment,
• effects of a programme (planned and unplanned, positive and negative, direct and indirect, short- and long-term etc.),
• information/data describing the degree of achievement of the objectives (expressed through objective indicators),
• information/data necessary for evaluation of macro-economic impact of an undertaking,
• information/data concerning the strategy (policy) towards the sector/field/issue which is a object of an evaluated undertaking,
• information/data describing external factors (indicated and non-indicated in the programme documents of an undertaking) influencing the process of implementation and effects of an undertaking (e.g. impact of the socio-economic context, institutional and political environment etc.).

The third phase is the specification of the scope of ex-post evaluation. It usually takes place in several steps. The first one is specifying the level of study (e.g. in case of evaluation of structural funds, whether the evaluation is performed from the level of NDP/CSF or NCS/NSRF, an operational programme, selected priority axes or whether it is limited to a specific measure, scheme or project) and the evaluation’s type, that is whether it will be of complete nature, relating to the whole intervention (as for example Ex-ante Evaluation of the Cross Border Cooperation Programme Lithuania–Poland 2007–2013; Ex post Evaluation of the Rural Development Plan for 2004–2006) or whether it will refer to a selected issue (as for example Assessment of progress in implementing Measure 2.6 of the IROP and trainings for PEC in Małopolska; Impact of implementation of the SOP ICE on the level of innovativeness of Polish enterprises), or whether it will cover horizontal issues concerning more than one programme (as for example Assessment of impact of EU structural funds on the condition of natural environment).

The second step is determining frameworks, how broad our study will be, e.g. whether we refer to one or several elements of an intervention; whether we refer to one or several groups of recipients (beneficiaries, institutional partners, potential recipients etc.); whether we take into account only the examined intervention or consider its potential synergies with other interventions or public programmes; whether we take into account regions, groups of people which are not/were not included in the programme but may experience its impact; whether we take into consideration issues connected with the next programming cycle etc.\(^7\). In other words, we adopt assumptions concerning the characteristics, nature of an examined reality/object (to what extent it is variable and what the cause of change is), the method of treating the study object from the point of view of its autonomy (i.e. whether we treat it as a closed, isolated system or as an open system) and, in further steps, concerning potential possibilities of becoming acquainted with this reality through a specific method\(^8\).

The third step is specifying the time perspective in which we would perform the assessment after the programme completion. In case of ex-post evaluation, the time of measurement depends directly on the intervention type, beneficiaries’ characteristics, possibility of measuring the effects and impact (short- or

---


\(^8\) Compare L. Słoma, *Metody i techniki badań socjologicznych* [Methods and techniques of sociological research. Wybrane zagadnienia [Selected questions], Uniwersytet Warmińsko-Mazurski [University of Warmia and Mazury region], Olsztyn 2002, p. 29.
long-term). Therefore, for example projects aimed at investment support for enterprises (e.g. co-financing of purchase of fixed assets such as machines or appliances) may not be evaluated sooner than within 12-24 months after the investment completion because the effects and their impact on operation and development of a beneficiary (enterprise) requires time for e.g., modification of the production process; training the personnel; implementation of new products and introducing them into market; increasing the production and turnover; lowering the enterprise’s debt; employing new persons and increase of investment outlays from the enterprise’s own assets and liabilities (excluding the grant). On the other hand, for short projects aimed for example at improving qualifications (e.g. trainings for employees of enterprises, unemployed persons etc.), at the same time (1-2 years after their completion) it would be too difficult for many beneficiaries to recall them in their memory. Therefore, such projects should be definitely evaluated earlier: between three months to one year after completion of participation in the programme.

The fourth step consists in the conceptual structuring the subject, which is specifying the meaning of all concepts which we will examine in order to systemise communication (with recipients of the evaluation results and respondents) and to avoid potential misunderstandings concerning the examined reality. Explication of concepts is a continuous process, starting at the level of evaluation conceptualisation and made more precise in the course of collecting and interpreting data. E. Babbie, citing C. Marchall and G. Rossman, indicates the so-called “conceptual funnel”, thanks to which the researcher’s interests become more and more concentrated. Therefore, for example, the increase of competitiveness or innovativeness of enterprises on a global scale may become narrower even to the level of individual situation of enterprise groups selected according to features such as: enterprise size, years of operation, sector, operation scale (local, regional, national, international etc.), size of undertaken investments in innovation, outlay for R&D activity etc. It should be remembered that in case of evaluation, this process is closely connected not only with the language of examined objects, broader socio-economic reality but also with the “language” of the intervention, programme itself.

The fifth step is formulating the problem area which is developing detailed survey questions (what information is available at the moment of designing evaluation, what are information needs of evaluators

9 See E. Babbie, Badania społeczne w praktyce, translated into Polish by W. Betkiewicz et al., Wydawnictwo Naukowe PWN, Warszawa 2003, p. 147.

10 Examples of questions used in ex-post evaluation of a project from the area of labour market:
1. What was the level of achieving programme objectives as far as planned products, results and impact are concerned?
2. Has the creaming effect not occurred – whether persons who were recruited for the programme were the ones really in need of such support?
3. How did the professional situation of beneficiaries change after completion of participation in the project?
4. Did participation in the project influence the increase of the participants’ adaptation abilities on the labour market?
5. What caused problems in recruiting beneficiaries included in a particular category (e.g. graduates, long-term unemployed)? Was it influenced by the regional specifics or unrealistic definition of operational objectives maladjusted to the region needs or to any other factors?
6. Can the change of professional situation of beneficiaries be considered permanent or temporary?
7. To what extent do the obtained effects persist after the end of financing?
8. Are there any important differences in the sustainability of the programme between voivodeships and if there are, what are the reasons for it?
9. Are there any important differences in the sustainability of the programme between sub-projects and if so, what are they?

and of the commissioning entity and whether these needs are convergent; whether we refer to the entire logic of an intervention – objectives, processes, effects, impacts etc.) and formulating hypotheses which we would verify (preceded by potential research diagnosis). It is one of the most critical moments of conceptualisation. On the one hand it is the identification of information needs of various groups of evaluation recipients, however, on the other hand – their systemisation in relation to the key evaluation criteria.

Operationalisation – evaluation design

After specifying in detail the study objective, object and general scope of evaluation we start the phase of operationalisation that is the activities which must be performed in order to get to know the examined intervention/programme.

Creation of specific research procedures covers the following areas:

• Selecting and defining indicators and variables whose values allow measuring the occurrence of phenomena which are of interest to evaluators (e.g. results, impact of a particular intervention). Selection of indicators should, according to J. Górniak, be rather a consequence of initial identifying and conceptualising the problem or indications and not the method for their discovery. However, when creating public policies, the indicators are treated too formally and seem to have only decorative function – added as a separate part of a programme without a clear logical connection with the programme objectives. This situation often makes it impossible to conduct effective ex-post measurement of the entire process of intervention (outlays → results → outcomes → impact) and forces evaluators to a kind of recreation and reselection of indicators at the level of general objectives (indicator of global impact) and detailed objectives (indicator of direct impact). However, it should be remembered, that the indicators are not the method (neither direct nor indirect) of determining the degree of implementation of particular dimensions of an implemented programme, their task is just to enable the construction of measuring tools of higher reliability and verifiable accuracy.

• Indicating a community in which the study will be implemented – so-called study population (who or what we want to examine). The basic population in evaluation study is usually a group of beneficiaries or entities involved in designing, managing and/or implementing a particular intervention/programme. Unfortunately, slightly more seldom evaluators in ex-post studies cover groups not directly connected with a specific intervention such as potential programme recipients, entities which applied for support in a given area but their applications were refused, control groups,

---

11 More about this issue in the article by M. Szalaj, Mechanisms of selecting assessment criteria for ex-post evaluation [Mechanizmy doboru kryteriów oceny w ewaluacji ex-post].


other institutions implementing interventions of similar nature etc. Determining final population is not always a simple task. The evaluator in this process may encounter a number of limitations, for example due to incomplete list of entities and/or their incomplete characteristics (e.g. individual characteristics such as enterprise size, sector of activity or information concerning support, amount of requested/obtained support, amount of public support obtained within last three years), their availability (e.g. ceasing operation by entities, circulation of personnel in state institutions, quality of information systems, etc.). The evaluator may also decide to limit on purpose the study population due to some variables, e.g. select only the beneficiaries who use the support for the first time or, on the contrary, have used it many times (e.g. entrepreneurs who in a way incorporate obtaining public financial means in development strategy of the company) or beneficiaries who used all elements of the programme (e.g. unemployed persons who used counselling, training or internships). Only after performing such correction tasks of a set, decision concerning covering the whole population with the study or conducting the study on a particular sample as well as selection of study methods and techniques may be taken.

- Choice of study methods and techniques, source bases and decisions concerning the principle of analysis of obtained empirical material. According to C. Robson, evaluation is not a new or different research strategy but a study which serves implementation of a particular objective (assessment of the outcomes of effects). Therefore, it uses study methods and techniques developed in social sciences, which are quantitative and qualitative methods. A number of conditions which should be taken into account while constructing a particular evaluation study are linked to the method selection. The criteria of such selection are for example: 1) characteristics of a given public intervention, 2) objective of an evaluation study, 3) scope and nature of searched information, 4) moment of evaluation, 5) commissioning entity and 6) entity implementing the study, 7) subject-object perspective (potential recipients, beneficiaries of intervention, other participants), 8) sources, quality and availability of data and also 9) time, financial, human resources, which do not directly determine the quality of collected data but determine indirectly the possibility of implementing the study. All decisions taken in relation to the selection of study methods and techniques will directly influence the quality of final results of the evaluation.

- Selection of assessment methods is a phase of project construction which is a kind of concept and which, at the level of study implementation in the phase of analyses and assessment, is always verified and completed with additional methods necessary to conduct full analysis and verification of assumed hypotheses as well as estimating the effects and specifying cause and effect relations. The choice of method at this phase is particularly important, mainly due to such methods as experts panel, benchmarking, micro- and macro-economic models since they require preparation and providing back-up in advance (e.g. in case of an experts panel it is necessary to choose good experts, develop principles of generating products; implementing benchmarking requires choosing and inviting units for comparison, and also development of adequate methods of data collecting; in micro- and macro-economic models it is necessary to develop a model and determine required data included in it, etc.).

---

16 More about sample selection, see E. Babbie, *Badania społeczne...,* op. cit., p. 200–244.
17 C. Robson, *Projektowanie ewaluacji,* op. cit., p. 147.
18 The second part of this publication was devoted to this issue.
The last two areas of operationalisation are the following:

- Estimating resources needed for project implementation. The estimation has to include 1) time resources - among others, works schedule should be specified: whether the study will not fall in the holiday period, how long the field implementation of the study will take, how much time is required for analyses, whether the results of subsequent phases of study be able to complement themselves (e.g. impact of the qualitative study on the construction of tools in the questionnaire surveys), 2) human resources, both on the side of the contractor (members of the team) and on the side of the commissioning entity (project manager and possibly an auxiliary expert, members of the evaluation group etc.), 3) financial resources – the study budget determined by the study size, choice of study methods, number of experts, etc, or also, probably much more often, determining the framework of the project.

