Evaluation in the Making

Contexts and Methods

Evaluation in the Making

Contexts and Methods

Edited by

Agnieszka Haber
Maciej Szałaj

MINISTRY
OF REGIONAL
DEVELOPMENT

PARP

Evaluation in the Making. Contexts and Methods

Agnieszka Haber, Maciej Szałaj – editors

The following publication is a selection of texts from three books previously published by the Polish Agency for Enterprise Development in the 'Evaluation' series:
I vol. 'Evaluation ex-post. Theory and Research Practice' (Agnieszka Haber – editor, PAED, 2007)
http://www.parp.gov.pl/index/more/2046
II vol. 'Environment and Craft of Evaluation' (Agnieszka Haber, Maciej Szałaj – editors, PAED, 2008)
http://www.parp.gov.pl/index/more/9658
III vol. 'Evaluation Towards the Demands of the Public Finance Sector'
(Agnieszka Haber, Maciej Szałaj – editors, PAED, 2009)
http://www.parp.gov.pl/index/more/12416

Texts for the following publication were selected by:
National Evaluation Unit (MRD) and PAED

PAED is not responsible for the personal opinions of the authors presented in this publication.

ISBN 978-83-7633-092-1

1st edition

Translation:
Contact Language Services
Ul. Marszałkowska 55/73 72a
02-516 Warszawa

Proofreading:
Elayne Fracaro-Bieniecka

Printed by:
Institute for Sustainable Technologies – National Research Institute
ul. K. Pułaskiego 6/10,
26-600 Radom

NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

Table of contents:

This is a publication devoted to evaluation - the key tool for improving the quality of administration - which is gradually becoming an integral element of evidence-based policy. This book is a selection of professional and academic considerations concerning theoretical aspects of evaluation along with its practical implementation – in particular in the analysis of EU intervention effects. The book – elaborated by the Polish Agency for Enterprise Development – fits perfectly into the mainstream works of the National Evaluation Unit at the Department of Structural Policy Coordination in the Ministry of Regional Development.

In the process of supporting the development of evaluation methodology, the National Evaluation Unit prepares manuals and guidebooks on the problems and issues relating to the evaluation of public intervention effects (e.g. infrastructural interventions). A grant competition for academic and scientific circles had been organised since 2008. It aims at the creation and development of innovative research methodology. To estimate the macroeconomic impact of the Cohesion Policy, on-going cooperation is being maintained with three independent research and development centres which prepare autonomous macroeconomic models for the administration.

Numerous trainings and information initiatives ensure the development of the evaluation potential of administration. We organise many discussion panels and conferences which help to build evaluation culture in Poland. Since 2005, an evaluation conference (www.konferencja-ewaluacja.pl) is being organised in cooperation with the Polish Agency for Enterprise Development. It is the most important initiative in this field of expertise in Poland (over 300-400 participants, both Polish and international speakers representing diverse circles). Moreover, since 2008, in cooperation with University of Warsaw, we run gratuitous post-graduate studies entitled *Academy for the evaluation of socio-economic development programmes*. This initiative – thanks to the involvement of European and world-scale experts as lecturers – gives students an opportunity to become acquainted with the most up-to-date theories on the methodology and techniques of evaluation and their practical uses in managing development policy.

I hope you will find the book interesting and inspiring in your work and that it will become an incentive to develop the methodology of evaluation studies further and to seek innovative research methods and techniques. I also hope that the papers presented herein serve as an impulse for the discussion over the most important challenges of development and practical applications of evaluation studies' methodology in managing public programmes.

Piotr Żuber, PhD
Director of Department of Structural Policy Coordination
Ministry of Regional Development

The book, which we are pleased to present, is a reflection of the debate going on in the Polish evaluation community. Our institution has always played an active role in this discussion. The Polish Agency for Enterprise Development has been involved in evaluation research for over ten years now. We were able, thus, to become one of the pioneers of this field among Polish public administration.

Publishing is one of the most important areas of our evaluation activity. The texts which you will find here are a selection of articles published under a PAED series devoted to issues linked with the area of evaluation research. The authors represent a group of researchers and scientists from leading Polish universities and institutions who have a rich experience in the evaluation of public interventions. This book has been divided into two main thematic axes dedicated to the most sensitive aspects of evaluation process – the methodological framework and socio-organisational context.

It is widely agreed that methodology is one of the most important elements of evaluation. Correct selection and skilful application of methods thereof are essential for the success of any research. PAED has always put an enormous stress on the methodological aspects of its evaluations. Over the past years, we have frequently applied innovative research approaches.

We need to particularly point here to the assessment concept based on the propensity score matching method, elaborated in cooperation with Professor Jarosław Górniak. It proved to be a solution providing particularly reliable knowledge on the effects of structural policy. The application of that method has raised a wide discussion on the accuracy of the approaches prevailing to date, and has set a specific trend in the practices of Polish evaluation practice. Other important directions for shaping our know-how were the applications of network analyses, which are currently one of the most dynamically developing areas of social science. We have also attempted to extend the scope of online methods use. The reflections accompanying these projects are presented in the chapter: „Methods of Evaluation".

Contexts of evaluation are another key theme of this book. Its popularity in the local debate results from the tension arising between the dynamic development of evaluation activity (reinforced by formal requirements of EU funds expenditure) and the weakness of institutional traditions as regards reflective public management. There is a risk that the evaluation practice will suffer from – misperceived – mimicry consisting in literal application of guidelines or codes of best practices. The problem was discussed on many occasions – also during the Evaluation Conference organised in cooperation with the Ministry of Regional Development. The texts published in the chapter "Evaluation in Organisations and Society" explore the question of optimum strategy for running evaluation studies and application of their results.

We hope that you will find this volume interesting and helpful. We are also positive that the topics discussed here will fit well into the areas of your expertise as the issues and concepts considered by the authors refer to universal challenges faced by the practitioners and recipients of evaluation studies, regardless of specific contexts of their activity.

We also strongly encourage you to share your opinions and reflections with the authors and publishers of the book. Any such comments will serve as an ideal foundation for further, international discussion.

Bożena Lublińska-Kasprzak
Chief Executive Officer
Polish Agency for Enterprise Development

Part I

Evaluation in Organisations and Society

Dariusz Zalewski

Possibilities and limitations concerning the use of evaluation outcomes by public institutions

"Diabolic exercise"

It has been explained in many ways what evaluation functions are. Therefore, there is no reason to quote well-known textbook definitions once more. The career behind this term seems to be inseparable from two phenomena; first of all – from the expansion of the public sphere and the growing significance of administration liable for the implementation of specific tasks. Second – from the disappointment by the efficiency of these tasks' implementation by administration organised according to classic rules formulated by M. Weber. Obviously, we need to remember that evaluation itself has a much longer history - it dates back to the beginnings of the 20th century, and such figures as Florian Znaniecki had their share in its development – than the critique of Weberian bureaucracy and, accordingly, the search for new organisational forms for the public administration for a more effective and efficient task execution. For the sake of completeness, it needs to be added, that evaluation researches, as a tool for evaluation of social intervention programmes executed, are used not only within public administration, but their development owes most to its specificity or imperfection.

The increased interest in evaluation researches in Poland is strongly related to the Polish accession to the European Union. The reason is banal and relates to the obligatory requirement to evaluate public programmes financed by the EU[1]. The charming appeal of evaluation is visible in almost every official EU document dedicated to this issue. The first sentence of one of the European Commission's documents is the following: "Evaluation may be regarded by some as a diabolical exercise. However, if evaluations are well conducted, and if the results of evaluations are used by decision-makers, they can contribute to improved public programmes, as well as to increased transparency, accountability and cost-effectiveness"[2].

Although evaluation has many functions and its value should be assessed through its complexity of functions and objectives set before it, the evaluation of rationality of incurred costs seems to be central for the achievement of specific objectives, due to the fact that public administration is the largest recipient of its outcomes. This idea has been well reflected in the following sentence: "Evaluation is to deliver criteria, methods and measures to assess rationality of public activities, functioning in the public sector analogue analogically to the market mechanism of the private sector"[3]. Obviously, we could argue that such

---

[1] *European Commission Initiative Sound and Efficient Management 2000* (known as SEM 2000) includes evaluation as a key element for the improvement of management culture of the Commission itself. The key innovation of SEM will be the requirement for the covering of all EU programmes with systematic evaluation, see *Evaluating UE Expenditure Programmes: A Guide*, European Commission, 01.1997, p. 7.

[2] *Evaluating UE Expenditure Programmes*…, op. cit.

[3] W.R Shadish, T.D. Cook, L.C. Leviton, *Foundations of Program Evaluation: Theories of Practice*, Sage Publications, Newbury Park, London 1991, p. 19, quoted after K. Olejniczak, *Theoretical Basis for Ex-post Evaluation*, in: A. Haber (ed.), *Ex-post Evaluation. Theory and Practice*, PARP, Warsaw, 2007, p. 19.

reduction of evaluation research functions is unjustified and comparing its role to the market mechanism is inaccurate, insofar as the state differs from the market and the logic of public administration's activities differs from the one of a private enterprise. However, the critique of public administration, inspired mostly by neo-liberal classical economics and public choice theory representatives, has led to the situation where it is the market, with its competition and striving for economic efficiency, has become the reference platform for reforms proposed for public sector management. The role of evaluation, conversly, would be answering the question whether public programmes executed are effective in the economic sense.

There is the reason to believe that expectations regarding evaluation are too high. If so, on the one hand, what are the possibilities and on the other hand, what are the limitations in using evaluation outcomes in administrative practice? The answers to these questions seem interesting in the context of the public sector's efficiency sought for, that "Holy Grail" for all reformers, as smartly put by V. Wright[4]. The analysis will concentrate upon external limitations in the possibility to use evaluation outcomes despite the awareness that its applicable value depends equally, if not more, on the evaluation process' internal factors (appropriateness of methodological assumptions, observance of research standards, researcher's competence etc.). Such selection has been dictated by popularly formulated (often contradictory) expectations towards evaluation research from various societies, in order for their outcomes to contribute to the increased efficiency of the public programmes executed.

This chapter has been structured so as to clearly present the expectations towards evaluation and its possibilities and limitations. Since the significance of evaluation may be different according to various public tasks' execution concepts, we will briefly present the proposals for changes in the public administration and potential consequences for the evaluation of social intervention programmes. It is recommended that this text inspired a discussion on evaluation possibilities and limitations, not only among academics, since they always find time to argue, but also among representatives of administration, liable for the execution of public tasks.

A few words on the administration reforms and their consequences for evaluation

The role of evaluation in the process of efficient expenditure of public funds, the extent to which it can contribute to rational allocation decisions, depend largely on who and how executes public services.

Nearly thirty years ago Milton Friedman with his wife Rose, published a book entitled *Free to Choose*, in which they accessibly presented the rules governing the rational allocation of financial resources, in regards of two pairs of variables: whose money we spend (own one or somebody else's) and who we spend it on (for ourselves or somebody else). As a result of joining those two pairs of alternatives we obtain reply matrix, consisting of four variants: own money for own purposes (1), own money for the purposes of others (2), other persons' money for own purposes (3) and other persons' money for the purposes of others (4) [5]. In criticizing the lack of allocation rationality of the public administration, the Friedmans concentrated on variants 3 and 4, proving that expenditure of public money is a premise signifying the waste of these funds, since there are no sufficient incentives to start rational strategies for acting so as to optimise benefits obtained by beneficiaries in relation to incurred public costs. The

[4] V. Wright, *Paradoxes of Administrative Reform*, in: W.J.M. Kickert, *Public Management and Administrative Reform in Western Europe*, Edward El-gar, Cheltenham 1997, p. 11.

[5] M. Friedman, R. Friedman, *Free Choice*, Wydawnictwo Panta, Sosnowiec, 1994, p. 111.

representatives of public choice theory, who base their analyses on microeconomic assumptions of neo-liberal economy, make similar statements. *Nobody spends somebody else's money as carefully as he spends his own –wrote* W. Mitchell years ago, analysing the functioning of the public sector as part of the public choice theory[6].

Comments made by the Friedmans rather did not contribute to the development of evaluation research, and their intention was to show that the public administration mainly cared for its own interests, instead of public ones. One of possible conclusions we could make out of the neo-liberal critique of the public sector's functioning would be that the outcomes of evaluation research could be helpful, if not decisive, in the discussion on the efficiency of programmes executed by public institutions. This is confirmed by the tone of various statements made by the representatives of political establishments, particularly visible for the EU politicians, who treat evaluation outcomes as the factor rationalizing the decision making process in the public sector. "Evaluation is an indispensable part of modern public sector management practise" - wrote E. Liikanen, member of the European Commission responsible for budget and administration[7].

Obviously, we must remember that evaluation is only a tool intended for the assessment of various social intervention programmes, undertaken not only by the public administration, and thus, on no account can be treated as a functional remedy for various problems, the functioning of the public sector breeds. However, if we would like to respond to such accusations, often justified, towards public institutions, or deliver the information about costs and effects of administration's functioning to the administration, it seems that we should refer to this activity's evaluation outcomes, which we can obtain as part of evaluation researches undertaken. We should however be aware not only of various limitations for the use of research results, but also of the fact, that significance of evaluation may be different depending on by whom and how the specific tasks traditionally belonging to the public sector will be executed as part of mutual bonds between the state and the market.

"Commercialisation of State" and evaluation

Proposals for radical changes in the public administration, which appeared, among others, in the 1980s, formulated mainly by the Chicago School economists and supporters of the public choice theory, referred to the rules behind the free market with its benefit maximisation logic. The common experience was that the public administration failed to solve various problems, although it had been established to solve them, and its allocation efficiency was significantly lower than that of private entities operating in the market. The justification of the maintenance of the public sector and its administration, included in the *market failure* slogan, was countered by the *government failure slogan*. Thus, comparing the public administration's activity to market enterprises (whether a right or not is another issue), neo-liberal economists proposed, first of all, deregulation of the public sector, and second, its privatisation[8].

Deregulating the public sector was mainly aiming at the reduction of state liability for solving various problems in order to extend the space for market activities. "If public deregulation is not necessary – argued George Stigler – than a lot of public sector programmes regulating the economy should be

[6] Quoted after: P. Dunleavy, *Democracy, Bureaucracy and Public Choice*, Harvester Wheatsheaf, 1991, p. 173.

[7] *Evaluating UE Expenditure Programmes*…, op. cit.

[8] J.E. Lane, *Public Sector Reform: Only Deregulation, Privatisation and Marketisation?* in: J.E. Lane (ed.), *Public Sector Reform. Rationale, Trends and Problems,* Sage Publications, London 1997, p. 3.

removed, including the agencies introducing those public regulations"[9]. In short, deregulation, followed by privatisation of public sector agencies, was to provide better competitiveness between units liable for state public functions' execution, while the evaluative function was to be provided by the introduction of market rules, with their ability to effectively solve resources allocation problems. The trouble is though, that the market or quasi-market institutions allocate their resources only where it is possible to obtain a high rate of return for the invested resources. Apart from that, there are no guarantees that competition between entities rivalling for the execution of specific public services will take place in an impartial manner. However, apart from numerous problems and criticisms regarding the Chicago School or public choice theory representatives' proposals, such as ones stating that deregulation requires new forms of regulation[10] and privatisation increases transaction costs, it is worth noting that one of the consequences of these proposals should be the decline in evaluation research significance. This role should be largely taken over by the market, with its clear economic efficiency rule. Privatisation constitutes a reversal of the state interventionism broadening process, which was one of the key premises under developing evaluation research or, in a broader sense, implementation of the social sciences' output in solving specific problems[11].