- The decision about the entity performing evaluation and, in case of public procurement, mode of selecting the contractor (according to the Public Procurement Law\(^{20}\)). Irrespective of this mode, while performing selection not only the following issues should be taken into consideration: 1) technical capabilities to implement the project, 2) experience in implementation of projects of similar nature and 3) particular study techniques but also 4) experience in studying a particular thematic field of evaluation and 5) members of the expert team. It is worth mentioning that these are the members of the team who are responsible for the quality and reliability of evaluation as well as relevance and utility of recommendations. Therefore for many commissioning entities it is the team of experts proposed by the contractor which is crucial for the quality of the offer. As an example, according to J. Szlachta, within the field of evaluation of structural funds, a good evaluator should possess competences in three areas: 1) methods, techniques and tools of evaluation, 2) various policies of the European Union, including mainly the cohesion policy, 3) professional knowledge concerning particular areas of intervention and economic policy of the State such as: human capital, entrepreneurship including support of small and medium enterprises, knowledge-based economy and information society, technical and social infrastructure, regional policy etc.\(^{21}\)

Finally, it is worth mentioning that every evaluation project should be founded on several key principles: usefulness, feasibility and methodological correctness of evaluation and also on the basis of principles of research ethics\(^{22}\). Only fulfilling of these conditions may open the road to its implementation.

**Implementation of evaluation study**

Implementation of evaluation study may be considered from the perspective of a commissioning entity, of a contractor and from the level of particular phases of implementation: 1) structuring, 2) observation, 3) analysis, 4) assessment, 5) reporting the results (including conclusions and recommendations from the assessment) and making them public.

---

\(^{20}\) See Public Procurement Law of 29 January 2004 (Dz. U. of 2006 No 164, item 1163, No 170, item 1217 and No 227, item 1658 and of 2007 No 64, item 427 and No 82, item 560).

\(^{21}\) J. Szlachta, *System ewaluacji funduszy strukturalnych w Polsce*, in S. Mazur (ed.), *Ewaluacja funduszy strukturalnych...*, op. cit., p. 86.

The structuring phase includes development of the evaluation project within the following fields: selection and detailed specification of criteria and elements which are planned to be subject to evaluation; selection of indicators with which a given phenomenon would be examined; methodology of evaluation study; selection and development of observation tools; determining the method of reaching the study respondents. At this phase a number of methods of evaluation design are used, e.g. SWOT analysis, logical matrix, impact maps, metaplan, feasibility study, consultations with partners, logical models. This phase is the foundation for further activities within an evaluation project. At this stage, both the commissioning entity and evaluators make the most mistakes. The most common are the following ones:

23 For more about this issue, see: European Commission, Evaluating..., op. cit.
• schematic selection of study methods,
• too broad evaluation project – long list of survey issues and absence of prioritisation of study areas,
• underestimating financial or human resources on the side of the commissioning entity or contractor,
• too short time planned for reliable preparation of the project, becoming acquainted with the subject by the team, development of study tools,
• no flexibility of the commissioning entity (in case of public procurement – limitations resulting from Terms of Reference and contractor’s offer),
• limitations in communication or no cooperation between the contractor/commissioning entity/steering group for evaluation,
• no clear strategy of evaluation (for whom is evaluation conducted? what is its objective? how to use the results? etc.),
• incorrect or incomplete identification of required sources of existing data – mainly databases but also programme documentation, monitoring data, previous evaluation results, official statistical data, expertises etc.,
• no clear division of competences and tasks between team members.

Observation phase covers collecting data required for analysis – existing data (programme documentation, monitoring data, previous evaluation results, research within a particular field, statistical data, expertises etc.) and data collected on the basis of field works. In this phase, qualitative and quantitative methods are applied\(^4\). The observation phase is a very important stage for taking particular care of the quality and reliability of data. The most common mistakes, except for formal errors, resulting from incorrect application of techniques or absence of reliability of implementation, are the following:
• no diversification of observation tools (not taking into account the principle of triangulation) or limiting the study exclusively to desk research analysis, selection of quantitative or qualitative methods only,
• flooding with data (too elaborated research tools, collecting data which will not be included in the analyses or contextual data which are not very useful from the point of view of an examined object).

The analysis phase covers: 1) interpreting collected data, their comparison and identifying differences, finding irregularities of ongoing processes and basic meanings and structures, 2) verification of hypotheses, 3) cause-effect analysis, and 4) estimating effects of a programme with the use of statistical methods, qualitative analysis, economic models, comparison groups, panels of experts etc. In research practice, usually a short period of time is planned for this phase or in case of delays this time is consumed by the previous phases. It results often in the situation where the analyses are superficial, single-dimensional, remaining at the level of the so-called "still frame" – description of the initial state without deepened interpretation. Except for the above, the following mistakes are also common:
• too much concentration of the evaluators on details without identifying the general picture of the results,
• no verification of hypotheses,
• no analysis of causes and effects,
• absence of estimating the programme effects,
• remaining at the level of gross effect in the analyses.

\(^4\) See the second part of this publication.
The assessment phase includes assessment of the programme effects in relation to previously formulated evaluation questions and on the basis of accepted assessment criteria. The assessment construction should be based on the assessment scheme which on the one hand helps the team in its reliable implementation, on the other hand - helps the recipients to understand its process of implementation and grounds for formulated judgements. The methods applied at this stage are for example: cost and benefits analysis, cost-effectiveness analysis, multi-criteria analysis, panel of experts, benchmarking.

Reliable and credible data and detailed analysis should always be the foundation of assessment. What is more, the assessment should be justified which means that it should be apt and stem from the facts. Both, the evaluators and commissioning entities should care about its impartiality, objectivity and, most of all, transparency. In addition to omitting the above principles, the most common mistakes at this stage are the following:

- evaluation of the programme effects without referring to the previously formulated evaluation questions or accepted assessment criteria,
- no strong bases for assessment – justification, founding on facts, study results, credible data,
- absence of translating the assessment results into reliable and useful conclusions and recommendations,
- no discussion/interpretation in the group of experts/recipient, due to which the assessment results may be limited only to a subjective interpretation of evaluators,
- no application of diversified tools for presenting the assessment results (adjusting results to recipients).

The last phase, often omitted in many works, covers the final product of the evaluation process – final report, its dissemination and promotion. It happens very often that this phase is not sufficiently well developed, while it is the form, scope, methods, selection of communication and distribution means of the report which are decisive for the use of evaluation results. The final report is a product which, just like any other product, requires setting and promotion and therefore, a number of methods developed in such fields as public relations or marketing may be applied at this stage. Without doubt, writing a good report is an art, but it is even more difficult to write it and present it in an approachable form and reaching the largest possible recipients group with the results. Therefore there is a necessity for the contractor and commissioning entity to cooperate and search for the best possible forms of presentation and communication channels. Today, it is not enough to organise one meeting during which the contractor would present the most important conclusions and recommendations, placing the report on Internet sites of an institution or sending the report through e-mail. Every day we are flooded with thousands of information items and we have to perform selection, therefore the evaluator’s task is not only to produce a good survey; it is equally important to condense and present information (conclusions, recommendations) so that they are interesting, understandable, clear and most of all – useful for recipients. Unfortunately, contrary to research projects implemented for commercial companies, this sphere of evaluation is still underdeveloped and this is why many valuable projects remain unnoticed, and this not only because of the environment external to a particular intervention but also – maybe too often – because of the environment inside a given institution or project. As a result, it is often the final phase which is crucial for the fulfillment of the actual functions of evaluation, that is for the potential success or failure of our activity.

The above look at evaluation as a process and its potential critical points may be complemented with an additional point of view - its phases, nature of actions, products, particular types of participants and recipients involved at various phases of implementing the project (see table 1).
## Table 1. Life cycle of evaluation and its key elements

<table>
<thead>
<tr>
<th>Phases of project life</th>
<th>Conceptualisation</th>
<th>Operationalisation</th>
<th>Project Implementation</th>
<th>Promotion of Assessment Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structuring</strong></td>
<td>initiating defining appointing project manager/entity responsible for implementation*  commissioning evaluation</td>
<td>specifying project structure planning project implementation resource planning</td>
<td>organising project implementation by an internal/external team, commissioning the study, contractor selection</td>
<td>implementation control coordination</td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Presentation of assessment results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phases of project life</strong></td>
<td><strong>Character of activities</strong></td>
<td><strong>Products</strong></td>
<td><strong>Direct participants/ key recipients</strong></td>
<td></td>
</tr>
<tr>
<td>initiating defining appointing project manager/entity responsible for implementation*  commissioning evaluation</td>
<td>conceptual planning organisational executive control coordination</td>
<td>evaluation design expertise, analyses, auxiliary works Terms of Reference or scope of contractor’s works contractor’s offer contract with contractor</td>
<td>commissioning entity project manager management of institutions involved in administering/ implementing initiative/programme/ project group for evaluation experts project manager project team experts commissioning entity project manager with project team potential contractors - bidders contractor</td>
<td></td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* in case of evaluations implemented externally.

Conclusion

Presumably, many theoreticians and practitioners would agree that the quality of evaluation, its final product and use of results is, to an equal extent, within responsibilities of evaluators and entities commissioning evaluation. The quality of the final product and potential success or failure of the project depends on cooperation, involvement and awareness of importance of particular steps in the evaluation process. Obviously, this involvement will not be the same in all phases. In specific phases of the process, the roles and tasks of the main actors would be different. Therefore, at the beginning of the stage of initiating the project, specifying evaluation objective and conceptualisation, the main role is to be played by commissioning entity, alone or accompanied with a group of experts for evaluation. Only after, an evaluator steps on the scene. He or she has to develop details, operationalise the idea and in the subsequent stages perform research and analytical works. The commissioning entity is only seemingly “unnecessary” at this phase. Its involvement should have supportive, coordination and control character. Only appearance of the final report activates it entirely until the end of the project. However, it is worth mentioning that, both the roles and particular tasks should be clearly specified, because there may be a situation (intended, due to pressures from the commissioning entity, or entirely unintended for example as a result of the evaluator’s sympathies for the commissioning entity’s team) in which the evaluator loses his or her assessment objectivity or the commissioning entity takes the evaluator’s role or tasks, which will always lead to prejudice the final evaluation.

We hope that the above discussion of evaluation in the context of many perspectives will help the reader to understand well its complicated process as well as implementation of the best solutions. Certainly this article will not cover all issues and may be only an introduction to developing the most relevant, individual research strategies and further discussion which would contribute to development of both the evaluation theory and practice.

Agnieszka Haber – is a graduate of the Sociology Institute and the Institute of Applied Social Science at the University of Warsaw. She also completed a postgraduate programme in project management at the Warsaw School of Economics. Since 2004 she has been managing the Evaluation Section in the Polish Agency for Enterprise Development. She is responsible for the implementation of evaluation of pre-accession and structural programmes in the areas of SMEs development, development of human resources and infrastructural projects implemented by the Agency. She has been working on issues of social research for over 9 years, and on evaluation and management of public interventions for 6 years. Since 2005 she has been organising a cyclical Evaluation Conference under the auspices of PAED and MRD. Since 2005, she has been a member of evaluation groups for subsequent programmes: ESF, OP IE, HC OP, OP DEP. She has been lecturing and conducting workshops on the evaluation of programmes of social and economic character for 4 years. She is the author of a number of publications and papers on the evaluation of public sector programmes.

Sources


Leszek Korporowicz

Interactive aspects of evaluation process: between analysis and animation of social change

There are many ways to show developmental possibilities that are present in the realisation of a well planned and used evaluation study. This chapter aims at showing interactive aspects of the whole undertaking – little noticed but having an immense meaning for the success of evaluation. They are primarily listed as elements of the so-called evaluation instruments, even though their meaning is much more important, and it is a proof of other dimensions of the broadly understood evaluation process. In order to picture them, the first part of the text includes a number of theses that identify the proper perspective to view evaluation interactions, treated as social relations, which allows for a sociological, axiological and anthropological meta-reflection of this integrated process. The second part of the chapter outlines the genesis and period in the history of evaluation studies that gave birth to the need of analytical examination of quality, importance and goal of evaluation interactions, in relation to successful attempts to socialise them and put them into the conscious initiating of change of the subject and environment of evaluation. The third part presents natural consequences of perceiving the qualities of interactive approach to study process and taking them into account in contemporary evaluation strategies of formative character. Finally, we brought up the problem of new challenges that evaluation faces in the knowledge based society evolving from the advanced information based society. In this context, social interactions take new forms of indirect electronic communication, set out new standards of social relations and model the real shape of new type of evaluation. Communication strategies applied within that paradigm will be defining evaluation functions and require understanding of new applications, perspectives and problems, even though evaluation, as a form of knowledge, is shaping the character of modern knowledge based society itself and becomes one of its identifiers.

Instruments as interaction, strategy and communiqué

Evaluation is usually perceived as a type of social study, undertaken in order to establish the value of a given goal-oriented action, from a point of view of certain criteria. It is worth stressing that they do not need to be limited just to effectiveness or efficiency of the evaluation. The evaluation study always has its genesis, it stems out of a more or less conscious development and research needs of an organisation, it is included in a real shape of its organisational and evaluation culture, it engages the stakeholders, and finally it becomes a part of a decision-making process, it becomes a part of a public discourse.

Everything that has its influence on the instruments of the evaluation should be perceived in a broader perspective of an evaluation process as a social process with its own change dynamics and axiology, taking place in a dense network of interactions between its individual and institutional participants. Analytical approach to the issue of evaluation instruments is about relating them to those
interactions. They create the necessity to perceive them as an integral element of the same instruments portfolio, and that means - also the professional competence of the evaluation team. In order to make it more approachable, one should set out a number of well-defined theses that connect the issues of evaluation research methodologies, their realisation and use, which is usually referred to as the evaluation instruments portfolio, with the crucial phase of planning, a sometimes long discourse that sets out their goals, more or less perceived social, educational, communication, promotional functions, or those most important - developmental ones.

Firstly: Activities and competencies that define evaluation instruments portfolio of a given study are a constant and integral component of a long and multiphase process. It’s set out not only the character of study procedures, but also it makes them a functional component of complex relations within the context of a given social evaluation project environment. Those important phases of the evaluation process are:

a) identifying the object and social subjects of the process,

b) negotiating and then selection of goals and functions of an evaluation study in its individual phases and for given participants,

c) selecting a model - existing or devised for the needs of a given evaluation process – adjusted to the expected outcome and its mode of use.

These activities are currently stressed to such a degree that they become an almost indispensable part of the instrument portfolio itself, which – within evaluation strategies aimed at the animation of their democratic and developmental functions – became a sort of social axiology of many activities that are implemented to stimulate the positive social change.

Other components of the evaluation process that are decisive for the key instrument skills are: the design of evaluation project as precise as possible, including - apart from the definition of the object of evaluation – also basic, key questions, evaluation criteria, study methods, methods of sample selection, as well as process management elements. It is supplemented by the phases of generating and collecting data, then its analysis and interpretation, reporting activities, communicating the reports - i.e. various forms of presentation, and finally the implementation and applicative phases.

Secondly: Revealing interactive aspects of evaluation process brings about the necessity to define the role of evaluation actors, then the rules of their interaction: they can be authoritarian, if they are defined only by the commissioning entity, technocratic if they are set by the evaluating team or populist, if the dominating role is played by the evaluation subjects and when they define the selection of the criteria, database and interpretation. Such interaction form is a degeneration of a process of socialising the evaluation and is not a form of its democratisation which has to be based on an efficient dialogue of all participants.

Thirdly: Interactive perspective of evaluation process shows the dynamic character of meaning that is given by its participants both to the aims and individual components, actions, notions, evaluations and interpretations of given behaviours, statements or even data. In this respect, the whole tradition of symbolic interactionism is extremely inspiring, as it sheds light on social constructs of accepted meanings, their re-evaluation that is performed both by the evaluators and by the participants of

---

evaluated “discoveries”, change of positions of key participants of a process as a result of its dynamics. This perspective forces us to observe the differences that come to light when we penetrate closer the world of values of the participants, the changes that take place in the process as a result of change of positions of other persons, as well as the poses, games and mystifications undertaken to shape the image of one's actions. Evaluation process interactivity thus shows itself not only in the analysis of data collected, but above all in all inside relations and in the context of actions performed directly.

Fourthly: The instruments portfolio used in evaluation studies always takes the format of a communiqué sent to its participants, regardless of evaluation goals declared. It is, in this respect, a separate interactive action that defines its real aim, regardless also of projected functions. Interaction between the actors of evaluation scene eventually define the controlling/reporting or supporting/developmental effect of the evaluation. This will not be changed even by the formal definitions of study methods that are “soft” or “hard” depending on the goal and use.

This communiqué can realize a range of interactive strategies closer to the form and practice described as “inter-action” or “re-action”. The first one underlines reciprocity, feedback and exchange (inter), as well as agency, subjectivity and openness (action), the second one – reproduction and duplication (re), as well as passiveness and apparent character of the action. Such are ritual forms of evaluation, executed out of administrative and procedural necessity, when we bring the communicative aspects of evaluation - also outside of a narrow definition of instruments portfolio - to the level of non-creative, technocratic messages.

Interactive aspects of evaluation process were not noticed in evaluation theory and practice until the late 1960s, when they started to be perceived more broadly as a type of social actions, sometimes even as social intervention entangled in particular contextual framework, and specific set of evaluation goal, both revealed and hidden ones.

**Interactive history of evaluation studies**

Almost all sources devoted to the history of evaluation refer to Ralph Tyler as the father of this concept. In 1929, he launched a systematic evaluation studies in the Ohio State University. From almost the very beginning they were oriented to achieve measurable results of analyses, and treated almost like a synonym to measurement procedures. The effect of these evaluation studies were perceived as a final procedure encompassing all actions that are to capture current/actual “state of affairs”, detached from its history and interpretation, detached from the research process, as if these actions - as well as the process itself - existed outside the evaluation instruments portfolio.

The role of the early evaluation study was to set out, as clearly and operationally as possible, the goals of evaluation program that would be translated into the desired behaviours or skills of its participants, and then to define the level of their realisation by comparing the intentions with actual results. The whole sequence of reaching them, the changes that took place on the way as well as interactive context of the dynamism was deemed irrelevant. Moreover, such understanding of evaluation aims very often resulted in boiling down the whole problem to the recording of change of only individual behaviours of single subjects. This brought about marginalisation of supra-individual effects of a program, not to mention the analysis of its institutional, organisational or social dimensions.

---

Thus, it was not without reason that at the end of 1960s the issue of behavioural reductionism was raised - it eliminates differently defined aims from the area of evaluation study, such as for example the dimension of social experience of the program participants, their role in defining the importance of skills gained, the change in the value system etc. Still, it should be strongly stressed that already in 1930s Tyler himself was one of the initiators of a bold broadening of the functions and aims of evaluation, as he had noticed many questions and simplifications that arose at that time, such as the role of class community and the role of the educational program in shaping the skill and knowledge level of pupils.

Therefore, the correction of evaluation strategy and thinking, as postulated by Tyler, did not find the recognition it deserved due to a dynamic (and unilateral) development of psychometric, sociometric and didactic measurement techniques. Such reductionism in evaluation, especially in the field of education, but also in the evaluation of many social programs might be unfortunately seen also in the contemporary phase of development of many Polish evaluation programs and trainings. This statement finds its proof if we take into account the surprisingly similar elements of social situation that created the need for evaluation both in the US and in Poland in 1990s.

Firstly, it had its roots in the need to curtail the corruption in decision-making community, due to the lack of objective, publicly-controlled and truly transparent criteria for awarding licenses, permits or granting the rights to undertake business or public activity. Secondly, in many areas of life, especially in education, there was an immense diversity of locally accepted standards that could hardly be deemed comparable. This by itself complicated the implementation and management of a national educational policy, and prevented many wider selection, examination and certification procedures that require unified criteria for procedure and analysis. Thirdly, there was a quite popular belief that the American educational programs are not adjusted to a dynamically developing modern industrial society, which was also one of the reasons for the Polish educational reform.

It is interesting to note that Tyler, standing at the head of the evaluation studies that were developing in the above conditions, was also able to perceive the limitations hidden therein. He recommended relating the evaluated programs to the context, to the specific conditions of groups realizing them and to the intentions and experiences of the realising team. Undoubtedly these were the first postulates to develop a social, interactive evaluation base, later dubbed as “democratisation” of evaluation studies.

However, even though the achievements of Tyler in evaluation studies were widely recognised and well documented, the fact of almost identical serious educational study, led by Florian Znaniecki around that time, which actually were of evaluation character both due to their background and aims, is almost totally forgotten. Florian Znaniecki was appointed a director of a research program at Columbia University between the years 1930-33. The report of that research was lying idly for decades in the University archives, even though it set foundations for a totally different approach to evaluation³. Contrary to Tyler, Znaniecki was starting from very different methodological assumptions, addressing the social experience of the studied in a way that is well known to humanist sociologists as a study strategy taking into account the postulate of “humanist factor”. It makes us look at the studied reality not only through the eyes of the researcher, but also with the eyes of the studied, and through meanings that he sets up in a given context of his life as well as the reality of the research process.