"Commercialisation of power" and evaluation

Proposals for changes in the public sector formulated by neo-liberal economists were undoubtedly a serious threat to the evaluators' interests, and to some extent, to evaluation researches themselves, as it could have transpired there was nothing to evaluate or no one to evaluate for. The reformative concept of public entrepreneurship, best-known supporters of which were, in the 1990s, D. Osborne and T. Gaebler, authors of a notorious book *Reinventing Government*, certainly does not pose such threat. The proposal itself, fitting into a broader strategy of New Public Management, was neither especially new nor original, and drew upon both the public choice theory assumptions (based on the neo-liberal model of unit rationality), and the neo-Taylorian organisation and management theory. What evidently distinguishes the public entrepreneurship concept from the public services' commercialisation concept, deriving from the Chicago School, is included in different approaches towards market forces, which were to reform the public sector. Insofar as the neo-liberal economists proposed to limit the state activity in order to extend the space for the private sector, the "managerists" proposed the inclusion of market mechanisms into administrative activity in order to create the so-called internal market within the public sector. In reference to the neo-Taylorian concept of effective management, they proposed a model of target-oriented administration, and in particular, they stressed the use of efficiency and effectiveness indicators, evaluation of programmes executed, remuneration according to obtained outcomes and increasing the quality of services provided[12].

---

[9] Ibid.

[10] G. Majone, *The rise of the regulatory state in Europe*, West European Politics, 17, p. 77–101.

[11] K.W. Frieske states that the beginnings of the involvement of social sciences representatives into "social practice" in the United States took place in the New Deal period, when the state extended the area of its interventionism, see. K.W. Frieske, *Sociology in Action. Hopes and Disappointments*, University of Warsaw, Warsaw 1990, p. 78–79.

[12] Ph. Keraudren, H. van Mierlo, *Theories of Public Management Reform and their Practical Implications*, in: T. Verheijen, D. Coombes, *Innovations in Public Management. Perspectives from East and West Europe*, Edward Elgar, Cheltenham 1998, p. 40.

In encouraging performing systematic evaluations of programmes executed by the public administration, Osborne and Gaebler indicated a number of problems that accompany the measurement of expected outcomes. Their comments, formulated on the basis of experience derived from administrative practice, were supposed to warn against certain errors to be avoided during measuring the outcomes. It is worth referring to them, without questioning their substantive reasonableness, if only to show that neither do they exploit the catalogue of possible errors likely to influence the inference stage based on evaluation outcomes, nor the reservations launched by various researchers towards evaluation.

- First of all, "there is a huge difference between measuring the process and measuring its outcomes"[13]. This remark sounds quite banal for each person who has been at least basically in the know of evaluation, but does not have to be that obvious for administration representatives. Basically, the authors were about the fact, that administration often concentrates on not measuring what should be the subject of evaluation. "The tendency to concentrate on a certain process is natural – managers measure what the organisations they are accredited to do, while in organisations driven by regulations and regimens people treat their work as a certain process, the course of which is defined by these regulations. If they carefully observe regulations and produce amounts they are expected to, this means they fulfil their tasks. They barely think of their work results and the influence it has on those served by a given agency. But even a perfectly performed work is a waste of time and money if it does not lead to the achievement of desired outcomes"[14].
- Second, "there is a huge difference between measuring performance and effectiveness"[15]. An effective achievement of the goals set should be important for each organisational unit, the authors say. Performance is not and should not be a goal towards which the organisation is heading, but only a measure for its achievement. Concentration upon performance (and its measurement) in separation from the activity's effectiveness, leads to a paradoxical situation of resource squander. "There is nothing more ridiculous than doing something you should not do at all but in a more productive way"[16]. Moreover, Osborne and Gaebler state that: "concentrating on performance also leads to the feeling of alienation among the public institution workers. When the authorities underline the cost of each labour unit, they often achieve only that the workers see their own segment, which hampers their intelligence and skills. (...) However, if their principals concentrate exclusively on performance - on how quickly they perform each labour unit - they begin to feel as if they were working on an assembly line"[17]. By the way, it is worth adding that it is clear that in their warning against negative consequences of measuring labour performance, authors go beyond neo-Taylorist tradition and refer to Organisational Humanism, marked by such names as John Dewey, Elton Mayo, Chester Bernard or Kurt Levin – inclined towards the exploration of human relations within organisations, among others, to test the relation between the satisfaction level and the level of fulfilling social and emotional needs of organisation members and the outcomes obtained by the organisation[18].

---

[13] D. Osborne, T. Gaebler, *To Govern Differently. How the Spirit of Enterpreneurship Pervades and Tranforms Public Administration*, Media Rodzina, Poznań 1992, p. 467.

[14] Ibid.

[15] Ibid., p. 468.

[16] Ibid., p. 469.

[17] Ibid.

[18] More on this topic in, among others, J.B. McKinney, L.C. Howard, *Public Administration. Balancing Power and Accountability, Praeger Publishers*, Westport 1998, p. 159–165.

- Third, "there is also an important difference between the 'programme results' and 'policy results' covering a larger scope"[19]. According to Gaebler and Osborne, what is important from the social point of view refers to the results of more broadly understood "policy"; particular programmes are only a part of. There is a persuasive example given by the authors. The institutions executing the programmes consisting in looking for jobs for social assistance clients, often calculated the number of persons for whom the jobs had been found and the amount of their pay, not taking into account the number of persons receiving an unemployment benefit, the period of collecting a benefit, etc. What were the consequences of it? "Persons directly supervising the programme might counter that they are responsible for the number of jobs found but not for the number of persons receiving a benefit and they will be right. It is the general policy in this area, i.e. rules governing who qualifies for the benefit, the relation between the benefit and the remuneration for unskilled and low-paid job and the number of jobs like that likely to be found, that influences the number of the latter. However, if the assistance department does not measure both sets of numbers, it may seem to its workers that they are doing a great job, finding jobs for their clients and thus getting rid of them, while in fact the line of clients is getting longer. This is what has happened in many states for 10 years now. (...) Such statistics underline the significance of measuring both programme results and the policy"[20].

Gaebler and Osborne also formulate a number of practical recommendations for the public administration, which should be interested in gaining valuable information from evaluation research.

- The first recommendation, to *carry out both a quantitative and a qualitative analysis*, has its own vast bibliography and there is no reason to do any additional clarifications in this matter[21].
- Secondly, *beware of easy bits amateurs*. This is about not allowing oneself to fall into a trap of easy solutions, which were to attest to the programme's effectiveness. Referring to the American public sector's practice, the authors state the following: "Service providers usually give numbers they are expected to give, even if they have to simplify their work or avoid the regulations. If they are to find jobs for a thousand persons per year, they will choose a thousand persons for whom finding a job is the easiest and then train them. (...) This is how social assistance centres worked in the first years of implementing the Act on Vocational Training Participation"[22].
- Third, *be prepared for strong opposition*, which means more or less that reliable information about performance and effectiveness may pose a serious threat to an existence of a given organisational unit. Authors give a couple of examples for such situation: the State of Florida, which liquidated the programme rewarding schools for improved work quality and the State of Arizona, which ceased to publish data concerning the rate of vocational and postgraduate schools' graduates who'd found jobs. The reason was banal: - "the college and vocational schools conceded that such information poses a threat to them, since it reveals how they actually prepare young people to taking up a job"[23].
- Four, *involve service providers and subordinates into the measurement method creation process*. This recommendation refers to models known from the organisation and management theory, such as *Organisational Humanism* referred to above, recommending cooperation to work out common solutions and avoid intra-organisational conflicts. In the words of authors, "the best way to break the

[19] D. Osborne, T. Gaebler, op. cit., p. 470.

[20] Ibid., p. 471.

[21] M. Jasiński, M. Kowalski, *False contradiction: qualitative or quantitative methodology?* in: A. Haber (ed.), *Ex-post Evaluation. Theory and Practice*, PARP, Warsaw, 2007.

[22] D. Osborne, T. Gaebler, op. cit., p. 476.

[23] Ibid., p. 477.

opposition against the introduction of methods measuring the work performance and quality is to involve service providers and all workers into the process of creating them"[24].

- Five, *submit measure methods to annual revision and modification*.
- Six, *do not use too many or too few measures*. This truly Aristotelean recommendation of temperance is based on the conviction that it is possible to develop as part of cooperative measure creation strategy. "If (…) workers and service providers are actively involved in creating measurement methods and they are allowed to regularly adjust them, they usually remain moderate in creating measures"[25].
- Seven, *beware anti-incentives*. This is about avoiding situations leading to inefficient outcomes, from the point of view of "overall policy". "Measurement methods primarily assumed in the Act on Vocational Training Participation encouraged service providers to look for easy bits. Such anti-incentives can scupper efforts leading to the creation of work performance measures. In order to avoid them, organisations should test them before introduction, trying to find all possible avoidance methods utilised by cunning service providers"[26].
- In the eighth recommendation the authors call for "concentrating on the maximum use of work performance data. The sole creation of measures does not guarantee that managers will use them to change the operating methods of institutions managed and that legislative bodies will utilise them to change directions and methods of financing programmes and institutions"[27]. In brief, it is not recommended to delay the use of available data and apply "the policy of completed facts" based on gathered information on efficiency and effectiveness of activity.
- Nine, commission the measurement to an independent, objective institution. If people are to rely on data then they must believe they are objective. Therefore it is good to make use of the independent institution's services"[28]. Implementation of this recommendation faces certain obstacles, though. The trouble is not only with the question whether the "belief " in the objectivity of delivered data, as suggested by Gaebler and Osborne, actually makes them objective. We can abandon this scientifically fundamental question of "objectivity" of reality and possibility to become familiar with it, since authors' intentions are very clear. What they were about was that they recommended that performance and effectiveness measurement of executed programmes were carried out by external organisations, independent of contractors. Objectivity of obtained outcomes' evaluation should then be treated as a function of independence and evaluating institution. This raises a problem with the so-called independence of an evaluating institution, which could deliver objective information to the administration. We could obviously say, in accordance with authors' suggestion that independence means as much as the lack of personal and equity links between the evaluation institution and programme executing organisation. But such reductionist interpretation of "institutional independence" is unreliable, insofar as it does not take into account the complexity of various points of view and interests in the public sphere. Barely are there good justifications for the situation where the so-called independent institution (be it private or public) evaluating the programme, would not take into account the contracting authority's expectations in its evaluation. Doubtlessly, taking into consideration contracting authority's viewpoint or interests should influence the final evaluation outcome, even if its outcomes are based on well-prepared research tools. The final evaluation result is

[24] Ibid.

[25] Ibid., p. 478.

[26] Ibid.

[27] Ibid., p. 479.

[28] Ibid., p. 478–479.

not only, or even not in the first place, a result of applying certain research standards, but of stipulating them in the pre-research phase.

Moreover, we could attempt to prove that decisions made in the public sphere are most often not shaped according to empirical stipulations, free of personal values and viewpoints of decision-makers, but in compliance with them. Outcomes of various evaluation researches used in discussions about penal policy, sexual education, abortion etc., serve mostly as an instrument strengthening the viewpoints or values of involved parties and the objectivity of outcomes incompliant with expressed viewpoint is usually questioned, rather than viewpoints changed. There is a lack of any premises (and empirical evidence) to believe that evaluation researches well carried out by independent institutions, e.g. concerning rehabilitation programmes for dangerous criminals, will make the capital punishment supporters abandon their views on penal issues or otherwise - will make avid supporters of capital punishment out of avid supporters of rehabilitation. In short, the problem of independence and objectivity of evaluation researches' outcomes is tangled up in the structure of interests and values defining the dynamics of decision-making processes in the public sphere.

Hopes and disappointments concerning evaluation

G. Majone, known for his scepticism towards the possibility of defining rationality of decision-making process in the public sphere according to purely empirical stipulations of evaluation researches, emphasises that "professional evaluation constitutes merely a small part of an overall public policy criticism and evaluation process, in which all politically active members of democratic community participate in different, but equally useful ways"[29]. Acknowledging the significance of evaluation, as a vast and developing branch of policy analysis, devoted to "gathering, verifying and interpreting information on the implementation and effectiveness of existing public policies and programmes", Majone states that it is not sufficient for the efficient provision of social services to prepare a theoretically optimal action plan. "It is even more important to get to know how the programme implementation actually goes, who benefits from it and who loses, whether the programme meets intended objectives and if not, how it could be improved or ceased"[30]. For example, the evaluation of programmes for the activation of the unemployed, particularly popular in the EU, does not usually include information about who benefits from them and who loses, since it is assumed in silence that increasing the number of jobs is in everybody's interest, therefore it is justified to include the costs of such training courses in citizen taxation costs.

We could argue that the controversy presented is more ostensible rather than real, since defining the "winners" and "losers" as a result of certain public programmes and policies, the level of meeting intended objectives or methods of implementing various programmes, belongs to the set of evaluation tasks. It is hard not to agree with Majone, though, who believes that: "It seems that many evaluators assume that these are purely empirical stipulations, neither linked to the choice of values, nor to personal views. In reality, values and views play a great role in evaluation, not only due to the ambiguity of outcomes in practice - difficulties with ascribing certain reasons to specific outcomes, with outcome measurement and the evaluation of intended effects, with discerning erroneous concepts and failures in executing them – but to an even larger extent due to inevitable differences in views concerning the types of evaluation

[29] G. Majone, *Evidence, Arguments and Persuasion in the Plitical Process*. Wydawnictwo Naukowe Scholar, Warsaw 2004, p. 249.

[30] Ibid., p. 248.

criteria, which are rational, justified or politically acceptable in a given situation. Such obscurities and differences in opinions cannot be solved by means of improved measurement and verification methods, but they can be presented and explained via discussion and mutual persuasion"[31].

Majone's arguments clearly point at the limited persuasive potential of evaluation researches' outcomes and, as the author suggests, it will be rather fruitless to look for ever more refined measurement methods. The trouble is that, among others, the **"output"** evaluation outcome depends on the "input" material available to evaluators and its collection is not only a matter of applying one research technique or the other, but also certain arrangements of viewpoints, values or interests between the parties interested in the programme execution and evaluation. Application of specific research standards is something completely different than stipulating them and in order to become aware of persuasive potential of obtained evaluation outcomes, it is necessary, in Majone's opinion, to discern the stipulation of standards and the application of standards"[32].

Discussing specific evaluation standards, e.g. arrangement of basic programme evaluation criteria among the interested parties, with all consequences to the final evaluation outcomes, may protect evaluators from various accusations, e.g. that they are not measuring something they ought to measure or they have not taken into account significant aspects of the programme under evaluation or the problem the solution of which was supposed to be the outcome of the programme. Such accusations may be raised particularly often when the evaluation outcomes pose a threat to more or less articulated interests of certain groups involved in the programme execution or affected by the effects of its implementation[33]. All those who implement programmes disturbing a specified interest system, also when they have empirical evidence that retaining status quo is disadvantageous for all and the proposed change is in line with Pareto progress criteria, should bear in mind those risks. However, referring to the argument that "numbers don't lie" may be as irritating and incomprehensible to one group or the other, as the past attempts of ardent Marxists to persuade "opposing" workers that they were the bearers of a "false class awareness", but this must be changed for their benefit as part of "objectified" historical processes.

Applying specific evaluation criteria e.g. efficiency and effectiveness of activity, conditions the final evaluation outcome in an obvious way. The application of specific criteria is to a large extent a consequence of social roles the actors involved in the evaluation process play. The same programme will be evaluated differently by an economist, for whom it is most important to look for efficiency and effectiveness of executed programme, and by a lawyer, for whom it is important to evaluate its compliance with the law in force. There is nothing wrong in evaluating the programmes from different points of view. It is important to hold on to diversity of various viewpoints since "as long as the opinions expressed from the perspective of one specific role are not presented and falsely interpreted as opinions referring to all possible roles or expressed on their behalf, we are dealing with a safe state of multiple or pluralistic evaluation"[34]. The only trouble is with the fact, that various viewpoints overlap each other in evaluation practice, and it is a mistake to assume a certain universality of conclusions resulting from evaluation in a

---

[31] Ibid., p. 249.