This postulate, already fully embedded in contemporary sociological research, is unfortunately still absent in most of evaluation studies. This actually interactive research postulate, extremely important in the phase of assuming the evaluation criteria, is being forgotten. Too often the criteria are set as externally accepted standards of a given type of actions: administrative, professional, technological, and often they are a result of possibility to translate them into the language of measurement. Znaniecki was in this respect far ahead of his contemporaries, putting forward theses that will appear on the grounds of evaluation studies only in 1980s, during the rise of the issue of evaluation exclusion - and not in a straightforward sense of underrepresentation in a study sample of certain evaluated groups, but in a sense of “melting in”, exclusion or over-interpretation of their vision of studied experiences, relations, key references, value concentration centres, emotions or interest.

The externally accepted assessment standards are created by unification of indicators that – for the sake of comparison - must often be intersubjectively or even interinstitutionally unified. Such strategy works well in evaluation of large scale programs, especially those that are under international supervision and coordinated with a close control of execution. However, it causes major havoc if it is related to culturally different conditions of realisation, when the subject of evaluation is a socially detached sphere of actions assessed by equally detached criteria that are not subject to social negotiation and which, in the evaluation study, mean something different than they mean in the experience of the evaluated organisation and its employees.

An interactive type of evaluation studies is thus based not so much on the parametrical, but rather descriptive and interpretative way of studying the phenomena dubbed by Zaniecki “cultural facts”, separating them from “natural facts”. His analysis concentrated on the adequacy of education in social and cultural competencies of young Americans in the context of totally new challenges, resulting from a rapidly modernising society and national structures. Znaniecki valuated educational processes, taking into account cultural change, change in social order, transformation and requirements that those changes evoked in the characteristics of social identity, type of cultural competence necessary in the process of efficient participation in the social, economic and scientific life. The analysis that he put forward broadened considerably the spectrum of evaluation criteria and located this type of analysis on various areas of research. Simultaneously it included the shifting character of social values that forms a natural base for all the processes and valuation procedures.

A very important difference in understanding the object of evaluation appeared at the end of 1960s. It was when an important inertia of evaluation models oriented towards one-time capturing of relations between the expenditures and effects of evaluated programs started to be noticed. Dominant approaches of that time did not encompass analysis of dynamics, or even a “dramaturgy” incorporated in implementation of evaluated programs. Dynamisation of the evaluation studies stemmed both from the need to broaden the “variables” that were taken into account and the scope of data recorded, but also from the need of capturing the phenomena that take place during the program studied, which can be an important result of it, left unnoticed in the case of study of only end results. Moreover, there is no reason to exclude everything that takes place during the program realisation and what has its place in the experience of its participants, and in a way that requires a considerable broadening of the range of methods used in evaluation.

The following works played a major role in changing the perception of the classical evaluation models: *Course Improvement Through Evaluation* by Lee J. Chronbach, 1963, *The Methodology of Evaluation* by Michael Scriven, 1967, and *The Countenance of Educational Evaluation* by Robert Stake, also published in 1967. They led to:
a clear formulation of a postulate to broaden the scope of data taken into account in evaluation studies, with ever more popular referencing to qualitative data,
• moving away from a final outcome-focused model towards more process-oriented approach,
• perceiving this process in its complexity, entanglement and context, in reference to social environment,
• a clear focus on the role of social interactions in the dynamics of the research process as constitutive for understanding its functions,
• avoiding alienation of evaluation actions resulting from incompatibility with the needs of the interested institutional and individual entities,
• taking account of the needs of evaluation analyses of holistic, multi-aspect character.

Similar stratifications, growing dilemmas and discussions that appeared in the evaluation study development in 1960s have further deepened in the beginning of 1970s. Most of them are still not concluded. They base on the dispute about the eligible scope of interests, functions, methods and applications of evaluation studies. The beginning of 1970s marked a very interesting debate on the balance of achievements and failures of evaluation programs realized often for huge sums of money on a relatively large scale. Already at that time the allegations were raised of technocratic character of evaluation, neglecting the ethical aspect, too instrumental role of evaluation in managerial procedures, socially arbitrary creation of evaluation criteria, lack of perception of evaluation process in its animating functions of social changes directed towards socialization and participatory evaluation that stimulates the process of organisational learning, development of public discourse skills and embedding into the norms of modern knowledge based society.

A conference organised in December 1972 by Churchill College in Cambridge, attended by British, American and Swedish evaluators, as well as representatives of many institutions and funds, brought about almost a revolution in perception of function and social role of evaluation. The most complete expression of re-evaluation and serious criticism that traditional evaluation models met with, as well as a kind of manifesto, was a work of M. Parlett and D. Hamiton Evaluation as Illumination. A New Approach to the Study of Innovatory Programmes, presented at the conference The authors, as well as many discussion participants, have in fact taken a lot of administrative power from the evaluators, at the same time strengthening the meaning of their role in the development of understanding of own actions as well as those of other evaluation interactions participants. The postulates were formulated, encouraging the following:
• dialogue with various groups and participants of the evaluation process,
• “illuminating” the complexity of human experiences that take place during the program and reconstructing their meaning,
• understanding analysis of the organisational, psychological and cultural aspects of the evaluated actions and the evaluation as such,
• communicativeness and interactivity with all parties and subjects of evaluation already at all stages of evaluation process (from planning phase through the stage when evaluation results are used).

This laid down fundamentals for dialogic evaluation which social rationality reaches far beyond the measurement evaluation of 1930s. Michael Q. Patton, one of the leading contemporary evaluators states that an evaluation model thus created moves away from the hypothetical-deductive methodology.

---

Interactive aspects of evaluation process: between analysis and animation of social change

Towards the “paradigm of incessant choices, stressing the multitude of methods, alternative approaches and... embedding them in the context of specific situations and evaluation requirements.” In this way, the evaluation practice is less following ready-made and adamant rules, than it is about flexible interaction with the needs, opportunities and challenges of any given community. Such interactive strategy of evaluation was further developed and used by Robert Stake, Egon G. Guba and Yvonne S. Lincoln, Helen Simons, evaluators who followed the heritage of the British Centre for Applied Research in Education in the University of East Anglia (Norwich), directed for many years by one of the founding fathers of the “democratic evaluation”, Berry MacDonald. It is a tradition corresponding with qualitative interpretative studies in social sciences, developed since 1970s, that show the importance of symbolic culture in various spheres of human activity, including the research activity. If we were to sketch a genealogical tree of interactive idea of evaluation studies, even though it does not stem from the classical theory of symbolic interactionism of the so-called Chicago school of George H. Mead, William I. Thomas, Robert Park in 1930s, as well as the second Chicago school after the 2nd World War, consisting of Herbert Blumer, Anselm Strauss and many other researchers, loosely linked with each other, the trees would most certainly not touch each other in many places, theoretically enriching evaluation, but at the same time increasing the social tangibility of interactionism.

The activity and achievements of this group of evaluators and research centres established at that time turn evaluation instrument portfolio - and evaluation as a such – domain stretching far beyond the ad-hoc social diagnosis. It is exactly the unveiling of dynamic, interactive components of the evaluation process that has made it an element of a broader social policy, for which it is not only a tool, but a strategy and stimulation of a pro-development thinking within the boundaries of organisational cultures, in inter-institutional relations, as well as in the sphere of intercultural programs that require a very dialogic thinking about the values and valuation, and that rise the bar very high for the understanding of interactive evaluation.

**Dialogic values of formative evaluation**

An especially important role of evaluation interaction that require from evaluators the skill to lead a dialogue and social negotiations, is present in the formative evaluation models – i.e. those that are embedded in the process of realisation of evaluated programs and those that brings results long before evaluation research is finished. Formative evaluation requires however a relatively advanced research and organisational skills, resulting from:

- constant dialogue with the program managers,
- careful observation of the evaluation process by constant reference to its basic goals, accepted criteria, the correctness of rules for sample selection,
- correction of research instruments,
- effective communication within the evaluating team, especially during observed conflicts, problems, dysfunctions, changes and modifications,
- effective communication with the actors of the evaluation scene in various phases of program realisation.

---

Discussion and incessant verification of criteria as important elements of formative evaluation and its democratising function show a crucial dimension of contemporary evaluation that is ignored by technocratic approaches. This is because questions about the core goals, sense, and users most adequate forms and techniques of evaluation appear constantly. These questions reveal the multitude of standpoints, conflicts of interests, varying visions of its use. In technocratic evaluation, the answers to these questions are given in a narrow group of the so-called evaluation specialists and decision-makers. At the same time, democratic visions of evaluation underline this critical moment of the evaluation process. It creates an opportunity to use the design and realisation phases of the evaluation to spur a social discourse between interested parties, revealing of their points of view, presentation of their needs and, indirectly, unveiling a number of problems that should be taken into account in the evaluation or outside the area of its functioning.

This discourse has to be counted among the results of evaluation, which overthrows a popular cliché that forces us to see the evaluation report as the only proper product of evaluation procedures. This dispute is encompassed by their indirect goals; often it allows for breaking of many communication barriers and can become an impulse for the change in organisational culture. For these reasons, it is so often that the initiators of the democratic evaluation have stressed its potential role in change management, in initiating the transformation of organisational consciousness, including consciousness of goals and forms of action. Evaluators must treat any interactions that appear at that time as an important part of the strategy and of the instruments portfolio as well. They often become not only data analysts, but also animators of mental and communicational change; they transform the attitude to the transparency of actions and criteria of its assessment.

Each phase of the evaluation process has its social dynamics, its actors, and is taking place in a specific context. Each phase generates also its own ethical problems. They are a part of a broadly understood evaluation culture, and at the same time they set out a number of norms that should apply to all participants of the evaluation process. Taking them into account is a part of social authorisation of evaluation activities and a source of many misunderstandings, even conflicts, disparate expectations, various perceptions of roles, and – in consequence – failures of many evaluation projects. Therefore also, especially in the conditions of poor evaluation culture and ignorance of its functions, the key importance is attributed to negotiation of goals, process and conditions of evaluation realisation. The explanation of these issues – which certainly should be initiated consciously by the evaluators – does not only define mutual expectations, but also builds norms that apply to both parties, legitimises procedures and evaluation actions, allows to overcome restrictive perceptions of evaluation and opens communication possibilities.

Interactive understanding of the evaluation process is thus well located in the well-known tradition of interpretative evaluation studies initiated, as it was mentioned above, by Florian Znaniecki in 1930s, then refurbished, developed and embedded in the reality of the contemporary era by the fourth generation evaluators. What separates such studies from symbolic interactionism is their noticeably applied character and tendency to undertake practical challenges which directly derives from the very identity of the evaluation studies undertaken not only with one, but with three goals in mind: practical correction of a physically undertaken task, not excluding instrumental aims of such studies, in order to understand the real value and dynamism of it and to free the its potential.

---

Evaluation interactions in knowledge based society

Absolutely new tasks appear both before the instruments portfolio and the whole evaluation process in the time of rapid development of communication and information technologies of the information based society and its advanced version – the knowledge based society. It is a “type of social organisation in which the creation and use of knowledge becomes a strategic factor of its functioning and development, and the knowledge itself a key, dynamically developing resource that decides on the civilisational shape of the living conditions, value system and the ways to participate in symbolic culture and social life”.

All macrostructural, economic and social causes of this development, such as the necessity for optimization, coordination and qualitative development of activities decide simultaneously about the necessary changes in forms, techniques and goals of evaluation studies. The indispensability of recognising such changes will be ever more visible via the rising need to ask not only about the object and method of realisation of evaluated activities, but - most importantly - about the goal, a type of socially accepted value and its cultural sense. Even though the logic of evaluation process phases and the ways of participation in it does not change, there will be significant modifications in types of interactions, shaping new competencies of evaluators. The general change springs up from the capability of mediated interactions that broaden the range of process participants, under the condition however that they have required communication skills. Access to the evaluees might be theoretically easier, in practice however it can be regulated by many factors that are not always known to us and predictable - such as the skill level for using network communication techniques, or the knowledge of the op–rational techniques of IT communication systems. All this generates hidden lines of social division, selection and marginalisation, which also applies to socialised informational resources available at various stages of information collection and processing.

An important interactive aspect of network type of participation in the evaluation process is the possibility, necessity and, at the same time, the problems of use of already existing informational resources, important for setting goals, key questions, criteria, study sample, method of data verification and other important evaluation actions. “Network society” consists of multiplied systems of multilevel interactions, information flows, finding channels of mutual influence, supplementation, as well as verification, reference and comparisons. It is the network society that makes the interaction a new quality of contemporary civilisation, forces us to broaden and enrich the scope of perception of the world, other people, actions and their results.

Evaluators of the information- and knowledge based society are actually animators of such flows, they direct them in order to generate knowledge and then introduce them into the area of their organisational experience. However, there is a new ingredient of their competence - it is the necessary skill in the information and communication technology that generates the flow channels, their efficacy, volume,
accessibility, but also purpose, accuracy and dependability. Immense information resources are becoming, on one hand, a very rich source that can be used by the evaluation team commencing the design of goal, function and mode of operation, on the other hand it may introduce a certain informational chaos the management of which will require a well-developed skill of managing information for the need of the project, realisation and using evaluation studies. It is worth to note that all the activities related to the data collection, transfer and processing are to lead towards a qualitatively different knowledge about the values of the program evaluated which is one of the (even though not the sole) important products of the evaluation process. The way in which information is processed into knowledge depends on the selected evaluation model, its functions and social embeddedness, on its internal capabilities to generate social experiences of development, learning and becoming an impulse for transformation.

Evaluation becomes one of the identifiers of the knowledge based society, as it is the knowledge that became a strategic resource decisive of capabilities and directions of development, and evaluation leads to gaining a certain type of knowledge. Evaluation, as knowledge is its stimulator and sometimes a very important condition. Due to the fact that the knowledge based society is something different from the knowledge based economy, new meanings start to apply to interactions acquired via information and communication technologies – both the interactions between process participants, between them and the context of their activity, inside the evaluation team, and finally among evaluation recipients for whom it is to be an accessible and inspiring source of knowledge and thought. Knowledge based society will increase the intensity and scope of various flows, relations and communications, at the same time increasingly demanding their selection, channelling - and this means the identification of their values, both for broader social goals and for the quality of evaluations performed within.

Some very important characteristics of the modern knowledge based society, ones that would certainly set new evaluation standards, are listed by Lech W. Zacher in his very analytical work Transformacje społeczeństw: od informacji do wiedzy (Transformations of societies – from information to knowledge): He writes that “… knowledge based society – within the frames of a permanent pan- and self-education – should place within their educational systems the “packages of new knowledge” that would serve its strengthening and development. Most possibly, it will not be easy to define what exactly will be included in this new knowledge, but still examples of new areas of knowledge can be supplied (.):

- systemic approach, holism;
- complexity theory (large systems);
- chaos theory (or rather concepts of chaos);
- network theory (of any type and range);
- risk and uncertainty analysis;
- idea, methods and procedures of the impact assessment and its types, such as the technology assessment, environmental impact assessment, social impact assessment etc;
- multidimensional perspective of reality: multi, inter- and trans-disciplinary approach;
- modelling, simulation, forecasting...;
- method of searching for synergies of processes, actions etc.”

Putting aside at the moment the issue of relation between evaluation and assessment which should not, for many reasons, be treated as equivalents, the above characteristics shows to what considerable extent the modern thinking of evaluation is detached from the notion of simple mirroring of reality. This becomes even more blatant in the context of traits that characterize the transfer from information based society into the knowledge based society, such as:

---

Interactive aspects of evaluation process: between analysis and animation of social change

- ability of objective, multi-criteria assessment,
- ability to include and calculate many variables,
- ability to alternate analysis and synthesis,
- associative and parallel thinking, ability to "connect issues" or identify the conjugates,
- ability to think about the future,
- ability to use intuition and imagination as useful tools of inquiry, opening new horizons.\(^\text{13}\)

From the evaluation point of view also the growing importance of *symbolic culture* in knowledge based societies is crucial. Its meaning is discovered gradually with the intensification of various phenomena of multicultural aspect of the personnel and the "cultural capital" of the organisation, the importance of intercultural communication processes in interpersonal, inter-group and inter-organisational dimension.

Another dimension important for evaluation process is the self-realising, non-instrumental dimension of knowledge as a value of post-modern characteristics of the work culture, which has a direct influence on the evaluation criteria, social perception of evaluation interactions as a method of knowledge collection and valuation, and then the means to share the previously collected and verified knowledge. Challenges that the evaluation faces in the society of selective information processing, i.e. in a systematically constructed knowledge based society, can be presented in a form of a matrix (see Table 1).

In knowledge based society, evaluation becomes a functional element, characteristic for the general increase of importance of the reflexive and analytical approach to many activities. Interactive implication of this change is a decisive step towards many methods of information exchange, interpretation and processing, both in a direct and mediated forms, down to the creation and sharing of knowledge thus created. The knowledge itself is of increasingly dialogic and developmental character, moving away from the notion of an unambiguous prophetic entity and static picture towards a more dynamic process that animates the processes of learning, the exchange of meaning and manifesting their goal.

<table>
<thead>
<tr>
<th>Manifesting characteristics of the knowledge based society</th>
<th>Appropriate evaluation characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technological characteristics:</strong></td>
<td></td>
</tr>
<tr>
<td>High level of use of advanced communication and information technologies in which the information collected almost immediately become a part of messages</td>
<td>Use of electronic communication and information technologies becomes one of the basic characteristics of the instruments portfolio and professional competencies of an evaluator</td>
</tr>
<tr>
<td>Wide access and constant update of information resources due to an increasing freedom of network flow of information</td>
<td>All phases and actions of evaluation process planning phase, as well as purely research-oriented action make use of and analyse the consequences of a network flow of information.</td>
</tr>
<tr>
<td>Progressing potential for managing excessive and chaotic information by technological means for their processing, selection and arrangement</td>
<td>Fundamental meaning of clearly defined and operational criteria of evaluation as knowledge generators</td>
</tr>
</tbody>
</table>

\(^{13}\) Ibidem, p. 231.
### B. Structural characteristics

<table>
<thead>
<tr>
<th>Rational means to generate knowledge that includes many perspectives and variables, their contexts, “linking issues”, “identification of conjugates”</th>
<th>Sensitisation to multidimensional and multifunctional character of an evaluation process with various potentials of use in various phases, in relation to its various participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever wider use of knowledge in decision-making processes by modelling forecasting, simulation and future projection.</td>
<td>Increasing importance of formative strategies of evaluation process and its appearance in the organisational culture, work culture and decision-making processes of defined undertakings/activities</td>
</tr>
<tr>
<td>Dynamisation of the perception of reality, problems and solutions approached from prospective, developmental angle, rationalizing change and transformation designing</td>
<td>Increasing importance of democratic strategies of evaluation process, treated as “invitation for development” in organisation’s strategic planning</td>
</tr>
<tr>
<td>Searching for synergies of processes, linking resources, animation of added value</td>
<td>Increasing role of dialogic, participative strategies and evaluation instrument portfolio techniques in its “reflexive practice” function, as well as sensitisation to synergetic evaluation criteria</td>
</tr>
<tr>
<td>Manifesting characteristics of the knowledge based society</td>
<td>Appropriate evaluation characteristics</td>
</tr>
</tbody>
</table>

### C. Interactive characteristics

<table>
<thead>
<tr>
<th>Decentralisation of knowledge creation and dissemination centres linked with the decentralisation of power and participation in horizontal communication systems</th>
<th>An increasing role of evaluation as learning process stimulator - stimulating the skill of constant identification, interpreting and use of new sources of knowledge as a form of public discourse both by evaluators and the evaluation subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mediation of knowledge transmission and selection processes</td>
<td>Increasing role of mediated forms of communication at every phase of the evaluation process, embedded in evaluators competencies, but also in the analytical area of reflexion on the consequences of use of “electronic communication hubs”</td>
</tr>
<tr>
<td>Intercultural and transcultural perspective of valuation of the knowledge creation, sharing and use process</td>
<td>Increasing importance of intercultural evaluation</td>
</tr>
</tbody>
</table>
Summary

Evaluation process, treated as a whole, fulfils many interlinked and inter-related functions. Its aims are being realised not only as final establishments and their effective applications, but they are spread among various phases via varied interactions that take place between its participants. Their awareness and goals that are ascribed to them are an integral part of the evaluation instruments portfolio, and they also take part in defining the social axiology of evaluation that forms their base. The importance of interactive aspects of evaluation can be seen most obviously by referring to the symbolic interactionism and concept of humanist factor, well known from the works of Florian Znaniecki who realized it in his studies on the border between education sociology, social change and evaluation already in 1930s on the Columbia University, giving humanist evaluation models that are inspiring even now and that have been suppressed for many years by first behavioural approaches, and then narrowly utilitarian, measurement-oriented goals of evaluation. Still, its further history proved the importance and potential of an integral approach to research process that has shown the evaluators of 1980s the creative, executive and democratising functions of evaluation interactions as an instrument, but also the values that animate the developmental character of change in the working culture, awareness, quality and social purposefulness of actions realized, moving beyond the control and reporting goals of evaluation. Contemporarily, as well as in the future, new challenges are being put before evaluation by the advanced information based society and the modern knowledge based society. General access to information, as well as informational chaos, the freedom of flow of information, mediation indirect acquisition and dependency on the information and communication technologies create a new reality for evaluation instruments portfolio, the need to find out about the new competencies of evaluators, new capabilities, but also threats. Evaluation enters a new phase of development of open society - a society that puts knowledge as one of its basic functional and axiological resources and that generates it in the conditions of global informational network, new systems of mediated information. These systems imbue the information and knowledge with fundamental importance in the creation of decisions in almost all areas of life, stimulate innovative, synergetic, reflexive, projecting, self-realisation oriented and cooperation-oriented approaches, open to values and coexistence of various cultures in dynamic processes of intercultural communication. The latter call for a especially detailed researching of anthropological base for evaluation interaction to an extent proportional to their internationality, permeation of national and cultural borders, and this means also touching upon diversity and change of the world of values. The identification of those values by the evaluators not only in the society, but also in their own working systems, in updated competencies, approaches and own instruments portfolio, will depend on their sensitivity and imagination.