[32] Ibid., p. 249.

[33] "The key evaluators' dilemma can be very easily defined: most public policy researches underline the fact that effective reforms must be granted acceptance from those groups they concern the most. However, many of those affected by reforms are either reluctant or hostile towards them and it is no surprise, since their jobs are at peril; only the most eccentric turkeys await holidays", see V. Wrigth, *The Paradoxes of Administrative Reform*, in: W. Kickert (ed.), *Public Management and Administrative Reform in Western Europe,* Edward Elgar, Cheltenham 1997, p. 9.

[34] G. Majone, op. cit., p. 251.

sense that they should be equally significant for actors playing different social roles. "Perhaps the biggest problem is when it is assumed that conclusions from the evaluation carried out for the use of a specific role are equally significant from the perspective of other roles using different evaluative criteria. Since the roles and criteria do not match, it is almost inevitable that evaluation conclusions will turn out to be inconclusive"[35].

A trouble appears particularly when there is an emphasis placed upon a little "mythologised" efficiency and effectiveness of achieving intended objectives, while in such cases it is almost always about the so-called rationalisation of public expenditures, and in fact, costs limitation. This problem is special, since the most important role in the public programmes evaluation process is most often attributed to the efficiency and effectiveness criteria, oftentimes forgetting that increasing the two, according to an economic rationality standard, can disturb many vital interests or values shared by main actors involved in the programme. On the other hand though, not taking these interests into account may end up in a situation where striving for decision rationalisation may lead to opposite effects and significant squandering of resources (the case of Augustów ring road).

The methods to increase efficiency and effectiveness recommended in evaluation reports may also be at variance with the work culture within an organisation executing a given programme or valid professional standards. Moreover, as indicated by V. Wright, analysing paradoxical consequences of public administration reforms undertaken, evaluation may produce totally unexpected outcomes from the point of view of the so-called public expenditure rationalisation, which is the basic objective. "The fourth paradox is the fact that while cost reduction is the intention of many reforms, the final result of some of them may turn out to be completely opposite. Let us take "evaluation" as an example, this contemporary madness ignoring labour force costs. Evaluation may identify resource waste areas but also reveals differences not only in performance, but also in general outcomes achieved: some people can be treated relatively more badly. These differences were previously hidden. The result will be an exertion of political pressure to remove such differences. In line with an intention, transforming citizens into clients with rights clearly specified by objectives may influence the programmes' quality improvement, but this also creates expectations and remedy pressure (for revealed inequalities – annotation D.Z.) often by means of legal action"[36]. If Wright's remark is correct, then hopes that due to constant evaluation of public programmes or administrative actions their economic efficiency could be increased, seem to be futile, insofar as they do not take into account the complexity of interpersonal relations and multiply false convictions about a simple relation between stimulus and effect, included in conventional, mechanistic organisation and management theories.

G. Majone maintains that apart from taking into account the roles allowing discerning evaluation criteria, it is also useful to refer to three different evaluation modes: output, input and process mode[37]. An output mode seems to be particularly attractive, since the evaluation carried out on the basis of obtained outcomes appeals the most to decision-makers' imagination. We need to remind that all attempts to rationalise public expenditures and to optimize the outcomes of researches undertaken in the public sphere are based on the tendency to measure and quantify the outcomes. It is not without reason

[35] Ibid.

[36] V. Wright, *Paradoxes of Administrative Reform*, in: W.J.M. Kickert, *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham 1997, p. 10.

[37] In the output mode, evaluation concentrates on the outcomes of a specific action. In the input mode, evaluation emphasizes resources, skills and people, while in the process mode, it is transferred onto methods used to transform inputs into outcomes; see G. Majone, *Evidence, Arguments*…, op. cit., p. 253.

that the platform of reference for the public administration operation critics is the market and a private enterprise with their attributed ability to measure achieved outcomes. This assumption is logical, as well as it is in-built in methodological evaluation canon, namely in order to be able to evaluate an executed programme in the best possible way, one has to measure its outcomes. Referring to purely economic rules of evaluating social intervention programmes is troublesome, though, not only because the logic of market and state and functions they have are different. The trouble is that, among others, the outcomes are not always quantifiable, objectives are not unambiguously specified and measurement tools often fail. However, it is even more problematic that the relation between registered outcome and its reason is often doubtful and we depend more on intuition than scientific evidence.

E. Babble accurately points out that: "one of potentially most onerous aspects of evaluation research is to stipulate whether an evaluated programme was a success or a failure. The aim of the foreign language teaching programme may to help students in better language learning, but how large is the sufficient progress? The aim of the programme for visiting spouses in prison may be to boost the moods among prisoners, but how much do they need to be boosted to justify the programme? As you can expect, there are no concluding answers to such questions. This dilemma has become a basis for the cost – benefit analysis in an obvious way. Unfortunately, situations you usually deal with during evaluation researches scarcely can be submitted to a simple economic calculation[38].

It seems though, that expectations the evaluators have towards cost – benefit analysis are also over-exaggerated in a sense that its outcomes are at times undermined as biased, since the analysis itself is rather based on subjective assumptions becoming an economic model. While remembering that the cost-benefit analysis is only one of the tools used in evaluation researches, we need to be aware of a limited value of outcomes. R. Formaini, in his suggestively entitled book *The Myth of Scientific Public Policy*, persuades that there is no such thing as an "objective" cost-benefit analysis; therefore outcomes we obtain from it are not "objective", in a sense that they do not deliver certain and admissible-to-everyone knowledge needed to evaluate an executed programme[39]. Accusations launched by Formaini are of an ontological-epistemological nature and concern the assumptions of neo-classical economy, which are incorrect in his view, regarding the nature of social order, upon which the cost-benefit analysis is based. "Cost-benefit analysis and neo-classical economy are symbiotically linked to each other by the fact that the cost-benefit analysis could not be carried out without assumptions offered by neo-classical methods. When analysts measure, define and aggregate costs and benefits, they do it within the neo-classical structure of assumptions made. It is legitimate then to argue that the cost-benefit analysis is a 'child' and the neo-classical economic theory a 'father'". (...) Thus, it is necessary first to get to know the basic assumptions of neo-classical economy to understand the cost-benefit analysis"[40].

The discussion on the ontological bases under models of analysis, countering the model of an "economic human being" with a "sociological human being" and outcomes obtained as part of them is a canon of scientific inquiries. There is no reason to reconstruct this discussion here. It only needs to remind us about the fact, that what may be an outcome objectified as part of applied theoretical assumptions, on the basis of which it is possible to make rational decisions for ones, for the others may be no more than a projection of assumed motifs or attitudes, which cannot form the basis for making generalised theoretical explanations or actions undertaken in practice. The accusation towards scientific *Public Policy*

[38] E. Babbie, *Social Research in Practice*, PWN, Warsaw 2004, p. 376.

[39] R. Formaini, *The Myth of Scientific Public Policy*, Transaction Publishers, 1990, p. 39–65.

[40] Ibid., p. 39.

made by Formaini, is nothing else than an attempt to expose shortcomings of the *homo economicus* model. It could be judged as legitimate also for evaluation researches, the supporters of which often claim that if they are carried out according to a carefully prepared methodology, their outcomes are the basis for making rational decisions. Formaini's conclusion is that *"Public Policy* based on scientific pillars, this dream that had grown since the Renaissance and perhaps reached its peak towards the end of this century, is a myth, a theoretical illusion. It has existed in our minds, analyses and methods only because we are looking to discover it, but we are bound to discover what we are looking for"[41].

Conclusion

A conclusion stating that all limitations that evaluation is prone to, make its outcomes of little use in decision-making processes, particularly concerning the rationalisation of public expenditures, seems to be as illegitimate, as stating that the outcomes of carefully carried out evaluation, sticking to all tenets, provide a certain backing for the rationalising decisions in the public sphere. However, stating that "truth lies in the middle" is burdened with the same probability of error as indicating either of extreme opinions. Moreover, the dispute should not take place according to the logic of "who's right" and on the basis of what sort of arguments we are likely to agree with one or the other side of this dispute. We should rather seek consensus in order for the evaluation researches' outcomes, even if we agree they are imperfect in many ways, to constitute a specific value for the public administration responsible for making certain decisions in the public sphere.

The problem is basically nil for "decisionists", to quote G. Majone, as rational decisions should be made on the basis of information brought by a carefully carried out cost-benefit analysis. However, since opponents indicate plenty of imperfections of this approach, mostly of ontological-epistemological character, there is a question what to do with a huge amount of "produced" information as part of social intervention programmes' evaluations. If we were to accept "decisionists'" opinions, questioning the usefulness of evaluation analyses for the decision rationalising process, without any reservations, then we would have to admit, to some extent, that striving to allocation rationality based on outcomes' evaluation starts to contradict itself in a sense that wishing to rationally expend public money we do not obtain any certain knowledge, for which we still have to pay.

The proposal of the evaluation research critics, carried out according to the economic model's assumptions provides for the inclusion of evaluators' efforts into a broader democratic debate devoted to the stipulation of public programmes execution criteria and standards and their evaluation. Public administration should be one of the debate's parties, which would not blindly use information received from evaluation undertakings. In other words, the value of evaluation would not consist in allowing to making a rational decision that would maximise benefits at specified costs or minimise costs at specified benefits on its basis, but in its valuable contribution to discussion between the parties representing different viewpoints or interests. "Currently – claims Majone – it is more necessary to facilitate broad dialogue among supporters of various criteria than developing 'objective' outcome indicators, which is a traditional evaluation analyses' objective"[42].

[41] Ibid., p. 1.

[42] G. Majone, op. cit., p. 264.

If we succeeded in making use of the evaluation outcomes in a broader debate about the legitimacy of social interventions, their costs and benefits, as well as criteria according to which the intervention outcomes would be evaluated, then we could be sure that evaluation has got its contribution in the democratic process of arranging various rationales or viewpoints. In present-day democracies suffering from legitimacy deficit, such function as evaluation seems to be very attractive for at least two reasons. The first reason is that while provoking a debate, the legitimacy of a decision-making process in public administration is somehow increased; the second reason is that such debate, based on outcomes obtained in accordance with scientific standards of empirical research, brings a specific knowledge potential which can contribute to a more sublime perception of public policy shaping process, in the words of Majone. This certainly does not mean that evaluation research outcomes, the public administration should perhaps use, will free us from various manipulations regarding the information resources, since science itself is not free from such temptations, however this is a topic for a different discussion.

It is important though, to broadly discuss possibilities and limitations of the use of evaluation research outcomes among public administration representatives, since the quality of social intervention programmes executed depends on public administration. If it were to transpire that administrative system decision-makers overestimate evaluation possibilities, then evaluation outcomes would not necessarily reflect an actual value of executed programmes. The last one, in the words of "anti-decisionists" is not always calculable. Scepticism towards evaluation should not obscure its possibilities since many social intervention programmes can be successfully submitted for evaluation, at the same time delivering a lot of valuable information about various aspects of executed programmes. The discussion around evaluation should therefore take place both among academics as part of interdisciplinary or paradigmatic exchange of opinions, and among representatives of administration, among others. Due to their experience, they can contribute much more to the processes of improving evaluation researches than it seems from the perspective of academic chair. What does it mean for the process of administrative decision-making? Certainly, evaluation cannot replace common sense and experience generated in everyday course of administrative practice, often referred to as incremental. We need to bear in mind though, that following specified schemes may bring about petrification of decision-making structures, while evaluation outcomes can protect us from such danger. The Aristotelean praise of temperance seems to be a reasonable proposal the public administration should also make use of.

Dariusz Zalewski – is a Doctor of Sociology, Assistant Professor at the Institute of Sociology, the Faculty of Philosophy and Sociology of the University of Warsaw (Department of Social Problems and Social Planning) and at the Institute of Labour and Social Studies (Department of Collective Labour Relations). He has dealt with the issues of the welfare state institution, public administration and social dialogue. He is an author and co-author of such publications as: *Social Welfare and Assistance. Dynamics of institution; Utopias of inclusion. Successes and Failures of Social Reintegration Programmes; Social Dialogue at the Regional Level. Evaluating Chances for Development*. He has participated frequently in research schemes devoted to, among others, the forms of representing group interests, methods of executing tasks by public administration, social dialogue at the central and regional level. As part of his professional interests, he also participated in evaluation researches, e.g. in the evaluation of Leonardo da Vinci youth exchange programme, co-funded by the European Union.

Sources

Evaluating UE Expenditure Programmes: A Guide, European Commission, 01.1997.

Babbie E., *Social Researches in Practice*, PWN, Warsaw 2004.

Dunleavy P., *Democracy, Bureaucracy and Public Choice*, Harvester Wheatsheaf, 1991.

Formaini R., *The Myth of Scientific Public Policy*, Transaction Publishers, 1990.

Friedman M., Friedman R., *Free Choice*,. Panta, Sosnowiec, 1994.

Frieske K.W., *Sociology in Action. Hopes and Disappointments*, University of Warsaw, Warsaw 1990.

Haber A. (ed.), *Ex-post Evaluation. Theory and Practice*, PARP, Warsaw 2007.

Kickert W.J.M., *Public Management and Administrative Reform in Western Europe*, Edward El-gar, Cheltenham 1997.

Lane J.E. (ed.), *Public Sector Reform. Rationale, Trends and Problems*, Sage Publications, London 1997.

Majone G., *Evidence, Arguments and Persuasion in the Plitical Process*. Wydawnictwo Naukowe Scholar, Warsaw 2004.

Majone G., *The rise of the regulatory state in Europe*, West European Politics, 17.

McKinney J.B., Howard L.C., *Public Administration. Balancing Power and Accountability*, Praeger Publishers, Westport 1998.

Osborne D., Gaebler T., *To Govern Differently. How the Spirit of Entrepreneurship Pervades and Transforms Public Administration*, Media Rodzina, Poznań 1992.

Verheijen T., Coombes D., *Innovations in Public Management. Perspectives from East and West Europe*, Edward Elgar, Cheltenham 1998.

Agnieszka Haber, Maciej Szałaj

Social capital – the prerequisite and the effect of successful evaluation

For the last three decades, social capital has been one of the most popular notions appearing in the field of social sciences. Apart from numerous references which can be found in expert literature, the term can often be encountered in the public debate on civic activity and efficiency of public institutions. Social capital is often a kind of a "lifebelt", helpful to many journalists and other commentators of daily events when they have to explain the original cause of development divide, reform failures or pathologies in public space. It can be safely assumed that social capital has settled in colloquial language. It has also found a fertile ground in the jargon of students of sociology, political science or other, more ephemeral scientific disciplines.

The presence of reflection, focused on social capital, has been visible for some time also in the area of evaluation practice and theory. It was reflected most clearly in the subject matter of the latest European Evaluation Society Conference, which was be based on a multi-aspect approach to relations between the evaluation and various forms of social capital. In this context, this chapter is a voice in the current discussion which involves international community of evaluation practitioners and recipients.

However, undisputed "social capital mania" is not the main reason why this text is published in the present volume**.** The perspective of social capital - provided that it is applied selectively - is of great cognitive value in terms of organisational (institutional) description of the evaluation study environment. It allows to reflect some crucial but– usually ignored - conditionalities for embedding evaluation in the practice of public institutions and their social partners.

An outline of the origin of the notion of social capital will be presented in the first part of the chapter along with a variety of its definitions and applications. Also, a proposal for selective understanding will be indicated. This will be restricted to key organisational aspects of the evaluation process that have a particular importance within the context of the Polish institutional culture.

In the second part, the double role of social capital in the area of evaluation will be discussed. The significance of social capital - both as one of the preconditions for conducting successful evaluation research, and one as the outcome of properly managed evaluations – will be mentioned.

The third part will serve as a summary where significance and characteristics of social capital are outlined against other types of capitals and resources available to public institutions. It will also be indicated how particular forms of the capital can be "invested" in developing evaluation system and culture, and thus in high quality evaluation research providing useful results for a wide range of recipients.

Social capital – origin, definitions, functions

The term "social capital" was introduced for the first time in early 1970s by Pierre Bourdieu, one of the most well-known European sociologists of the previous century. Several years later, Americans took up considerations on social capital, adopting initial assumptions somewhat different than Bourdieu. Thus, two main and dominant to this day approaches have been formed. Their common elements are a direct derivative of the adopted analogy towards other, "classic", forms of capital (physical, financial) and concentrate on the ability to invest in social relations, which can generate unusual profits that no other form of capital could provide.

Bourdieu, unlike American authors, emphasised both positive and negative aspects of investing in social capital, which he defines in neutral categories as a "set of real and potential resources which are linked to possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition – or to put it differently with membership in a group - which provides each member with support in a form of capital possessed by the collective, reliability which gives them access to the credit in the widest sense of this word"[1].

Social capital, according to Bourdieu, is an attribute of individuals, but through aggregation it can strongly influence the whole of social relations, particularly the form of social structure and hierarchy. The example provided by Bourdieu was the class system, based on social capital, with institution of knighthood for securing privileged position of one of social groups.

Whereas according to the American tradition, social capital is perceived to belong to whole communities, which was reflected in positive connotations related with this term. Under this approach, social capital as a resource, available without distinction to all community members (regardless of what subgroups, coalitions or coteries they belong to) is expected to bring general social profits. James Coleman defines social capital as: "a social good (...) group of entities, having two characteristics in common: they all consist of some aspect of a social structure, and they facilitate joint actions of both individuals and institutions which remain within the structure. Just like other forms of capital, the social capital is productive – it enables achieving certain goals, which would be unattainable if this capital was absent (...)"[2].

Francis Fukuyama brought considerations over social capital further. The main point of his research is the influence of social capital on development through reduction of transaction costs and thus through increased efficiency. In his opinion: "social capital is important for efficient functioning of modern economy and sine qua non condition for stable liberal democracy"[3].

This concept - very popular today won wide renown in early 1990s as the works by Robert Putnam, the continuator of Coleman's thoughts, were published.

Putnam provided an empirical illustration of social capital functioning on the example of contemporary Italy and United States. The main theses of his works focus on the realisation that economic and legal factors are not the only ones that determine the level of socio-economic development as well as the efficiency of democratic institutions. According to Putnam's thesis, it is social capital that makes a difference in many cases. He defined it as a term referring to "such characteristics of social organisation like trust, standards and relations which can increase social efficiency by facilitating coordinated actions".

---

1 P. Bourdieu, *The Forms of Capital*, in: J.G. Richardson (ed.), *Handbook of Theory and Reseaerch for the Social education, Greenwood Pres*, New York 1986, p. 248.

2 J.S. Coleman, *Social Capital in the Creation of Human Capital*, American Journal of Sociology, no 94, p. 98.

3 F. Fukuyama, *Social capital, civil society and development,* Third World Quarterly 2001, no 1, p. 7.

Assuming this, Putnam brings the common element for all definitions of social capital, saying: "just like other forms of capital, the social capital is productive – it enables achieving certain goals, which would be unattainable if this capital was absent"[4].

The concept of social capital, understood as a unique resource of the community, was an attractive reply to cognitive challenges aroused by disintegration of bipolar balance of world powers and turbulence observed in transition countries, which - as many experts expected - should be smoothly adopting patterns of advanced Western democracies. As a result, social capital was included as heuristic device in numerous research projects commissioned by key international institutions. Their results became a part of public debate and have been quite often in policy making worldwide

Through reports bearing the logos of honourable institutions, social capital made its way to the language of manifests and programmes, covered, in time, with certain pathos. Precision of understanding the notion itself was also obliterated as were its relations with processes of vital importance for the modern world. Social capital was also defined with certain irony as "something of cure-all". In order to preserve its basic values and explain the term of social capital at the same time, it is necessary to strip the unnecessary pathos off the interpretation. This shall be done in the next part of this work.

Social capital in the processes of evaluation exercises

Connecting social capital with issues of evaluation requires great care for the precision of applied terms. The scope of interest here will be limited therefore to instances of social capital within public institutions that conduct evaluation and stakeholders who are recipients of its results. It will also be assumed that social capital constitutes a supra-individual property, thus belonging to specific communities, understood, for the purpose of these considerations, both, as personnel of particular institutions or other groups involved in creating and implementing given public undertakings, and as the society. Our understanding of social capital is marked with clearly positive valuation, although we do not deny that there are instances of the negative social capital within the scope in question (e.g. corruption mechanisms, so called "set-ups), appropriating the state, not transparent public-private relations, exclusions etc.)[5]. Assuming this allows proposing a thesis on two functions that social capital fulfils in terms of evaluation research and being at the same time both a necessary prerequisite and outcome of successful evaluation.

Social capital as prerequisite for successful evaluation

Successful evaluation should fulfil at least several criteria. First, an evaluation research should fulfil specified quality requirements in terms of accordance and credibility of the methodology applied. Second, an appropriate arrangement of evaluation process is necessary, so it is set within practices of action embedded in a given institution. Hence, the importance of social capital should be discussed by reference to the following aspects of evaluation process:

1) planning, implementation and management of evaluation process,
2) institutional perception of evaluation process outputs.

---

[4] R.D. Putnam, *Demokracja w działaniu: tradycje obywatelskie we współczesnych Włoszech [Making Democracy Work: Civic Traditions in Modern Italy]*, *Znak*, Kraków 1995.

[5] See more T. Kazimierczak, *Kapitał społeczny a rozwój społeczno-ekonomiczny – przegląd pojęć,* in: T. Kazimierczak, M. Rymsza (ed.), *Kapitał społeczny. Ekonomia społeczna*, Instytut Spraw Publicznych, Warsaw 2007, s. 54–55.

Each evaluation process creates unique connections and relations. Newly created bonds modify the existing social capital and create chances for using resources and their exchange between inter-organisational units (see Fig. 1). More or less dynamic networks of relations and contacts with many participants are created, which play a role in the strategy for evaluation projects implementation, as well as in the later stage of receiving and using its results. Networks based on cooperation[6] used in evaluation projects allow not only for quick identification of information needs, adjusted to actual needs of institutions, but also for effective planning (conceptualisation), preparation (operationalisation) and implementation of evaluation process

Fukuyama indicated that ability of an organisation to make use of development opportunities lies in the social capital, including skilful cooperation inside an organisation and within the external environment in order to implement common interests. Such ability to join into groups constitutes a considerable derivate of social capital serving in such cases as a source of trust between different parts of the organisation. This kind of trust should be regarded here also as an important and measurable economic value[7].

Within the context of the Polish administration culture evaluation, however, is often perceived as a somehow awkward practice, imposed by formal requirement related to EU structural funds and procedures.. One of the key barriers hindering institutional settlement of evaluation in this environment is the perception of evaluation projects as a control activity aimed at ensuring accountability of persons responsible for implementation of the given area of public activity[8]. Accountability, however, in the aforementioned sense, is connected with the paralysis of institution caused by distrust, which makes it impossible to take up cooperation between individuals or units directed at deriving benefits from evaluation process.

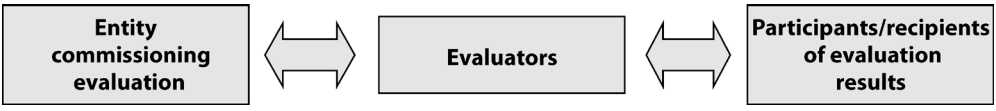


Fig. 1. Actors of the evaluation in an organisation

Source: Authors' own work.

The above problem reflects a broader phenomenon which characterises Polish public culture. Trust deficiencies constitute one of major characteristics of Polish society[9]. Literature on this subject often indicates that the quality of public institutions is closely related to the level of trust. This regularity is also observed in relation to evaluations which became doomed to fail when put in the "hostile" context of

[6] Every participant develops the network through bringing new contacts, knowledge, abilities and competence, thus further catalysing and developing organisational resources and possibilities of a given evaluation project.

[7] F. Fukuyama, *Trust: The Social Virtues and The Creation of Prosperity*, PWN, Warsaw 1997.

[8] It should be stressed that here we discuss the accountability in a meaning closer to the one already known from the public debate and not the one referring to social responsibility of public officers. Enforcing social responsibility is an immanent element of every democratic system, being also one of the original functions of the evaluation used often to this day.

[9] See more:
1) European Social Survey, http://www.ads.org.pl/pobieranie-zbioru-danych.php?id=15;
2) Kapitał intelektualny i społeczny w Polsce – raport doradców premiera, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html;
3) World Values Survey, http://www.worldvaluessurvey.org/.

distrust. In the best case they encounter serious barriers, first - within the native organisation. Because the institution which commissions the evaluation is not really interested in its results, and evaluators which conduct the study encounter a specific wall of distrust , preventing appropriate planning of the study (e.g. through getting to know the intricacies of the studied programme, diagnosing information needs of an institution). Success, within the meaning of an institution operating in this manner, would be a formal implementation of evaluation study, with minimal popularisation of its results, not to mention the practical use of recommendations following the study.

The term social capital gives an interesting analytic perspective in this case, due to stressing the importance of trust in the process of its creation, maintaining and multiplying. Using this option, we can state that social capital (especially in the form of relations based on trust) constitutes a resource necessary for conducting successful evaluations. Its absence, even when evaluators are highly competent and sufficient budget is secured, will always be a limitation impossible to overcome even by strengthening other available resources.

Weakness of trust relations is without a doubt the evidence of the absence of necessary conditions for conducting successful evaluation; however, it is not an argument for desisting from all evaluation efforts. According to rhetoric often used by advocates, the evaluation should be an "invitation to development" for public institutions and civil society institutions. Experiences of countries where evaluation is a integral part of public management, indicate that this slogan has some empirical confirmation. Their adaptation to our domestic conditions is possible if we assume that deficiencies of social capital will be supplemented which of course is a difficult but not unattainable challenge.

In order to transform social capital into both, an attribute of entities dealing with evaluation, as well as the ground for efficient evaluation, a reverse of thinking is necessary; a reverse which will teach organisations how to work on social relations in their complex environment.

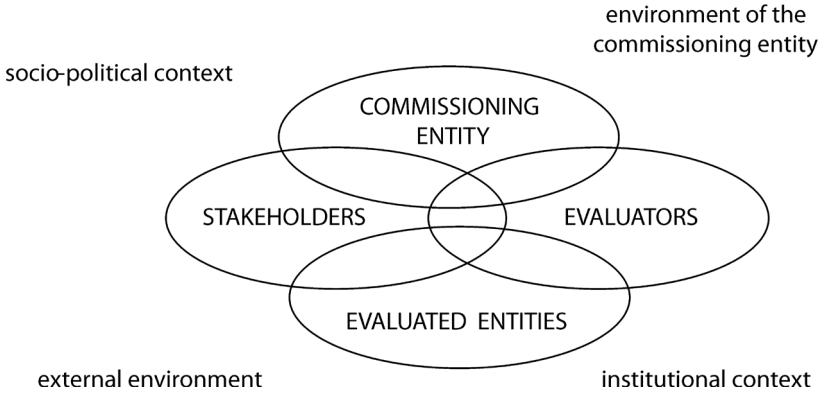


Fig. 2. Environment and context of evaluation

Source: Authors' own work.

Key factors which help organisations to understand complex environment and derive the greatest possible benefits from developed social capital should be the basis of such thinking[10]:

[10] V. Allee, *Return on Knowledge*, Executive Excellence 2000, no 9. Provo, Birmingham 2000, p. 2, in: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

- **Multidimensional view.** Organisations should be the beneficiaries of their own organisational behaviours and at the same time understand assumptions, values and ideas which support and develop those behaviours.
- **Dynamic relationships.** Creating maps of relations which take place between organisation participants allows to perceive emerging trends and to understand existing relationships, contacts and correlations in the organisation.
- **Commonly shared patterns.** Dealing with complexity is easier when participants of the organisation share the same way of thinking, behaving or cooperating.
- **Group work.** Individuals left on their own cannot comprehend the complexity or the entirety of the system. Comprehending it requires group effort. Learning in the "knowledge economy" is a process consisting in cooperation and great social capital outlays.

Accumulation of social capital, in a form which interests us, requires a long-term and systematic work on inter-institutional relations. An important role falls in this respect to evaluation units located currently within structures of most institutions in Polish administration both at the central and local level. Activities of these units should have the form of a large scale animation with the purpose of building permanent relations within institutions and with key external partners.

Formulating this in more practical categories, building trust towards evaluation should include a whole catalogue of activities which can potentially contribute to "disenchanting" evaluation, i.e. overcoming the notion of its strictly controlling, accountable character.

Collecting social capital for evaluation exercises requires taking up multidimensional activities at various levels of institutional hierarchy, inter alia, in such areas, as building the abovementioned trust and creating knowledge-based culture, cooperation, activity, innovativeness, experimenting, learning from errors and accepting failures. The simplest activities include:

- trainings/analyses concerning principles of evaluation, addressed separately to persons who manage a given institution and to other key persons who are actual or potential collaborators of the evaluation unit;
- development of competence of evaluation unit employees, in such areas as: 1) verification of reliability of external evaluators work as well as planning and methodology of evaluation exercises, 2) interpersonal and managerial competence necessary for correct management of a complex process of evaluation and animation of processes: a) creating permanent relations of trust b) communication, c) building cooperation network;
- constructing dynamic cooperation networks with scientific institutions, evaluation units and professional associations, oriented at similar groups of institutional problems or the same area of research issues;
- building dynamic cooperation networks within the institution: creating relations of trust through involving persons responsible for assessed area in processes of consulting and distribution of evaluation results;
- creating platforms for exchanging information and experience in areas of evaluation projects (conferences, meetings, professional seminars, but also Internet websites, Internet project databases etc.);
- devoting sufficient time for the exchange of information which facilitates precise formulation and receipt of feedback from each and every partner of the evaluation process (activities in this area allow for decreasing incomprehension, wrong interpretation, wrong reception of a message and strengthen cooperation).

The catalogue of activities outlined above is only a set of examples of initiatives used for building relations of trust which allow accumulation of social capital necessary for successful evaluations. The range of activities in this area is very wide and their conditional selection should be preceded every time by an in-depth analysis of specific nature of a given case. However, it should be stressed that building relations inside the institution should be seen as one of full-scale activities for creating - so frequently brought up - "evaluation capacity" of public administration. Similarity of certain activities to techniques typical for marketing practices, human resources management or public relations is not depreciating at all, provided that fidelity to professional principles is maintained. Evaluation studies do not lose their credibility then, but gain much more as regards their usefulness and adjustment to recipients' needs.

Social capital as an outcome of successful evaluation

Social capital, understood as a necessary condition for successful evaluation is an issue generally much less known and thus we have devoted more attention to it. Connecting the term social capital with evaluation is more often focused on the second role from among those mentioned before, i.e. multiplying social capital through evaluation results. It is a trend of analyses coming from the popular concept of democratic (participatory) evaluation. As indicated above, involving institutional partners in evaluation processes is a practice used for building relations of trust, necessary as a prerequisite.

In the second role we discussed, social capital is used in the environment characterised by sufficient resources of social capital at the "initial stage". Participatory democracy is a form of opening institutions to external partners through building trust, cooperation and communication. In a model interpretation, external partners are invited to actively participate in the process of evaluation.

Direct involvement of social partners in the processes of managing public undertakings - and in evaluation as a part of these processes - causes an increase in social capital through strengthening the significance of civil society institutions. As Robert Picciotto indicates, evaluation "allows overcoming asymmetry between citizens as principals and the State as an agent (...), as well as helps to solve dilemmas of collective action and develop cooperation between citizens and the authorities"[11].