Leszek Korporowicz, hab. Ph.D, is a professor at Jagiellonian University (Cracow), pro-rector of the Bogdan Jański Academy; sociologist, and founding member and the first President of the Polish Evaluation Society. Former member and currently an expert of the National Accreditation Commission, initiator and consultant of many studies, postgraduate courses, including the postgraduate course in Social Programs Evaluation in the Warsaw University Institute of Sociology, organiser and coordinator of international evaluation training seminar, for 15 years an expert of many ministerial committees on evaluation, including the Ministry of Regional Development. He is an author of numerous publications on methodology and social function of evaluation, especially perceived as social change and democratization process.
Sources


Part II

Methods of Evaluation
In recent years, along with the entrance of policy programmes financed by the European Union funds to the political arena, the issue of programme evaluation has turned up on the agenda. Evaluation was first applied because of external requirements made by European Union institutions. It gradually becomes a more evident element of the policy programme cycle. Introduction of the performance budget in Poland, has led to a radical increase of evaluation’s significance, which, in the context of public finance, is often manifested as the performance audit. In the context of changes in the Polish system of public finance, the evaluation of programmes co-financed by structural funds is a true testing ground, from which experiences should be quickly moved to the whole “battlefield” in a form of all units of the public finance sector. Therefore, these considerations are the part of a wide trend of preparations for changing the model of planning, implementation and evaluation of public expenditure in a manner oriented at achieving social effects and not only compliance with rules.

Experiences with evaluation in Poland resulted in a number of publications, which among other things, considered the extent of the evaluation and its functions (Górniak, 2007; Ferry & Olejniczak, 2008; Korporowicz, 2008; Olejniczak, 2008). This is a Polish rebound from the international debate on evaluation, or broader: on the analysis of public programmes and policies.

Understanding of the evaluation coming from these considerations is diverse, which results from different theoretical and methodological perspectives, as well as from arranging accents among functions performed by evaluation? In this paper, we assume a rather wide and pragmatic understanding of evaluation as the undertaking aimed at defining, on the basis of processed information, to what extent a given solution (e.g. public intervention: policy, programme or project) fulfils established criteria, including particularly, to what extent it achieved objectives for which it was undertaken and what are the relations between inputs, actions and results of this solution (Górniak, 2007). Evaluation is an integral element of the public policies cycle. Classically, it was the closing element of this cycle, which gave possibility of analysis, summary and assessment of the implemented programme. Currently, the term evaluation is also used for assessing the programme before deciding on its implementation (ex-ante) and during the programme implementation (on-going, mid-term), where there is still a possibility of modifications in terms of the strategic orientation and the mode of implementation management. Evaluation undoubtedly

---

1 It is probably possible to make subtle distinctions between evaluation and performance audit, but it is easy to notice the prevalence of their similarities in objectives and methods over differences, especially regarding mainstream of both approaches. There are many trends in the evaluation theory, but the approach of the European Commission to the evaluation, reflected in various guides and publications, is undoubtedly related to performance audit methodologies.
has an important function in the process of good governance; it is used for accountability of policy
makers and for the effectiveness and efficiency of these programmes complimented by a good system
of meta-analysis toward results. Also, is used for collecting knowledge on the conditions of success or
failure of public interventions and creates a base for a learning process in institutions responsible for
these interventions. In order that this knowledge is significant and useful, it should identify very well all
important consequences of implemented public interventions.

A Nobel Prize winner, Professor J. Heckman, points out that evaluation of public policies is faced with
three main problems:

1. “Evaluating the impact of historical interventions on outcomes including their impact in terms of the
   well-being of the treated and society at large.”
2. “Forecasting the impacts (constructing counterfactual states) of interventions implemented in one
   environment in other environments, including their impacts in terms of well-being.”
3. “Forecasting the impacts of interventions (constructing counterfactual states associated with
   interventions) never historically experienced to various environments, including their impacts in
   terms of well-being.” (Heckman, 2008, p. 8; see also: Heckman, 2005b).

Each of these key evaluation issues is related to the answer to questions on the mechanism of causal
relations involved in the implemented intervention. Until this time, we can handle the first of these
issues in a rather satisfactory means; this can be applied to some extent for solving the second problem.
The third one puts us before particularly difficult challenges and requires having theoretical concepts
included in possibly complete structural models of causal dependencies in the programme. Some
representatives of social science, e.g. economists, who represent the Austrian School, *inter alia*, guru of
this school, Ludwig von Mises (1966), reject the possibility of prediction in social science. Although their
arguments are valid in relation to attempts of predicting the course of history, they do not apply in the
same extent to the area, which Popper (Popper, 1949) called “piecemeal social engineering” - introduction
of concrete, precisely directed changes which improve social institutions, with readiness to correct
them further in case they fail. Every action is based on prediction of its consequences. It applies to the
same degree to the public sector as well as to the activity of enterprises on the market. Striving for the
improvement of the effectiveness is a natural aspect of planned action. The assessment of the actual
effects of actions previously undertaken and creating realistic concepts of future programmes based on
causal dependencies identified to date between undertaken actions and achieved effects is used for this
purpose.

This article points out a specific aspect of evaluation, which remains out of focus in a situation when
problems with measuring effects and even the appropriate definition of objectives in public programmes
are still common: examination of causal relations between undertaken actions, which are reflected in the
intervention outputs, and effects in the form of results and impacts of the programme. In the article,
I present arguments in favour of this aspect of evaluation as well as some problems related with the
examination of causal relations in public programmes. I also point out the significance of the examination

---

2 According to J. Heckman, for obtaining the correct answer, also the previous two areas require construction of
complete structural models based on theory. It is disputed whether it is really necessary (Heckman, 2005b; Heckman,
2005a; Sobel, 2005).

3 Application of terms: output, result and impact is adopted here according to slightly modified convention
presented in methodological documents of the European Commission; modification pertains to the term result,
the scope of which is broadened not only to instant but also to postponed (permanent) changes on the level of
direct beneficiaries, which are defined by the EC as local impacts (Górniak & Keler, 2008b).
of programmes’ consequences (understood in causal terms) in the situation of poorly defined objectives, with which we have to deal very often in practice.

**Why is it necessary to examine causal relations between public interventions and effects?**

The effects of public programmes (policies) - successes in their implementation - are connected with many factors (see a classic work in this respect by Sabatier & Mazmanian, 1980). Those include certainly: the clarity and precision of objectives, ensuring appropriate resources and appropriate identification of cause and effect relationships significant for the programme. This last issue is the particular subject of this article.

Identification of causal relationships is an important task of problem diagnosis, frequently neglected with bad results for programming solutions. A simple example of a diagnostic problem - without providing a conclusion: does the fact that children attend kindergarten is conducive for achieving success in school. As the indicator for success in school we can operationally assume e.g. results of the test in the sixth grade, or - after general introduction - in the third grade. The argument of the causal impact of kindergarten preparation on educational success is often brought in the debate on making the kindergarten education universal. A diagnostic research should be designed in such way, so that there is a possibility of analytical separation of the effect of passing the kindergarten education itself from the impact of various other factors which influence both, the success in school and the choice (also an opportunity of choosing) of going through the kindergarten or omitting it. Similarly, studying the impact of the size of classes on results of teaching, we cannot simply correlate results of common, comparative competence tests and the size of classes, but we have to design the research so that we can control the impact of all significant factors which decide on results of teaching and the size of classes. In diagnostic analyses, conclusions on causal impact are too often based on simple association of two variables, which as a consequence leads to basing projects on oversimplified or even wrong assumptions.

Every public intervention, regardless of its origin is a specific “theory”: it is based on a hypothesis that in given conditions, undertaking actions resulting from the intervention project will lead to achievement of the assumed results. The structure of dependencies, assumed in the project, between the adopted objective and applied means - programme instruments, is called the “programme theory”. It is completed by the set of assumptions regarding the best way for implementation management. Identification of the “programme theory” and the “implementation theory”, which dominates in the group of decision-makers and persons, who plan the intervention, allows for better understanding of the policy or the programme and its in-depth evaluation (Birckmayer & Weiss, 2000). Evaluation, particularly the final one, verifies whether the programme theory can be maintained in the light of achieved effects, or should it be dismissed. Concepts of actions, which successfully pass the evaluation, create the body of knowledge, which improves effectiveness of public policies. Indeed, a spontaneous process of verification takes place anyhow in the form of social assessment of the effects of conducted policies. Evaluation is a methodologically controlled and systematised way of such verification conducted by experts.

If we understand the task of evaluation like this - especially the ex-post one - than it must be based on the examination of causal dependencies in terms of the first task from the list of Heckman's demands.

---

4 The problem can be narrowed down to the issue of equal opportunities or precise it differently; here, the idea is not the actual diagnosis but methodological illustration.
towards evaluation. In case of ex-ante evaluation we are faced with fulfilment of at least the second demand, i.e. extrapolation of hitherto prevailing dependencies to new conditions. In many situations, due to the lack of the knowledge based on appropriately conducted evaluations we must rely on intellectual experiment supported by estimations of partial dependencies. Such intellectual experiment is based on a theoretical model, which should be operationalised in a form of structural equations system - here we are grasped by Heckman's third demand, which is the most difficult to implement. It is a vision of conducting policy which we can define as predictive governance. This term is an analogy to the name developed in the strategic enterprises sector defined as predictive enterprise. Is the predictive governance possible? In my opinion, behind the controversy between the "statistical" and the "econometric" school in the causal analysis in public programmes is the difference regarding a belief in the ability to "control the future". Despite all reservations towards such perspective we share, the examination of causal effects is possible and needed.

Evaluation certainly has a falsifying potential - of indicating ineffectiveness of interventions. It is weaker in confirmation that the achieved effects are result of a given intervention and not independent factors, and is very weak in predicting effects of new interventions. This causes certain reservation of decision makers towards evaluation. However, it must be stressed that even if the evaluation can only indicate that the programme does not have causal effect, it will allow for saving public funds, which can be spent for other purposes (which of course should be evaluated as well). However, evaluation should determine why the programme does not have an effect, because the ineffectiveness of intervention itself does not mean that the problem, for which the programme was undertaken, can be abandoned. Lack of effects may result from the wrong "implementation theory" and not from defective "programme theory".