The scope in which the evaluation contributes to the accumulation of social capital is related to the subject of democratic evaluation, i.e. democratic functions of evaluation. Social capital is then strengthened by providing information on effects of public policies and public spending. Including evaluation results in the public debate encourages citizens to take up collective actions aimed at ensuring their common interest. As Picciotto observes: "whenever the market is inefficient in allocating rare goods (and it basically is like that in case of public funds - authors' note), participation is related to hierarchy. In such situation, evaluation is forcing transparency and minimises «stealing a ride». Participation in evaluation is related to decentralisation and privatisation. Creation of active civil society in a global scale created a new type of evaluation, closer to a function of social advocate. (...) Openness of decision-makers to opinions of employees, clients, makes it easier to agree on premises and weak points of public programmes. Independent evaluation helps to expose the false renown of groups of interests or allows supporting their arguments"[12]. This way, through the access to information, transparency of public institutions performance and effects of public policies, active society may redefine its opinions, better understand the essence of problems and thus actively affect public policy.

[11] R. Picciotto, *Economy of Evaluation*, Państwo i Rynek 2006, no 2; see also: Intellectual and social capital in Poland – report of advisers to the prime minister: http://pliki.innowacyjnosc.gpw.pl/Kapital_Intelektualny_Polski.pdf .

[12] R. Picciotto, op. cit.

Evaluation which is deeply rooted in social capital is knowledge on knowledge. By promoting transparency and responsibility, it contributes to the increase in quality of public services and constitutes a counterbalance for highly defined interests. Nothing undermines politicians and bureaucrats' claims as much as the actual assessment of results of their actions. Thus, evaluation brings benefits in the form of improvement of expenditure efficiency and sensitivity of organisations regarding clients' needs[13].

Social capital as an attribute of building evaluation capability

The idea of social capital gains particular significance in the developing vision dynamic public administration orientated at, efficient solving of public problems, where basic, strategic resource is knowledge and the value of an institution is decided by its intellectual capital. This idea is the key to solve problems inside organisations as well as to maintain bonds with "external world". W. Dyduch, M. Szczepankiewicz, E. Szczepankiewicz indicate that similar to other forms of capital, social capital serves the purpose of increasing efficiency of functioning and development of an organisation, mainly through facilitating cooperation among participants[14].

This concept is particularly important regarding the evaluation. Mainly because on the Polish market we now have to deal with: 1) relatively young, dynamically forming market of demand/supply for evaluation studies, 2) relatively new institutional structures appointed for tasks related with the assessment of public interventions, 3) deficiencies in baseline resources (human, financial, information, structural, intellectual capital, cultural capital etc.), and thus 4) low evaluation capability.

The interpretation of human capital adopted herein comprises treating it as a component of the individuals' ability to cooperate within social groups, organisations and social institutions of various types (not only economic) aiming at the achievement of common goals. Skilful use of social capital, i.e. knowledge and abilities possessed by individuals, large or small social groups, is a source of future satisfaction, reinforcing efficiency of activities taken up by an organisation in the area of evaluation as well as a catalyst for development, innovativeness of various evaluation approaches.

In the literature on the subject it is clearly indicated that organisations with a great amount of social capital are more inclined to entrepreneurially take up risks and introduce new activities than organisations which limit spontaneous group creation. It is so because social capital is a form of social structure in an enterprise and strengthens proactive behaviour of people within this structure[15]. Understanding social capital as an involvement of individuals in creating a network of connections constitutes a premise for explaining a tendency for innovation[16]. Social connections are the reason for creating reciprocity of activities within the network, and the network, in turn, is a ground for the development of trust, increased inclination to take up risks and innovative actions. Networks facilitate coordination, communication as well as cooperation and reinforce reputation of participants, which allows for solving problems within a

---

[13] Ibidem.

[14] W. Dyduch, M. Szczepankiewicz, E. Szczepankiewicz, *Kapitał społeczny podstawą zdobywania przewagi konkurencyjnej w Nowej Ekonomii*, in: M. Moszkowicz (ed.), S*trategie i konkurencyjność przedsiębiorstw po dziesięciu latach transformacji*, Politechnika Wrocławska, Polanica-Zdrój 2001, p. 193–202.

[15] M. Bratnicki, *Dylematy i pułapki współczesnego zarządzania*, in: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

[16] P. Bullen, J. Onyx, *Measuring Social Capital in Five Communities in NSW*. Internet, materials from website http://www.mapl.com.au in: W. Dyduch, op. cit.

group[17]. Success of organisation settled in social environment depends on skilfully built network of social contacts, which supports innovative actions[18].

Orientation on instantaneous building of abilities and skills which allow the implementation of tasks set forth before the evaluation requires methodical building of evaluation resources. In current conditions of sufficient financial and material capital as well as market and institutional deficiencies regarding knowledge, information and human resources, it seems that it is possible only through[19] - participation in networks of institutional and project connections based on reciprocity, cooperation and trust based on social capital.

There is no doubt that it constitutes a foundation for benefits and efficiency of entrepreneurial organisations, inter alia, because it helps decrease costs of seeking and flow of information. Networks of social contacts decide on who is first to learn about, about such issues as methodological and technical innovations, new resources of knowledge, contractors' activities, new entities, conducted research and analyses, key experts in a given field etc. Moreover, due to the possibility to build new and efficient connections quickly, they allow to maximise the use of available resources (through the largest number of network participants), develop innovative approaches and solutions as well as decrease transaction costs (of coordination and cooperation).

Although a mixture of network, reciprocity, trust and standards is undoubtedly a chance for creating an efficient "evaluation environment", which has commonly shared and developed resources (including most important ones – information, knowledge and intellectual resources), available to every participant of the network, it is not easy to create new network connections. Mainly because organisational units are not always ready to become network participants or because they do not always want to share their resources, which can be their source of advantage.

Building sustainable relations therefore depends on the change of the way of thinking and activity of key leaders – aware animators of a number of systematic activities for the benefit of building a widely understood evaluation potential.

Conclusion

Experienced entities which commission and implement evaluation know that the success of evaluation, that is, its purposeful, reliable, efficient and effective implementation and usage depends mainly on non-material resources.

Social capital constitutes without a doubt one of the fundamental resources, necessary for conducting good evaluation on one hand and "consumption of results" on the other. However, in the light of subsequent studies on the system of social values in Poland, taking into account deficiencies – existing in such areas as trust, pro-social activity of citizens, ability and will to cooperate etc., thus the deficiency of social capital – derivation from this capital is currently highly limited for evaluation processes. This situation is a challenge not only for evaluators, public institutions and non-governmental organisations but also for the society.

---

[17] C. Sirianni, L. Friedland, *Social Capital and Civic Innovation: Learning and Capacity Building from the 1960s to the 1990s*, Internet, materials from website: http://www.cpn.org/sections/new_citizen-ship/theory.html, in: W. Dyduch, op. cit.

[18] L. Steier, R. Greenwood, *Entrepreneurship and the Evolution of Angel Financial Networks*, Organization Studies 2000, no 21, EGOS, p.165, in: W. Dyduch, op. cit.

[19] Compare: P. Bullen, J. Onyx, op. cit.

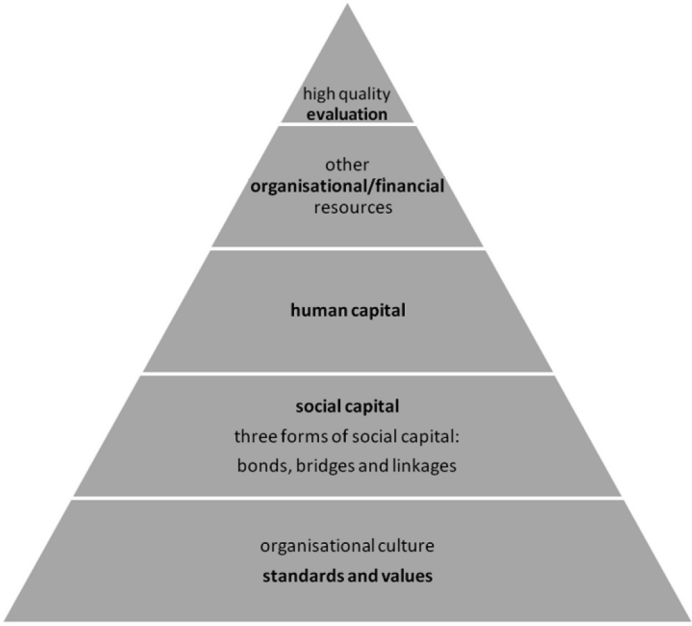


Fig. 3. Foundations of high quality evaluation

Source: Authors' own analysis.

We hope that this paper on evaluation in the context of social capital will allow the Reader to better understand complex relations and environment of evaluation projects; however we are fully aware that the above can only serve as an introduction to further discussion.

Agnieszka Haber – is a graduate of the Sociology Institute and the Institute of Applied Social Sciences 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.

Maciej Szałaj graduated from the University of Warsaw (majored in sociology) and completed a postgraduate programme in project management at the Warsaw School of Economics. He works as a senior expert in the Evaluation Section of the Polish Agency for Enterprise Development. As a manager of evaluation research projects, he specializes in interventions supporting the SME sector and in

measures aimed at developing the adaptability of enterprises. He participated in creating one of the first evaluation units in Polish public administration. He was a member of Steering Group for Evaluation of the SOP/HRD (Sectoral Operational Programme Human Resources Development). Currently, he is involved in developing evolution system at the level the OP/HC (Operational Programme Human Capital). He is an author of articles exploring various issues linked with evaluation practice and theory. He also serves as an academic teacher and public administration trainer. As a panelist, he has participated in several seminars and conferences. Since 2008, he is a board member of the Polish Evaluation Society.

Sources

Allee V., *Return on Knowledge*, Executive Excellence 2000, no 9. Provo, Birmingham 2000, in: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

Bourdieu P., *The Forms of Capital,* in: J.G. Richardson (ed.), *Handbook of Theory and Reseaerch for the Social education*, Greenwood Pres, New York 1986.

Bratnicki M., *Dylematy i pułapki współczesnego zarządzania*, in: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

Bullen P., Onyx J., *Measuring Social Capital in Five Communities in NSW Internet*, materials from site http://www.mapl.com.au, in: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html

Coleman J.S., *Social Capital in the Creation of Human Capital*, American Journal of Sociology, no 94.

Dyduch W, Szczepankiewicz M., Szczepankiewicz E., *Kapitał społeczny podstawą zdobywania przewagi konkurencyjnej w Nowej Ekonomii*, in: M. Moszkowicz (ed.), *Strategie i konkurencyjność przedsiębiorstw po dziesięciu latach transformacji*, Politechnika Wrocławska, Polanica-Zdrój 2001.European Social Survey, http://www.ads.org.pl/pobieranie-zbioru-danych.php?id=15.

Fukuyama F., *Social capital, civil society and development*, Third World Quarterly 2001, no 1.

Fukuyama F., *Trust: The Social Virtues and The Creation of Prosperity*, PWN, Warsaw 1997.

Intellectual and social capital in Poland – report of advisers to the prime minister, http://pliki.innowacyjnosc.gpw.pl/Kapital_Intelektualny_Polski.pdf

Kazimierczak T., *Kapitał społeczny a rozwój społeczno-ekonomiczny – przegląd pojęć*, in: T. Kazimierczak, M. Rymsza (ed.), *Kapitał społeczny. Ekonomia społeczna*, Instytut Spraw Publicznych, Warszawa 2007.

Picciotto R., *Ekonomia ewaluacji*, Państwo i Rynek 2006, nr 2.

Putnam R.D., *Demokracja w działaniu: Tradycje obywatelskie we współczesnych Włoszech*, Znak, Kraków 1995.

Sirianni C., Friedland L., *Social Capital and Civic Innovation: Learning and Capacity Building from the 1960s to the 1990s*, Internet, materiały ze strony: http://www.cpn.org/sections/new_ citizenship/theory.html, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

Steier L., Greenwood R., *Entrepreneurship and the Evolution of Angel Financial Networks*, Organization Studies 2000, nr 21, EGOS, w: W. Dyduch, *Kapitał społeczny organizacji pożywką dla przedsiębiorczości i innowacyjności*, http://www.zti.com.pl/instytut/pp/referaty/ref42_full.html.

World Values Survey, http://www.worldvaluessurvey.org/.

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.

---

[1] See for example Ministerstwo Rozwoju Regionalnego, *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, Warszawa 2007, http://www.funduszestrukturalne.gov.pl/NR/rdonlyres/4E41CBAE-F849-48BE-BAC9-0614753930AA/38773/PlanewaluacjiNSROwrzesien2007180907. pdf, p. 3.

[2] Compare C. Robson, *Projektowanie ewaluacji*, in L. Korporowicz (ed.), *Ewaluacja w edukacji*, p. 147–152, Oficyna Naukowa, Warszawa 1997; H. Simons, *Polityczne implikacje teorii ewaluacyjnych*, in L. Korporowicz (ed.), *Ewaluacja...*, op. cit., p. 65–92.

[3] Compare OECD, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris 2002, http://www.oecd.org/dataoecd/29/21/2754804.pdf, p. 21.

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 nature[4].

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 implications[5] (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 recipients[6] 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,
- intervention objectives,

---

[4] More about this issue in the article by K. Olejniczak, *Teoretyczne podstawy ewaluacji ex-post* in: *Ewaluacja ex-post. Teoria i praktyka badawcza*, A. Haber (ed.), PARP, Warsaw, 2007; S. Mazur (ed.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007, p. 16–23; D. Nevo, *Konceptualizacja ewaluacji edukacyjnej*, in L. Korporowicz (ed.), *Ewaluacja...*, op. cit., p. 54.

[5] See more C. Robson, *Projektowanie ewaluacji*, op. cit., p. 158; H. Simons, *Polityczne implikacje teorii ewaluacyjnych*, op. cit., p. 65–92.

[6] See table 1.

- 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

[7] See other example of developing a scope of project based of the four capital model: http://www. http://konferencja-ewaluacja.pl/2007/materialy.php (presentation by Prof. G. Gorzelak, slide 4).

[8] Compare L. Słoma, *Metody i techniki badań socjologicznych. Wybrane zagadnienia*, Uniwersytet Warmińsko-Mazurski, 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[9]. 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[10] (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?

   See also for example: PARP, *SIWZ Ewaluacji ex-post Phare 2003 Spójność Społeczna i Gospodarcza – komponent Rozwój Zasobów Ludzkich*, PARP [Polish Agency for Enterprise Development], Warsaw 2007, http://bip.parp.gov.pl/_gAllery/10/72/1072.pdf, p. 5-9; PARP, *SIWZ Ewaluacji sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego*, PARP, Warszawa 2007, http://bip.parp.gov.pl/_gAllery/17/24/1724.pdf, p. 1–4.

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[11].

Operationalisation – evaluation design

After specifying in details the study objective, object and general scope of evaluation we start the phase of operationalisation[12] 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[13]. 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[14]. 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[15].
- 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, other institutions implementing interventions of similar nature etc. Determining final population is

[11] More about this issue in the article in this publication by M. Szałaj, *Mechanisms of selecting assessment criteria for ex-post evaluation*.

[12] See an example of operationalisation of foresight study methodology: P. Kopyciński, Ł. Mamica, *Operacjonalizacja metodologii badań foresight*, Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie, Kraków 2006, http://www.foresight.msap.pl/download/operacjonalizacja.pdf.

[13] J. Górniak, S. Mazur, *Wskaźniki – instrument pomiaru skuteczności i efektywności polityki spójności*, http://www.konferencja-ewaluacja.pl/2007/materialy.php, p. 3.

[14] Compare J. Górniak, *Ewaluacja cyklu polityk publicznych*, in S. Mazur (ed.), *Ewaluacja funduszy strukturalnych*..., op. cit., p. 18.

[15] Compare ibidem, p. 19. More about this issue in the article by K. Keler & J. Górniak, *Indicators in ex-post evaluation of public programmes*.

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[16] 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) [17]. 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.)[18].

  The last two areas of operationalisation are the following:

[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] For more about this subject, see: European Commission, *Evaluating Socio-Economic Development*, Sourcebook 2, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/index_en.htm.

- 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[19]). 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.[20]

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[21]. Only fulfiling 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 recomendations from the assessment) and making them public.

---

[19] 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).

[20] J. Szlachta, *System ewaluacji funduszy strukturalnych w Polsce*, in S. Mazur (ed.), *Ewaluacja funduszy strukturalnych...*, op. cit., p. 86.

[21] Compare D. Nevo, *Konceptualizacja...*, op. cit., p. 63.

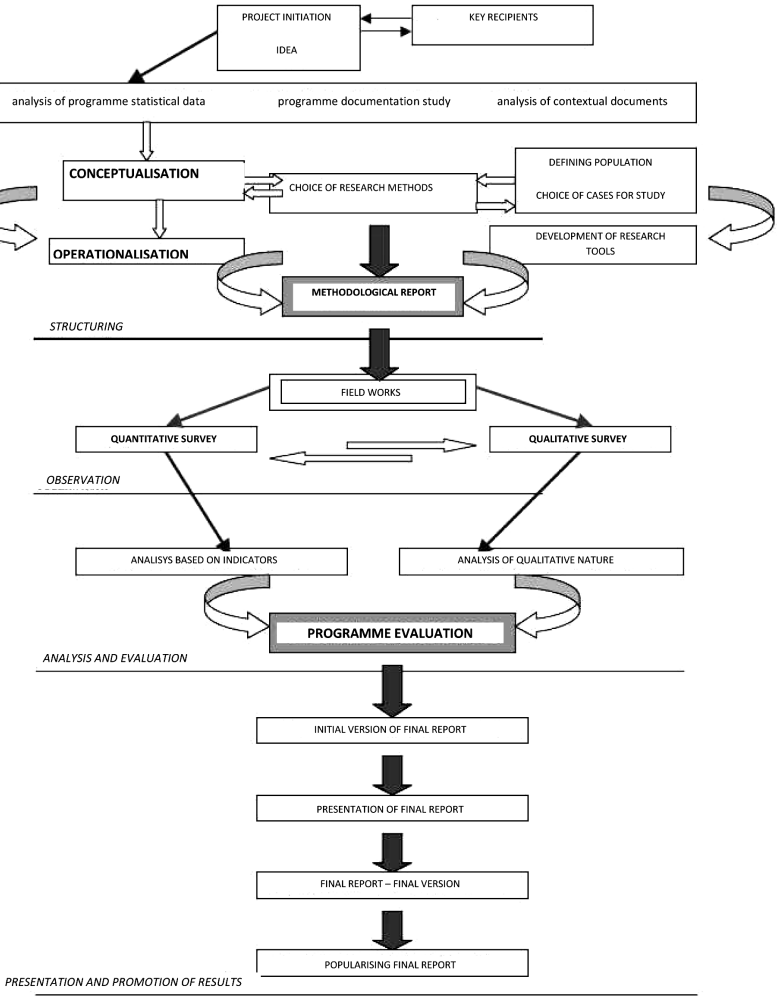


Fig. 1. Evaluation process

Source: Author's own compilation based on: E. Guba, Y. Lincoln, *Effective Evaluation*, Jossey-Bass, San Francisco 1981; E. Babbie, *Badania społeczne...*, op. cit.

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[22]. 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:

[22] 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. The observation phase is a very important stage for taking particular care of the quality and reliability of data. The most commom 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.

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/recipients, 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 ommited 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 recipents. 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

Phases of project life		Structuring		Observation Analysis Assessment	Presentation of assessment results
	CONCEPTUALISATION		**OPERATIONALISATION**	**PROJECT IMPLEMENTATION**	**PROMOTION OF ASSESSMENT RESULTS**
	initiating defining appointing project manager/entity responsible for implementation*/ commissioning evaluation	specifying project structure planning project implementation resource planning	organising project implementation by an internal/external team, commissioning the study, contractor selection	implementation control coordination	project closure
Character of activities	conceptual	planning	organisational	executive control coordination	reporting related to recipients promotional implementing
Products	evaluation design expertises, analyses, auxiliary works	Terms of Reference or scope of contractor's works	contractor's offer contract with contractor	methodological report assessment scheme partial reports from studies initial report and final report	final report with recommendations presentation of results table of recommendations publication of results/press articles
Direct participants/ key recipients	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	project manager with project team from the side of entity commissioning evaluation contractor with evaluation team key actors/ respondents – beneficiaries/ entities applying for support/project	contractor commissioning entity institutions involved in implementation of evaluated intervention (management, ordinary employees) key actors/respondents – beneficiaries/ entities applying for support/project

* in case of evaluations implemented externally.

Source: Author's own elaboration, compare: R.W. Griffin, *Podstawy zarządzania organizacjami*, translated into Polish by M. Rusiński, Wydawnictwo Naukowe PWN, Warszawa 1996; M. Trocki, B. Grucza, K. Ogonek, *Zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2003.

Conclusion

Presumably, many theroreticians 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 analitycal 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 symphaties 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 Sciences 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

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

European Commission, *Evaluating Socio-Economic Development*, Sourcebook 2, http:// ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/indexen.htm.

Górniak J., *Funkcje ewaluacji*, in S. Mazur (ed.), *Ewaluacja funduszy strukturalnych - perspektywa regionalna*, p. 16–23, Uniwersytet Ekonomiczny w Krakowie [Economic University in Cracow], Kraków 2007

Górniak J., Mazur S., *Wskaźniki - instrument pomiaru skuteczności i efektywności polityki spójności*, http://www.konferencja-ewaluacja.pl/2007/materialy.php.

Griffin R.W., *Podstawy zarządzania organizacjami*, translated into Polish by M. Rusiński, Wydawnictwo Naukowe PWN, Warszawa 1996.

Guba E., Lincoln Y., *Effective Evaluation*, Jossey-Bass, San Francisco 1981, http://www.konferencja-ewaluacja.pl/2007/materialy.php.

Kopyciński P., Mamica Ł., *Operacjonalizacja metodologii badań foresight*, Małopolska Szkoła Administracji Publicznej Akademii Ekonomicznej w Krakowie, Kraków 2006, http://www.foresight.msap.pl/download/operacjonalizacja.pdf.

Ministerstwo Rozwoju Regionalnego, *Plan ewaluacji Narodowych Strategicznych Ram Odniesienia na lata 2007–2013*, Departament of Structural Policy Coordination, Ministry of Regional Development, Warsaw 2007.

Nevo D., *Konceptualizacja ewaluacji edukacyjnej*, in L. Korporowicz (ed.), *Ewaluacja w edukacji*, Oficyna Naukowa, Warsaw 1997.

OECD, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris 2002, http://www.oecd.org/dataoecd/29/21/2754804.

PARP [Polish Agency for Enterprise Development, PAED], *SIWZ Ewaluacji ex-post Phare 2003 Spójność Społeczna i Gospodarcza – komponent Rozwój Zasobów Ludzkich*, PAED, Warsaw 2007, http://bip.parp.gov.pl/_gAllery/10/72/1072.pdf.

PAED, *SIWZ Ewaluacji sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego*, PAED, Warszawa 2007, http://bip.parp.gov.pl/_gAllery/17/24/1724.pdf.

Robson C., *Projektowanie ewaluacji*, in L. Korporowicz (ed.), *Ewaluacja w edukacji*, p. 147–152, Oficyna Naukowa, Warsaw 1997.

Simons H., *Polityczne implikacje teorii ewaluacyjnych*, in L. Korporowicz (ed.), *Ewaluacja w edukacji*, p. 65–92, Oficyna Naukowa, Warsaw 1997.

Słoma L., *Metody i techniki badań socjologicznych. Wybrane zagadnienia*, Uniwersytet Warmińsko-Mazurski, Olsztyn 2002.

Szlachta J., *System ewaluacji funduszy strukturalnych w Polsce*, in S. Mazur (ed.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Uniwersytet Ekonomiczny w Krakowie, Kraków 2007.

Trocki M., Grucza B., Ogonek K., *Zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warsaw 2003.

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[1].

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

[1] L. Korporowicz, *Metodologia badań ewaluacyjnych. O potrzebie integracji i kumulacji doświadczeń*, in: *Ewaluacja ex-ante – podsumowanie doświadczeń administracji polskiej*, Ministry of Regional Development, Warsaw 2007; K. Olejniczak, *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*, Scholar, Warsaw 2008; A. Haber (ed.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polish Agency for Enterprise Development, Warsaw 2007.

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[2]. 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.

[2] J.M. Owen, *Program Evaluation. Forms and Approaches*, Sage Publ., London, Thousand Oaks, New Delhi 1999; N. Norris, *Understanding Educational Evaluation*, CARE University of East Anglia, Kogan Press, London 1990.

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 evaluation[3]. 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.

[3] F. Znaniecki, *Education and Social Change*, E. Hałas, P. Lang (ed.), Frankfurt am Main 1998; M.Q. Patton, *Qualitative Evaluation and Research Methods*, Sage Publ., Newbury Park, London, New Delhi 1990; L. Korporowicz, *Nowe twarze ewaluacji. W kierunku inspiracji, analizy, czy inspiracji*, in: J. Kubin, J. Kwaśniewski (ed.), *Socjotechnika; kontrowersje, rozwój, perspektywy*, IPSiR UW, Warsaw 2000, p. 85–95.

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*[4], 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 fundaments 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

[4] M. Pariert, D. Hamilton, *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, 1972.

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"[5]. 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[6]. 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[7], 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.

[5] M.Q. Patton, *Creative Evaluation*, Sage Publ., Beverly Hills 1981, p. 270.

[6] M. Scriven, *Evaluation Thesaurus*, Sage Publ., Newbury Park, London, New Delhi 1991; E.G. Guba, Y.S. Lincoln, *Fourth Generation Evaluation*, Sage Publ., Newbury Park, London, New Delhi 1989.

[7] E. Hałas, *Interakcjonizm symboliczny*, Wydawnictwo Naukowe PWN, Warsaw 2006.

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 cosiousness 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[8]. 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.

[8] M.Q. Patton, *Qualitative Evaluation and Research Methods*, Sage Publ., London1990; K. Aspin-wall, T. Simkins, J.F. Wilkinson, M.J. McAuley, *Managing Evaluation in Education*, Routledge, New York-London 1992.

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[9]. 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 mac rostructural, 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[10].

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[11]. 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,

[9] K. Doktorowicz, *Europejski model społeczeństwa informacyjnego. Polityczna strategia Unii Europejskiej w kontekście globalnych problemów wieku informacji*, Śląsk University Publishing, Katowice 2005.

[10] M. Castells, *Społeczeństwo sieci*, Wydawnictwo Naukowe PWN, Warsaw 2007; A. Mattelart, *Społeczeństwo informacji*, trans. J.M. Pomorski, Universitas, Cracow 2004; D. Batorski, M. Marody, A. Nowak (ed.), *Społeczna przestrzeń Internetu*, Academica, Warsaw 2006; M. Truszkowska, J. Wojtkowiak, *Społeczeństwo informacyjne*, Gdańsk University Publishing, Gdańsk 2008; A. Bard, J. Soderqvist, *Netokracja. Nowa elita władzy i życie po kapitalizmie*, Wydawnictwa Akademickie i Profesjonalne, Warsaw 2006.

[11] L. Korporowicz, *Ewaluacja i społeczeństwo*, in: E. Hałas (ed.), *Rozumienie zmian społecznych*, Towarzystwo Naukowe KUL, Lublin 2001.

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."[12].

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:

[12] L.W. Zacher, *Transformacje społeczeństw. Od informacji do wiedzy*, C.H.Beck, Warsaw 2007, p. 232.

- 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[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 1

Manifesting characteristics of the knowledge based society	**Appropriate evaluation characteristics**
Technological characteristics:	
High level of use of advanced communication and information technologies in which the information collected almost immediately become a part of messages	Use of electronic communication and information technologies becomes one of the basic characteristics of the instruments portfolio and professional competencies of an evaluator
Wide access and constant update of information resources due to an increasing freedom of network flow of information	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.
Progressing potential for managing excessive and chaotic information by technological means for their processing, selection and arrangement	Fundamental meaning of clearly defined and operational criteria of evaluation as knowledge generators

[13] Ibidem, p. 231.

Table 1 – cont.

B. Structural characteristics	
Rational means to generate knowledge that includes many perspectives and variables, their contexts, "linking issues", "identification of conjugates"	Sensitisation to multidimensional and multifunctional character of an evaluation process with various potentials of use in various phases, in relation to its various participants
Ever wider use of knowledge in decision-making processes by modelling forecasting, simulation and future projection.	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
Dynamisation of the perception of reality, problems and solutions approached from prospective, developmental angle, rationalizing change and transformation designing	Increasing importance of democratic strategies of evaluation process, treated as „invitation for development" in organisation's strategic planning
Searching for synergies of processes, linking resources, animation of added value	Increasing role of dialogic, participative strategies and evaluation instrument portfolio techniques in its "reflexive practice" function, as well as sensitisation to synergetic evaluation criteria
Manifesting characteristics of the knowledge based society	Appropriate evaluation characteristics
C. Interactive characteristics	
Decentralisation of knowledge creation and dissemination centres linked with the decentralisation of power and participation in horizontal communication systems	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
Mediation of knowledge transmission and selection processes	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"
Intercultural and transcultural perspective of valuation of the knowledge creation, sharing and use process	Increasing importance of intercultural evaluation

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

Aspinwall K., Simkins T., Wilkinson J.F., McAuley M.J., *Managing Evaluation in Education, Routledge*, New York-London 1992.

Bard A., Soderqvist J., *Netokracja. Nowa elita władzy i życie po kapitalizmie*, Wydawnictwa Akademickie i Profesjonalne, Warsaw 2006.

Batorski D., Marody M., Nowak A. (ed.), *Społeczna przestrzeń Internetu*, Academica, Warsaw 2006.

Castells M., *Społeczeństwo sieci*, Wydawnictwo Naukowe PWN, Warsaw 2007.

Doktorowicz K., *Europejski model społeczeństwa informacyjnego. Polityczna strategia Unii Europejskiej w kontekście globalnych problemów wieku informacji*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2005.

Guba E.G., Lincoln Y.S., *Fourth Generation Evaluation*, Sage Publ., Newbury Park, London, New Delhi 1989.

Haber A. (ed.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polish Agency for Enterprise Development, Warsaw 2007.

Hałas E., *Interakcjonizm symboliczny*, Wydawnictwo Naukowe PWN, Warsaw 2006.

Korporowicz L., *Ewaluacja i społeczeństwo*, in: E. Hałas (ed.), *Rozumienie zmian społecznych*, Towarzystwo Naukowe KUL, Lublin 2001.

Korporowicz L., *Metodologia badań ewaluacyjnych. O potrzebie integracji i kumulacji doświadczeń*, in: *Ewaluacja ex-ante – podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego, Warsaw 2007.

Korporowicz L., *Nowe twarze ewaluacji. W kierunku inspiracji, analizy, czy inspiracji*, in: J. Kubin, J. Kwaśniewski (ed.), *Socjotechnika; kontrowersje, rozwój, perspektywy*, IPSiR UW, Warsaw 2000, s. 85–95.

Mattelart A., *Społeczeństwo informacji*, trans. Pomorski J.M., Universitas, Cracow 2004. Norris N., *Understanding Educational Evaluation*, CARE University of East Anglia, Kogan Press, London 1990.