A very popular expectation towards evaluation is included in the title of a well known book by Patton: "Utilization-Focused Evaluation" (Patton, 1997). The author emphasizes, that "the central challenge to professional practice remains – doing evaluations that are useful and actually used!" (p. xiv). Such approach will certainly reach for quality methods and for participation, but will not resign from objective and comparative measurement as well as sound standards of research methodology, in the spirit of "multiplism", i.e. mutual completing and controlling effects of applying various quantitative and qualitative research methods. Correct identification of programme theory and determination of causal dependencies between intervention and effects is essential for good evaluation.

A pragmatic approach towards evaluation is presented by other popular authors. Pawson and Tilley emphasize in their vision of "realistic evaluation" the need for identification of causal relationships within the programme in order to obtain answers why the intervention has effects (or not), who benefits from them and in what conditions. The starting point has to be, of course, the ability to measure the effects as well as determining their causal links with undertaken activities. It is advised to use different available research modes (methodological pluralism) for identifying the structure, people and their opinions, which setup conditions for the programme (including opinions on "what does really work"). It is necessary to determine, what causes the programme to work in the given social, economic, geographical and environmental context. What are the regularities which occur in programme implementation? And finally it is important to determine the change caused by the programme and to separate it from the impact of other social factors (Pawson & Tilley, 1997). While Heckman's approach is focused on the issue of operationalisation of causal concepts in the form of structural model and its estimation, the literature on realistic evaluation is focused on the issue of elaborating the adequate concept of the programme theory.
The European Commission stands clearly on the grounds of such a realistic and pragmatic approach. Methodological manuals clearly emphasize the need for determining the effect of intervention in the evaluation. Strongly emphasized is the proper measurement of the programme effects, including the problem of adequate use of indicators. Thus, the emphasis is put on quantitative methods, objective and allowing for comparison of evaluations across member states. Particularly, the need for determining the net effect is stressed - the change directly caused by the intervention, which cannot be attributed to an independent impact of other factors. One of the key tasks of evaluation, as the European Commission defines it, is to determine the causal relationship between the undertaken intervention and the observed results. Determining the causal impact of the intervention requires:

- Measuring the total (gross) effect of intervention as a general change in the scope defined by the objective (change at the level of adopted indicators).
- Separating the independent change (deadweight) from the change, which can be attributed to the impact of the intervention – calculating the net effect.

Therefore, determining the net effect requires estimation of what would have happened if the intervention was not undertaken. The manual published in 1997 by the European Commission formulates this as follows:

**When we say that certain effects were produced or caused by a programme, this means that if the programme had not been there or had been there in a different form or degree, those effects would not have occurred, or would not have occurred at the same level.** This means that it is important to have a clear idea of what would have happened without the programme. This is called the counterfactual situation.” (Nagarajan & Vanheukelen, 1997, p. 43)

Results, which would have taken place even in the situation when the public expenditure programme had not been in place (a counterfactual situation), are called the deadweight effect. Moreover, the substitution or displacement effect - situation, when programme effects regarding certain persons, groups or areas were obtained at the expense of certain persons, groups or areas - has the impact on the observation of intervention results. This type of effect is very difficult to measure. Let us add that its occurrence is more probable in the situation when the intervention breaks into area where market competition is normally functioning rather than focusing on the areas where the market fails as a regulator. The example of interventions that violates market competition; and therefore, is usually ineffective in terms of the net effect are direct subsidies for enterprises. They divert the attention of companies from competing for clients, replacing it with competing for the favour of units which grant public funds. The effect of such orientation of intervention could particularly decrease the quality of the whole offer of the sector (this may soon be the case on the market of training services; where subsidised trainings are likely to push out of the market those which clients would choose in competitive conditions).

The emphasis of determining the net effect clearly indicates that evaluation is based on causality foundation. It requires objectivism and quantitative measurement. However, the measurement of effect requires completing with good understanding of project’s mechanism, of what was conducive and what hindered its implementation. It is good to include the knowledge of the group of stakeholders, as wide as possible, as well as the knowledge available mainly through qualitative methods.

**Net effect as a measure of the causal effect of the intervention**

As if follows from the above considerations, determining the causal impact of the intervention requires:
• measuring the gross effect of intervention as a general change in the scope defined by the objective (indicators),
• separating the change independent of the undertaken intervention from the change, which can be attributed to the impact of the intervention (net effect).

Therefore, the net effect is the difference between the total effect of the programme in a group subjected to the influence, and the effect in case when a group is not subjected to the influence. Determining the net effect (measurement of the causal impact) requires measuring the size of the effect in a situation, which cannot take place in the target group, i.e., of counterfactual state5.

The net effect is the measure of the actual effect of the intervention at the level of direct beneficiaries - the result of intervention. Measurement of the net effect allows for determining the actual effectiveness of public expenditures.

Identification of cause and effect relationships and size of effects as well as their actual effectiveness allows for appropriate planning of future interventions.

The counterfactual state describes what would happen to beneficiaries if they did not receive support. Persons subjected to the intervention as well as those who were not subjected may show two types of effects: those actually observed and those we cannot observe because we cannot observe these persons in a counterfactual situation, i.e. contrary to the situation they are in. For the group subjected to the influence, we have the observed average result of the intervention: \( E(Y|W=1) \) and non-observed (latent) average result of not subjecting to the intervention: \( E(Y0|W=1) \). For the group not subjected to the treatment, we have the observed average result of the intervention: \( E(Y0|W=0) \) and non-observed (latent) average result of subjecting to the intervention: \( E(Y1|W=0) \).6

Within the framework of the analysis of counterfactual states, evaluation aims at determining the difference between the effect of intervention and the effect of the lack of intervention. Comparison between the actual situation among beneficiaries of the intervention after its introduction and the actual situation of persons not subjected to the intervention (total effect):

\[
E(Y|W=1) - E(Y0|W=0)
\]

is only an approximation of the achieved programme effect. It illustrates the change among those covered by the intervention and ignores the possible impact on the size of the measured final effect of all these factors, which had impact on the inclusion in the group of beneficiaries. These factors may have impact on selection by intervention providers as well as by beneficiaries themselves or their representatives, e.g. parents of minors (auto-selection).

Examination of the net effect should be treated as an attempt for estimating the difference between the effect of intervention in the group subjected to the intervention \( (W=1) \), and the counterfactual effect of the lack of intervention \( (Y0) \) in this group \( E(Y0|W=1) \), estimated with the observable effect \( E(Y0|W=0) \)

\[
E(Y1|W=1) - E(Y0|W=1)
\]

The key problem of evaluation is not the estimation of \( E(Y0|W=0) \) - effect of the lack of intervention in the group not subjected to the influence (factual effect), but \( E(Y0|W=1) \), i.e.: what would happen if those treated were not treated, and as a consequence, determining the net effect for beneficiaries as

5 A very good review of concepts of causal conclusions in social analyses is in the article: (Winship & Sobel, 2004).
6 Basing analysis of the net effect on the difference of average effects is characteristic for so called statistics school; J. Heckman presents different approach to analysis of causal impact of intervention on the ground of econometrics approach (Heckman, 2008).
Causal effect in evaluation

a difference between the factual effect of the intervention and the counterfactual state of its absence for this group. The better the assessment of the counterfactual state (more comprehensive model specification), the higher the internal accuracy of the examination, i.e. ability to grasp the true value of the causal effect in the group subjected to the influence.

A quite popular approach to the problem of estimating the net effect is regression analysis; wherein, the result measured by the outcome indicator is a dependant variable, and an independent variable is the fact of participation or non-participation in the programme as well as a number of control variables describing the features of examined units. Such a model would be based on observation data, which does not come from examinations based on randomized experiments but from descriptive research, such as: surveys, administration data or public statistics.

The common regression analysis, which normally allows for determining the net effect of the intervention, used in the models as a dummy variable (0/1), is problematic, due to the occurring correlation of the residual component with the variable representing the influence:

\[ Y_i = \alpha + \tau W_i + X'_i \beta + \varepsilon_i \]

Independent variable \( W \) is usually correlated with the error component \( \varepsilon \), because the same factors have impact on (auto) selection to treatment and on result (so called selection bias); moreover, persons subjected to the intervention may have features which are more conducive to success than persons from the control sample. As a result we obtain estimation of the effect of programme influence (\( \tau \)), which is biased and inconsistent (increase of the sample does not decrease the bias). In this case, the deadweight effect is not well separated and as a consequence, the assessment of the net effect is disturbed.

In the last twenty years, we noticed a great development of works on techniques for estimating the counterfactual effect for groups covered by intervention. On the one hand, it is the continuation of ideas formulated in the ninety-twenties by Jerzy Neyman and developed especially by Donald Rubin, which aim at recreating the counterfactual situation through measurement in the control group not covered by the intervention. This approach, known as the statistical approach, deals with constructing sampling methods for control samples and adjusted methods of calculating the intervention effect. At the same time, the econometric approach is being developed, the leader of which was previously quoted, James Heckman; within it, it is postulated that the analysis of the intervention effect should be based on the structural models, specified as good as possible, which explain the observed effects.

A classical approach within the statistical approach is based on randomized samples (experimental approach). A randomized experiment is traditionally treated as a perfect arrangement for measuring causal impact of the intervention. Randomisation consists in random division of the target group to the group subjected to the intervention and the control group, not subjected to the intervention; it allows for eliminating systematic impact of other factors outside the intervention. However, in case of public programmes, there are a number of organizational, legal and ethical limitations, which hinder the application of randomization. Experimental examinations are expensive and difficult. Moreover, in case of randomization, there may be phenomena which disrupt assessment of the net effect. The problem, of course, is reduced to the answer on the question, whether in case of the control group obtained through randomisation, \( E(Y_0|W=0) \) actually represents \( E(Y_0|W=1) \). We must take into account the fact, that in case

---

7 There has been a strong trend on the ground of evaluation literature since ninety-sixties, which emphasizes using of experimental and quasi-experimental approach, founded by D. Campbell (Shadish, Cook, & Campbell, 2001; Alkin, 2004).
of programmes we are not dealing with a laboratory experiment, but, in best case, with so called field experiment, stretched in time and not separated from the impact of other environmental factors on test group and the control group. In relation to this, so called “randomization bias” of estimations may occur (Heckman & Smith, 1995): the effect of the intention to participate is measured rather than the effect of participating in the programme (participants may drop out and non-participants may find substitutes).

Research on factors affecting the net effects leads to the conclusion that unbiased (causing no impact on evaluation results) sample selection is of key importance. This problem is well shown by the above attempt to apply statistical method of controlling the impact of factors when estimating the impact of intervention with the regression analysis. In relation to limitations of randomized samples, the approach was developed, popular in the field of epidemiological research, based on matching control samples according to the features of examined units. The matched samples method consists in selecting the control sample from among persons not subjected to the intervention is such manner, that each of them is most similar (identical) in terms of features which have impact on participation in the intervention to the appropriate person from the sample subjected to the intervention, with which it forms a pair.