Olejniczak K., *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*, Scholar, Warsaw 2008.

Owen J.M., *Program Evaluation. Forms and Approaches*, Sage Publ., London, Thousand Oaks, New Delhi 1999.

Parlett M., Hamilton D., *Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes*, Centre for Research in Educational Sciences, University of Edinburgh, 1972. Patton M.Q., *Creative Evaluation*, Sage Publ., Beverly Hills 1981.

Patton M.Q., *Qualitative Evaluation and Research Methods*, Sage Publ., Newbury Park, London, New Delhi 1990.

Scriven M., *Evaluation Thesaurus*, Sage Publ., Newbury Park, London, New Delhi 1991.

Truszkowska M., Wojtkowiak J., *Społeczeństwo informacyjne*, Gdańsk University Publishing, Gdańsk 2008.

Zacher L.W., *Transformacje społeczeństw. Od informacji do wiedzy*, C.H.Beck Publishing, Warsaw 2007.

Znaniecki F., *Education and Social Change*, E. Hałas, P.Lang (ed.), Frankfurt am Main 1998

.

Part II

Methods of Evaluation

Jarosław Górniak

Causal effect in evaluation

Evaluation of policy programmes in Poland in view of new challenges

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[1]. 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 complaience 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, includingt heir 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[2]. 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 1the 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[3]. 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[4]. 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 assotiation 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 cotrovercies 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". Dispiete 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". Also, programme objectives can be maintained while the concept of effective intervention changes.

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 centrall 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 elaborationg 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 (dadweight) 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 state[5].

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(Y1|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):

$$E(Y1|W=1) - E(Y0|W=0)$$

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 impact 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).

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 ε, 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 (τ), 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[7]. 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(Y0|W=0) actually represents E(Y0|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 nineteen 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[8] – *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

[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[9]. 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 nonmmeasurable 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[10]. 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, Jelonek, 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 otcomes 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[11]. 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 civering 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 engineerin*g.

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

Alkin, M.C. (2004). *Evaluation roots: tracing theorists' views and influences.* Thousand Oaks, Calif: Sage Publications.

Birckmayer, J.D., & Weiss, C.H. (2000). *Theory-based evaluation in practice: what do we learn?* Evaluation Review, 24(4), 407-431.

Bryson, A., Dorsett, R., & Purdon, S. (2002). *The use of propensity score matching in the evaluation of active labour market policies.*

European Commission. *Evalsed: the resource for evaluation of socio-economic development*. The Guide. 2009, from http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/index_en.htm.

Ferry, M., & Olejniczak, K. (2008). *Wykorzystanie ewaluacji w zarządzaniu programami unijnymi w Polsce.* [Use of evaluation in managing EU programmes in Poland.] Warsaw: Ernst & Young.

Górniak, J. (2007). *Evaluation in the Public Policy Cycle*. In S. Mazur (Ed.), *Evaluation of Public Interventions* - Regional Perspectives (pp. 12–27). Kraków: The Malopolska School of Public Administration, Cracow University of Economics.

Górniak, J., & Keler, K. (2007). *Wskaźniki w ewaluacji ex-post programów publicznych*. [*Indicators in ex-post evaluation of public programmes*] In A. Haber (Ed.), *Ewaluacja ex-post w teorii i praktyce* [*Ex-post evaluation in theory and practice*] (pp. 155–182). Warsaw: PAED.

Górniak, J., & Keler, K. (2008a). *Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych*. [*Evaluation and indicators in public policy planning and implementation cycles*] Zarządzanie Publiczne, 1(3), 109–124.

Górniak, J., & Keler, K. (2008b). *Rola systemu wskaźników w ewaluacji* [*Role of indicator system in evaluation*] in K. Olejniczak & B. Lendzion (Eds.), *Teoria i praktyka ewaluacji interwencji publicznych*. [*Theory and practice of public intervention evaluations*] Warsaw: Wydawnictwa Pedagogiczne i Naukowe.

Górniak, J., Mazur, S., Jelonek, M., Strzebońska, A., & Szczucka, A. (2008). *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*. [*Conditions and scope of efficiency audits (evaluation) in Polish administration*] Expert opinion commissioned by the Ministry of Finance.

Heckman, J.J. (2008). *Econometric causality*. Cambridge, MA: National Bureau of Economic Research.

Heckman, J.J. (2005a). *Rejoinder: Response to Sobel.* Sociological Methodology, 35, 135–162.

Heckman, J.J. (2005b). *The scientific model of causality*. Sociological methodology, 35(1), 1.

Heckman, J.J., & Smith, J.A. (1995). *Assessing the case for social experiments*. The Journal of Economic Perspectives, 9(2), 85–110.

Jakubowski, M. (2008). *Zastosowanie modeli EWD do ewaluacji polityki edukacyjnej*. [*Application of EWD models in evaluation of education policy*] http://coin.wne.uw.edu.pl/mjakubowski/publikacje/Jakubowski_EWD_metoda ewaluacji.pdf.

Jakubowski, M., & Sakowski, P. (2006). *Quasi-experimental estimates of class size effect in primary schools in Poland*. International Journal of Educational Research, 45(3), 202–215.

Konarski, R., & Kotnarowski, M. (2007). *Zastosowanie metody propensity score matching w ewaluacji ex-post*. [*Application of propensity score matching in ex-post evaluation*]. In A. Haber (Ed.), *Ewaluacja ex-post. Teoria i praktyka badawcza* [*Ex-post evaluation. Theory and research practice*] (pp. 183–209). Warsaw: PAED.

Korporowicz, L. (2008). *Cele ewaluacji ex ante. Myślenie o rozwoju.* [*Objectives of ex-ante evaluation. Thinking about development*]. *In Ewaluacja ex ante – podsumowanie doświadczeń administracji polskiej.* [*Ex-ante evaluation - summary of Polish administration experiences*] Warsaw: Ministry of Regional Development.

Mises, L. v. (1966). *Human Action: A Treatise on Economics*, Third Revised Edition. Henry Regnery Company.

Nagarajan, N., & Vanheukelen, M. (1997). *Evaluating EU expenditure programmes: a guide: Ex post and intermediate evaluation*. Office of Official Publications of the European Communities.

Olejniczak, K. (2008). *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki* [*Theory and practice of public intervention evaluations. Academic Guide*]. In K. Olejniczak & B. Lendzion. Warsaw: Wydawnictwa Pedagogiczne i Naukowe.

Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text 3rd edition*. Sage Publications.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation.* Sage Publications Ltd.

Popper, K. R. (1949). *The Poverty of Historicism*. Beacon Press/Free Press.

Sabatier, P., & Mazmanian, D. (1980). *The implementation of public policy: A framework of analysis*. Policy Studies Journal, 8(4), 538–560.

Scriven, M. (1991a). *Prose and cons about goal-free evaluation*. American Journal of Evaluation, 12(1), 55.

Scriven, M. (1991b). *Evaluation Thesaurus*. Sage Publications, Inc.

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

Sobel, M. E. (2005). *Discussion: The scientific model of causality*. Sociological Methodology, 99–133.

Strawiński, P. (2008). *Quasi-eksperymentalne metody ewaluacji.* [*Quasi-experimental methods of evaluation*] In A. Haber & M. Szałaj (Eds.), *Środowisko i warsztat ewaluacji* [*Environment and craft of evaluation*] (pp. 193–220). Warsaw: PAED.

Winship, C., & Sobel, M. (2004). *Causal Inference in Sociological Studies*. In M. Hardy & A. Bryman (pp. 481––503).

Roman Konarski, Michał Kotnarowski

Application of *propensity score matching* in *ex-post* evaluation

Introduction

One of the important purposes of ex-post evaluation is to estimate the causal effect of undertaken intervention. From the methodological perspective, a true experiment would be the most appropriate research plan to reach this goal. However, in most cases of programme evaluation conducted in the social sciences, it is impossible to apply a true experimental plan. Social experiments are too expensive to implement and often require unethical coercion to participate in comparative groups. On the one hand, people participating in an examination may not consent to submit to the experimental manipulation. On the other however – lack of such manipulation may be unethical, especially when it is beneficial. For example, a significant ethical problem will arise if an examiner assigns participants at random to a group covered by a training programme and to a control group (deprived of such programme) in order to estimate the effect of the programme.

Using *quasi*-experimental or non-experimental data to draw causal inferences is connected with significant problems[1]. Correlational examinations are characterised by the absence of random selection (randomisation) to the levels of exogenous (independent) variable, which does not allow adopting the assumption of equivalent comparative groups. In the last thirty years, the counterfactual state approach, developed by statisticians[2] and econometricians[3] has become dominant in the causal analysis in evaluation research. The propensity score matching (PSM) method, developed over twenty years ago by Rosenbaum and Rubin[4] is closely related to the counterfactual state approach.

In this chapter, we present an outline of the counterfactual state approach and closely connected with it the PSM *(propensity score matching)* method in the causal analysis of intervention effects. The chapter consists of six parts. The second part is an introduction to the counterfactual states approach. The third part presents the application of the PSM approach to the selection of a control group in evaluation research. The fourth part presents an example of the PSM method applied in the analysis of the net effect

---

[1] P.W. Holland, *Statistics and causal inference*, Journal of the American Statistical Association 1986, No 81(396), p. 945––960; R.J. LaLonde, *Evaluating the econometric evaluations of training programs with experimental data*, American Economic Review 1986, No 76, p. 604–620.

[2] D.B**.** Rubin, *Estimating causal effects of treatments in randomized and nonrandomized studies*, Journal of Educational Psychology 1974, No 66, p. 688–701; D.B. Rubin, *Assignment to treatment group on the basis of a covariate*, Journal of Educational Statistics 1977, No 2, p. 1–26; D.B. Rubin, *Bayesian inference for causal effects: The role of randomization*, Annals of Statistics 1978, No 6, p. 34–58; P.R. Rosenbaum, D.B. Rubin, *The central role of the propensity score in observational studies for casual effects*, Biometrika 1983, No 70(1), p. 41–55.

[3] J.J. Heckman, V.J. Hotz, *Choosing among alternative non-experimental methods for estimating the impact of social programs: The case of manpower training*, Journal of the American Statistical Association 1989, No 84, p. 862–880.

[4] P.R. Rosenbaum, D.B. Rubin, *The central role...*, op. cit.

of the regional component of the Phare 2002 SEC HRD programme. In the fifth part, we present the results of the estimation of the net effect for this programme. The sixth part summarises the possibilities of using the PSM method in evaluation research.

Counterfactual states

The counterfactual state approach enables conceptual interpretation of the problem of drawing conclusions on causal effects in non-experimental situations[5]. Let us assume that Y is a measure of the result of an intervention, efficiency of which is the subject of our assessment. Intervention may take the form of an educational programme or the assistance to the programme participants (beneficiaries) in finding employment. The examples of programme outcomes may be income, employment status or educational achievements of the beneficiaries. Next, let us assume, that although a specified person could be assigned *a priori* to one of two alternative states, a given person may be subjected only to one of these two conditions. Each of these two alternative states is characterised by a specific set of conditions, which has different impact on the outcome of intervention Y. For the purpose of this presentation we will define these two states as the intervention condition and the control condition. The intervention condition is participation in a programme and the control condition is the lack of participation in a programme.

In a typical evaluation research project, a defined group of beneficiaries participates in a programme (intervention), while the other, usually a larger group does not participate. Lack of participation in a programme is described as a control condition. A fundamental difference between the experimental situation and the evaluation exercise rests in the process of selection to the intervention and control conditions In cases of the experiment such selection is the result of randomisation. In cases of the analysis of programme effects, participation in a programme is the result of a self-selection process, resulting from characteristics (observable or not) of the programme beneficiaries. The key assumption of the counterfactual state approach is that individuals assigned to the intervention condition (programme) and the control condition (non-participation in a programme) has potential outcomes in both states; although, in reality they are observed only in one of the two conditions. As a result, examined individuals have observable results in a condition in which they were examined and only potential results in a condition they were not subjected to.

The framework of counterfactual states and counterfactual outcomes is summarised in Table 1, where every person i has a true result Y_i as an effect of subjecting to two alternative conditions, defined as the intervention condition (t) and the control condition (k). The outcome Y_i may be defined on an interval scale (e.g. level of income) or a discrete scale (e.g. employment status). Results Y_i^t and Y_i^k are potential outcomes of person i observed as a result of subjecting this person to the intervention condition and the control condition. Under the counterfactual framework we assume that both potential outcomes exist (if they are possible) in theory, but only one potential outcome, Y_i^t or Y_i^k, may be observed at the individual level.

[5] See: P.W. Holland, *Statistics*..., op. cit. In the literature on the counterfactual state approach there is a relatively large diversity in the applied notation. In our presentation of this approach we rely heavily on the presentation and notation developed by Winship and Morgan, see: C. Winship, S.L. Morgan, *The estimation of causal effects from observational data*, Annual Review of Sociology 1999, No 25, p. 659–706.

Table 1. Summary of counterfactual states and outcomes

		State (condition)	
		Intervention (t)	**Control (k)**
Group	Intervention (T) ($i \in$ T)	Observed outcome Y_i^t	Counterfactual outcome (not observed) Y_i^k
	Control (K) ($i \in$ K)	Counterfactual outcome (not observed) Y_i^t	Observed outcome Y_i^k

Source: Authors' own work.

The causal effect of an intervention (programme) on the outcome of every person i is defined as a difference between potential results observed in the intervention state and in the control state. Because both results Y_i^t and Y_i^k are possible in theory, we can define, at least in theory, the individual causal effect of a programme as

$$\delta_i = Y_i^t - Y_i^k$$

However we should point out that it is impossible to observe both outcomes Y_i^t and Y_i^k at the level of an individual person i. As a consequence we cannot directly calculate causal effects of an intervention at the level of individual programme participants.

The problem of calculating causal effects δ_i arises because outcomes Y_i^t and Y_i^k can be observed only for mutually exclusive subgroups in the population: the intervention group (T) and the control group (K). In order to observe the outcome Y, every person must be selected to one of two conditions. Individuals selected to the intervention group ($i \in$ T) have the value of selection variable $T_i = 1$. Whereas individuals selected to the control group ($i \in$ K) have the value of selection variable $T_i = 0$. As a result the observed outcome of person $Y_i = Y_i^t$ when $T_i = 1$, and $Y_i = Y_i^k$ when $T_i = 0$.

Under the counterfactual states approach it is assumed that although it is impossible to observe an individual causal effect of intervention δ_i, it is possible to define and estimate the average causal effect at the level of the population as

$$\overline{\delta} = \overline{Y}^{\mathrm{t}} - \overline{Y}^{\mathrm{k}}$$

where $\overline{Y}^t$ is the expected (average) population value of all individual results $Y^t_{i \in \mathrm{T}}$ of persons in the intervention condition, and $\overline{Y}^k$ is the expected (average) population value of all individual outcomes $Y^t_{i \in \mathrm{K}}$ of persons in the control condition. Because $\overline{Y}^t$ and $\overline{Y}^k$ are not observed (are missing) for mutually exclusive subsets of the population, $\overline{Y}^t$ and $\overline{Y}^k$ cannot be calculated for the whole population. It is possible however to estimate the values of $\overline{Y}^t$ and $\overline{Y}^k$, if we assume that it is possible to equate the intervention group (T) with the control group (K) so the two following two conditions are fulfilled: $\overline{Y}^t_{i \in \mathrm{T}} = \overline{Y}^t_{i \in \mathrm{K}}$ and $\overline{Y}^k_{i \in \mathrm{T}} = \overline{Y}^k_{i \in \mathrm{K}}$.