In the case of matching an attempt in terms of all significant features, a technical problem of the lack of sufficient numbers of appropriate control cases is quickly encountered. In reply to this problem, a technique of selecting control samples was drawn up, based on so called propensity assessment – *propensity score matching* (PSM). PSM is technique which allows for selecting matched samples and avoiding the necessity of matching cases in terms of values of each variable used in constructing the sample - matching is based on the result obtained from the logistic regression model of participation in the intervention on features which determine this participation (Bryson, Dorsett, & Purdon, 2002).

The method of matched control groups, including PSM, also has its weakness. It is based on assumptions which are not always fulfilled; these assumptions are as follows:

- control sample represents counterfactual state of the sample subjected to the intervention,
- all variables which have significant impact on participation in the intervention have been taken into account,
- these variables independently influence each unit,
- participation in the intervention is a decision made by each unit regardless of decisions made by other units.

This method ignores any impact of the intervention on the general equilibrium, i.e. what is the impact of the programme on behaviours and results obtained by units which are not eligible for intervention, including displacement and substitution effects, e.g. whether the conducted trainings leading to the employment of beneficiaries did not cause that they did not reach non-beneficiaries, which they would reach otherwise. It is insignificant when the scale of projects is small in relation to the size of a given market.

Two works have been published by the PAED (Konarski & Kotnarowski, 2007; Strawiński, 2008) regarding PSM techniques and related manners of calculating the net effect, so I will not discuss them in detail in this paper. Nevertheless, it is worth to emphasize that to this date, six evaluation studies have been conducted at the PAED request, regarding determination of the net effect. The two initial studies have been delivered in 2007. One of them determined the lack of the effect in the case of subsidies directed at enterprises, while the other - a relatively weak, but existing effect of trainings directed at enterprises.

---

8 Propensity is understood as probability of being assigned to the group covered by the interventions due to possessed characteristics.
unemployed. In the case of both examinations there were many implementing problems encountered by this method. Many of them could be avoided if reporting and monitoring systems used in public administration and in the programme implementation system were planned better.

What to do when objective is unclear?

The condition for the analysis of the causal effect is to determine the expected outcomes. It is reflected in well defined objectives of the intervention. There is still much work to be done in this respect. Objectives, especially in strategies and programmes, but often also in public projects, are formulated in general terms; they are expectations regarding changes in individual areas of social life (or rather title slogans), which only to a limited degree are subjected to the impact of policies. Sometimes, the place of general objective is even taken by the general category of actions (e.g. "Development of social and spatial processes for the improvement of life quality" - quote from a document) or entries of planned actions (e.g. "promoting civil attitudes and prosocial activities"). Generally, objectives are formulated in such way that they are difficult to monitor and assess the level of implementation: they are rather general than specific, usually nonmeasurable and there are no established targets for them, they are not time bounded, they have no relation with allocated resources, which hinders assessment of their feasibility.

This situation also pertains to programmes financed by European funds, although there have been certain improvement in this respect in the new programming period. However, above all, the greatest challenge is the whole system of domestic expenditures in public finance sector. Currently, works are in progress on implementing the performance budget, which means the transition from planning and reporting activities in terms of expenditures to defining objectives, including them in plans, connecting with expenditure and reporting their implementation. Current test variants of performance budget do not look well in this respect. Apart from the small number of evaluations of projects financed by the European Union funds, there are no assessments of net effects of activities conducted for public funds in terms of the above discussed causal analysis (in case of activities financed by national funds there are no assessments at all). However, in order to conduct such analyses, first the objectives in policy documents should be defined properly.

Does the fact that objectives are practically not formulated in the classic layout of the budget and many public programmes or previous developments of the performance budget have serious weaknesses in this respect, mean, that public funds are spent without objectives (= make no sense from taxpayer perspective)? Not at all. They are spent with purpose; although, the effectiveness is another problem. These objectives are encoded in the logic of activities, which are actually implemented with public funds.

The first source of objectives should be a governmental strategy, which has support of parliament majority (there is no such clear strategy). However, the appropriate analysis may be an important way of grasping actual objectives resulting from the need of the state functioning.

In relation to problems indicated above, which occur in the scope of defining objectives of long term programmes and measures of their effects, first evaluations of currently operating programmes

---

9 I wrote about this with K. Keler when discussing functions of indicators in evaluation ([Górniak & Keler, 2007; 2008a; 2008b]). First of these articles is published in this volume.

10 Detailed analysis of performance budget in terms of the manner of formulating objectives and susceptibility to evaluation was conducted under author’s supervision in the expert opinion prepared at the request of the Ministry of Finance (Górniak, Mazur, Jelonke, Strzebońska, & Szczucka, 2008).
should be conducted in an open form - so called goal-free evaluation - GFE. This term was proposed by M. Scriven, one of the leading theorists of evaluation (Scriven, 1991a; Scriven, 1991b). It does not mean releasing programmes from the necessity of establishing objectives, but directing evaluation to all outcomes, which occurred as a result of programme implementation.

GFE is based on observation and measurement of outcomes of actually undertaken activities and examining programme beneficiaries in terms of all consequences of intervention, including satisfying important needs. We ask what programme achieved rather than what it tried to achieve. It is also necessary to simultaneously diagnose the needs. Due to releasing from the assumed objectives, it allows for wider grasp of expected and unexpected consequences of activities. It also facilitates identification of all activities undertaken with the support of institutions’ resources, also those which may be ignored when we rely on the assumed logic of the programme. It serves well the reconstruction and update of objectives and may be used complementary for goal-based evaluation. It facilitates avoiding pressures and political involvement, which tends to optimistically bias the result by avoiding indications of deficits in achieving planned objectives. It is however, a method which requires great professionalism from evaluators. The best effects are yield when both approaches are combined: goal-free and goal-based evaluation, if goal-based evaluation is feasible (i.e. objectives are well defined and measurable). It is good to simultaneously conduct both forms of evaluation, and then confront and synthesize their results. Also, GFE may help in bringing the strategy to fulfil conditions for a goal-based evaluation.

In evaluation methodology adopted by the European Commission it would mean assigning a special role to the utility criterion. It tells us to pay attention to all consequences of activities which implement the strategy, those included in the form of objectives and those not planned. These effects are referred to identified social needs. Registry of all consequences, those desirable and undesirable, included in adopted objectives, as well as those who were not expected in a documented manner, allows determining benefits of the programme, and referring registered effects to their costs yields the efficiency measure. Effectiveness (and also efficiency) in achieving assumed objectives (if they were defined in a manner fitted for evaluation) becomes a particular case of such wide presentation. This approach has a value consisting in:

- facilitation of registration and measurement of all significant effects from the perspective of defined problems;
- providing knowledge on effects, which can be realistically expected and covering them by proper indicators.

The problem, which needs solution in this approach, is the high level of professional requirements for evaluators. Such type of evaluation also assumes greater involvement (participation) on the part of units participating in various roles in planning and implementation of the programme, which is often treated in the evaluation theory as an advantage.

---

11 Criterion of utility is often misinterpreted as subjective assessment of benefits gained by beneficiaries in the programme. It happens so although the guide for conducting evaluation, prepared by the European Commission (EVALSED: The Guide) does not raise any doubts:

“The criterion of utility judges the impacts obtained by the programme in relation to broader societal and economic needs. Utility is a very particular evaluation criterion insofar as it makes no reference to the official objectives of the programme. It may be judicious to formulate a question of utility when programme objectives are badly defined or when there are many unexpected impacts. This criterion must nevertheless be used with caution to ensure that the evaluation team’s selection of important needs or issues is not too subjective. One way of safeguarding against this risk is to involve other stakeholders, and in particular, intended beneficiaries in the selection of utility questions”.

Source: (European Commission).
With the developed map of objectives, we can start to determine indicators, conduct measurement of effects and then examine causal effects in a manner discussed above. Systematic examination of causal relationships between interventions and expected (and unexpected but important) changes is a foundation for using evaluation for improving public programmes and constructing a knowledge base for creating more effective interventions, which eliminate various defects of social institutions in a process of cautious piecemeal social engineering.

Without improvement of the process of designing interventions, especially without good diagnosis and clear identification of objectives, we are faced with a situation similar to practicing science based on unfalsified hypotheses: The place of empirical evidence is taken by faith, authority and power. It follows from the experience, that one of the sources of problems with setting up proper objectives is, to some degree, avoidance of accountability, including, in case of programmes financed by the European Union funds, fear of the necessity of returning financing in case of bureaucratically executed compliance of reported results with precisely determined targets. However, equally important, if not more important reason are deficits of the process of planning programmes: diagnosis, strategic analysis, definition and agreeing on objectives and conceptualization of programme theory. There is a lack of competence, expert resources, completed methodology and good models. Situation in this respect is continuously improving, mainly due to conducted evaluation, including those which examine net effects of programmes. However, we still have a long way to go.

**Conclusion**

This article is an encouragement for the ambitious project of building a knowledge base grounded in evaluations of causal effects of public. Is this a realistic demand? Yes, although in an evolutionary perspective. The last part of this paper indicates a way for unfolding objectives from activities conducted by public administration. This is a first step, which allows for focusing the attention on effects to be achieved. Pilot projects should help in developing assumption for constructing such public information system, which would allow for less expensive collection of information necessary for conducting evaluation. Before that, we should conduct as many as possible of this type of studies and pilot implementations with structural funds assigned to evaluation. We should also take care of creating networks of good, competitive units in the country, which will be able to conduct such examinations, preferably with the support of top universities. It will have a multiplier effect at universities, which will reflect in the education of staff and students. So far, too many funds are spent on evaluations, which are often conducted hastily, with simple means, and do not contribute significantly to the sustainable quality improvement of policy making process in Poland.

**Jarosław Górniak**, Ph.D, is a professor at Jagiellonian University and Tischner European University, creator and director of the Centre for Evaluation and Analysis of Public Policies at JU, director of the Institute of Sociology of Economics, Education and Methods of Social Research in the Institute of Sociology at JU, deputy dean of Philosophy Department at JU. As a sociologist and economist, he is a specialist on analysis of public policies and evaluation, sociology of economics and organisation and methods of social research and expert on statistical data analysis. He is an experienced investigator of market and opinions, ten years ago his scientific and research interests moved towards issues of the analysis of public policies
and evaluations. He was, inter alia, an international consultant of UNDP, expert of the coordination team for ex-ante evaluation of community documents for 2007-2013 in the Ministry of Regional Development. He directed research projects on evaluation and indicative measurement of effects of public programmes (inter alia, in the scope of the labour market, social exclusion, human capital development and regional development). He is a consultant on strategic programming and evaluation, as well as on research methodology and data analysis in management and in public administration. Member of the Board in CBOS and a member of the steering committee of METRIS project (Monitoring European Trends in Social Sciences and Humanities) at European Commission DG Research.

Sources