If we take the above assumptions, a consistent estimator of the average intervention (causal) effect is the difference between the estimated average outcome in the group of participants and the estimated average outcome in the group of non-participants of the intervention:

$$\hat{\overline{\delta}} = \hat{\overline{Y}}^{\mathrm{t}}_{i \in \mathrm{T}} - \hat{\overline{Y}}^{\mathrm{k}}_{j \in \mathrm{K}}$$

As a consequence, the selection process, that assigns individuals to the intervention and control groups, determines the level of bias with which the estimator $\hat{\bar{\delta}}$ estimates the average effect of the programme $\bar{\delta}$. The PSM method tries to eliminate (or at least considerably reduce) the bias resulting from the inevitable differences between the intervention group and the control group.

The counterfactual states approach to the analysis of intervention causal effects takes a critical assumption, known as the stable unit treatment assumption (SUTVA)[6]. SUTVA determines that the intervention effect for a specified participant of a programme does not depend on the participation of other individuals in the programme. This assumption cannot be verified empirically. An example of a violation of SUTVA may be a situation when a change in a local labour market takes place as a result of the participation of a large group of individuals in a training programme for the unemployed. As Winship and Morgan[7] notice, the simplicity of the counterfactual approach in the analysis of intervention causal effects vanishes in such situations.

The PSM method

The PSM method (propensity score matching) offers a particularly effective way for correcting bias of the estimators of the net effect of intervention created by selection. The purpose of the PSM method is to equate the intervention and control groups on measurable characteristics included in a vector of correlates X. Matching the control group to the intervention group is achieved through a selection of the control group from a larger pool of individuals not subjected to the intervention. The matching of the control group is conducted so that the distribution of measurable characteristics in X is maximally closed (balanced) in both groups. The most attractive property of the PSM method is the possibility of replacing the set of biased correlates in X with a single value, which is the estimated propensity score.

Probably the most obvious way of equating both groups on a set of the measurable correlates in X would be to find for every person i in the intervention group, a person j with exactly the same vector of measurable correlates in the group of individuals not subjected to the intervention. However, finding exact matches may be impossible, especially when all or some observed variables in X are defined on a continuous scale. A compromise solution would be to categorise the variables in X and to match all persons from the intervention group and the control group, who are in the same cell of the created contingency table. Unfortunately, such an approach is also not practical. If we categorise 20 continuous variables only into two categories for each variable, then the potential control group would be distributed among $2^{20} = 1048576$ matching categories. It is therefore very likely that with such a large number of possible categories, we can have many individuals in the intervention group without a perfectly matched unit in the pool of the potential control units.

The inability to match a control case to every case subjected to the intervention is defined as the common support problem. That is why in practice, we resign from the perfect matching and we select persons to the control group, who are to some degree "similar" to persons subjected to the intervention. The enormous insight of Rosenbaum and Rubin[8] rested in their observation that the multidimensionality of the problem of selecting a control group may be significantly reduced through equating the control group with the group subjected to the intervention on the basis of the propensity score, which we

[6] P.R. Rosenbaum, D.B. Rubin, *The central role...*, op. cit.

[7] C. Winship, S.L. Morgan, *The estimation of causal effects...*, op. cit.

[8] P.R. Rosenbaum, D.B. Rubin, *The central role...*, op. cit.

can informally define as the inclination for participation in the intervention condition. The propensity score is a scalar obtained through modelling the probability of selection to the intervention condition as a function of observable correlates (variables) included in the vector X. As a result the propensity score defines the probability of selection to the intervention condition, conditionally on the values of the observable variable in X[9]. In other words, the propensity score for an individual i is the probability that the person i, characterised by a set of values of the observed variables in X_i, will be assigned to the intervention group

$$P(X_i) = Pr(T = 1 / X_i)$$

where $T = \{0,1\}$ is an indicator of participation in the intervention group, X_i is a q-dimensional vector of values of characteristics observed before the intervention and

$$0 < P(X_i) < 1$$

If participation in the program is fully a function of the observed characteristics in the vector X (selection takes place only on the basis of the observed features), then conditionally on the correlates in X, participation in the programme is random from the outcome Y perspective. This assumption is called the conditional independence assumption (CIA).

CIA results from the balancing property possessed by the propensity score. Rosenbaum and Rubin[10] demonstrate analytically that if the intervention group and the control group have similar distribution of propensity scores, they will also have similar (balanced) distribution of the measurable variables in X. As a result the balancing property matching on the propensity cores mimics randomization in the experimental plan. Fulfilling both, the SUTVA assumption and the CIA assumption allows for estimating the counterfactual result for the group subjected to the intervention. However, similarly to the possibilities of empirical verification of the SUTVA assumption, the empirical verification of the CIA assumption is not fully possible. Fulfilling the CIA assumption depends mainly on identification and control of all variables which determine participation and the outcome of the intervention.

The implementation of the PSM method is basically a two stage procedure, where we first estimate propensity scores and then select a control group on the basis of the previously estimated propensity scores. However, for the clarity of this presentation of the PSM method, we will divide the implementation of this method into eight separate stages similar to the seven stages proposed by Rosenbaum and Rubin[11]. The eight stages are summarized in Figure 1. The implementation of the PSM method begins with the definition of the population of the individuals for whom we want to estimate the effect of the intervention. The next stage consists of the selection and measurement of the variables, which determine the process of programme selection and programme outcome. It is necessary that the measurement of these variables takes place before the natural process of selecting individuals to the group subjected to the intervention condition. We assume that this process is not random and it is responsible for the measurable (and immeasurable) differences between the group of programme beneficiaries and the group not subjected to the programme. We would like to diminish the consequences of this process of selection through the application of the PSM procedure. The purpose of the fourth stage is to select a representative sample N_T from the population of all persons subjected to the intervention and to select

[9] Ibidem

[10] Ibidem. See also: P.R. Rosenbaum, *Constructing matched of sets of strata*, in: Observational Studies, p. 200–224, Springer-Verlag, New York 1995.

[11] P.R. Rosenbaum, D.B. Rubin, *The bias due to incomplete matching*, Biometrics 1985, No 41, p. 103–116.

a relatively larger sample of "potential" controls N_{PK} from the control population of individuals not subjected to the intervention. A control group will be selected from this group of "potential" controls on the bases of the propensity scores in the further stages of the PSM method. The fifth stage involves estimating the propensity scores for the $N_T + N_{PK}$ individuals previously selected to the intervention group and the "potential" control group. In the sixth stage the selection of matched (balanced) group N_K of control units takes place. In the seventh stage we verify the matching quality of the control group. Usually, the last stage of implementing the PSM method consists of the assessment of the net effect of intervention as the difference between the average result in the group subjected to the programme and the average result of the selected counterfactual (control) group.

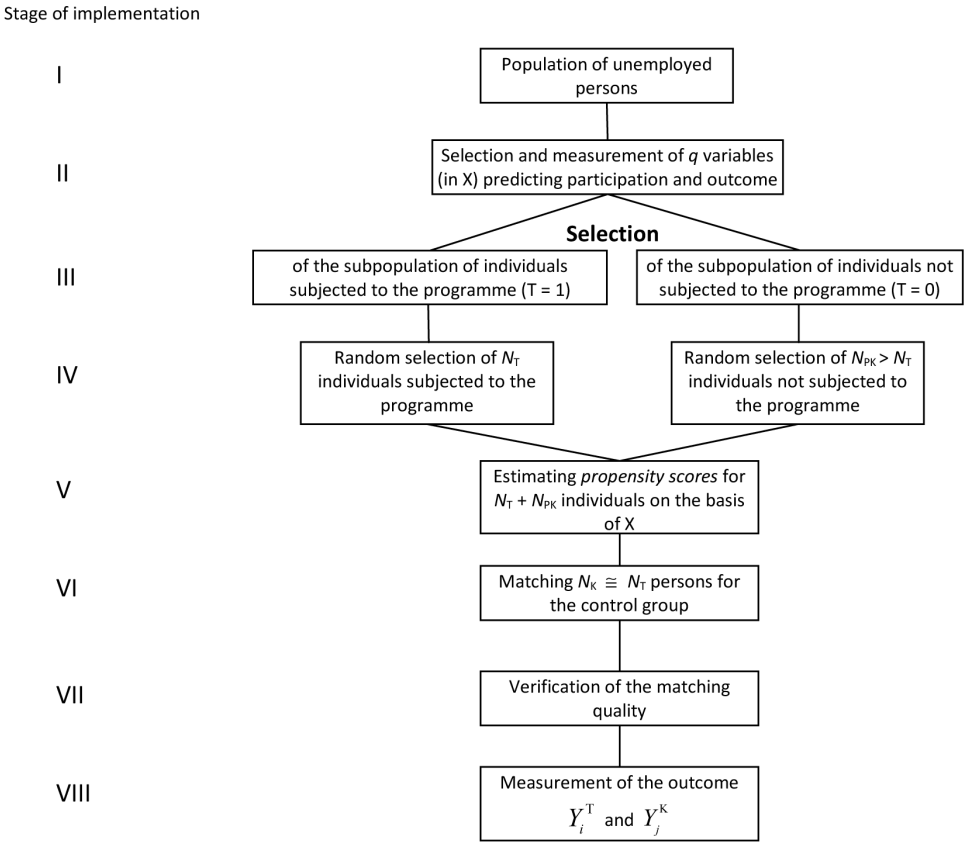


Fig. 1. Summary of stages of implementing the PSM method

Source: Authors' own work.

In the next part of the chapter, we will discuss in detail four stages of implementing the PSM method, which are particularly critical for accuracy of the assessed effect of the intervention. These four stages are: the selection and measurement of the correlates of program participation and programme outcome, the estimation of the propensity (score) for programme participation, the selection of a control group, and the assessment of the quality of the matched control group.

Selection and measurement of participation correlates

Selection of biasing correlates of the estimated net effect is one of the critical stages of implementing the PSM method. Two basic criteria of selecting these variables is the putative impact on the decision to participate in the programme and the programme outcome as well as the time of measurement in relation to the time of the undertaken intervention. First, the efficiency of matching the control group with the group subjected to the intervention is directly connected to the selection of appropriate correlates. Variables which determine the decision to participate in the programme and programme outcome may be observable or not observable. In the ideal situation, all variables which influence participation in the programme and its outcome should be included in the model estimating the propensity scores[12]. Variables which do not have an impact on either participation in the programme or its outcome are obviously inessential for the differences between the group subjected to the intervention and the control group. Also, there is no need for controlling variables having impact only on participation in the programme, because the result of the programme does not depend on them. Similarly, if variables have impact only on the result of the programme, there is no need for controlling these variables, because their distributions will not differ between the group subjected to the programme and the control group. As a result, variables requiring control are only variables which determine both the participation and outcome of the programme. The distributions of such variables will differ between the group of programme participants and the group of non-participants and at the same time will have a significant impact on the observed outcomes in both groups. The selection of variables should be based on theoretical considerations and previous empirical research. It is important to carefully consider all factors which may have an impact on the selection to the programme and the programme outcomes, and select a maximally inclusive list of measurable variables. Selected correlates are the basis for estimating the propensity scores, and the degree of fulfilling the CIA assumption depends on the quality of these variables. Heckman, Smith and Clements[13] proved that omitting important correlates may significantly bias the estimation of the net effect of an intervention.

Second, the measurement of selected variable values should take place before the commencement of the evaluated programme. We should especially avoid correlates measured after the programme has begun, because the values of these correlates might have changed as a result of the undertaken intervention[14]. Rosenbaum[15], Heckman, LaLonde and Smith[16] show why including such variable in the participation model may lead to a significant bias of the estimated programme effect. To understand that let us consider a variable that indicates the motivation level of the unemployed. The motivation level of an unemployed person is probably one of the most important factors in both the decision to join a programme and the programme outcome, which may be defined as finding employment. If motivation level of programme participants changes under the influence of participation in the programme, and we match the control group to the motivation level of individuals subjected to the programme at programme

---

[12] P.R. Rosenbaum, *The consequences of adjustment for a concomitant variable that has been affected by the treatment*, Journal of the Royal Statistical Society 1984, series A, No 147, p. 656–666.

[13] J.J. Heckman, J. Smith, N. Clements, *Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts*, Review of Economic Studies 1997, No 64, p. 487–537.

[14] P.R. Rosenbaum, *The consequences...*, op. cit.

[15] Ibidem.

[16] Heckman J.J., R.J. LaLonde, J.A. Smith, *The economics and econometrics of active labor market programs*, in: O.C. Ashenfelter, D. Card (ed.), *Handbook of Labor Economics*, vol. 3, p. 1865–2097, Elsevier, Amsterdam 1999.

completion, the result of such matching could be the masking (bias) of the estimated programme effect. That is why Rosenbaum[17] claims that the simplest and most secure solution is to avoid variables measured already after the intervention has begun, even if a given correlate is a key determinant of the decision to participate in the programme and its outcome.

Estimation of the propensity to participate scores

In practice, the scores of the inclination for programme participation (propensity scores) $P(X_i)$ of individuals in the intervention group and the control group are unknown and must be estimated. Although there are various methods for estimating $P(X_i)$, definitely the most extensively used statistical model is the logistic regression model. For example, Rosenbaum and Rubin[18] as well as Rosenbaum[19], use the logistic regression model when presenting the PSM method.

Estimating $\Pr(T_i = 1 \mid X_i)$ is achieved through the following logistic regression model:

$$\Pr(T_i = 1 / X_i) = \frac{exp\left[\beta_0 + \sum_{r=1}^{q} \beta_r X_{ir}\right]}{1 + exp\left[\beta_0 + \sum_{r=1}^{q} \beta_r X_{ir}\right]},$$

where β_0 is the regression constant, and $\beta_r (r = 1,2..., q)$ are regression coefficients for every q predicator of the indicator of participation in the intervention group. The estimated regression model may include main effects as well as interaction and/or quadratic effects[20].

Two issues related to the specification of the estimated participation model should be considered. First, we must consider which of the previously selected determinants of participation and programme outcome to include in the final regression model. Rubin and Thomas[21] advise not to exclude variables with statistically insignificant coefficients from the model. They[22] claim that including statistically insignificant variables in the model may improve the fit of the control group and only minimally complicate the estimation of individual propensity for participation (propensity scores). Also Zhao[23] confirmed in the series of simulation research, that including variables insignificant for predicting participation does not bias estimations of individual propensity scores.

Second, we must consider whether the tested model should be "additive" and include only main effects, or should it include interaction effects and/or quadratic (or higher order) effects. Although

---

[17] P.R. Rosenbaum, *The consequences...*, op. cit., p. 663.

[18] P.R. Rosenbaum, D.B. Rubin, *Constructing a control group using multivariate matched sampling methods that incorporate the propensity score*, The American Statistician 1985, No 39(1), p. 33–38.

[19] P.R. Rosenbaum, *Constructing matched of sets of strata*, in: Observational Studies, p. 200–224, Springer-Verlag, New York 1995, chapter 9.

[20] Details of estimating the logistic regression model can be found in classic publications, such as (in the order of technical advancement of presentation): S. Menard, *Applied Logistic Regression Analysis*, Second Edition, Sage Publications, Thousand Oaks 2002; D.W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, Second Edition, Wiley, New York 2000; D.R. Cox, E.J. Snell, *Analysis of Binary Dat*a, Second Edition, Chapman and Hall, New York 1989.

[21] D.B. Rubin, N. Thomas, *Matching using estimated propensity scores: relating theory to practice*, Biometrics 1996, No 52, p. 253.

[22] Ibidem.

[23] Z. Zhao, *Sensitivity of Propensity Score Methods to the Specifications*, Discussion Paper No. 1873, Institute for the Study of Labor, Bonn 2005.

Rosenbaum and Rubin[24] estimate a model which includes quadratic effects, there are not many other examples of using models more complex than the additive model. The fact that the probability of numeric problems taking place in estimating such models may be an evidence against a routine usage of models which include interactions and/or effects of the higher order[25]. Moreover, Zhao[26] proves that including statistically insignificant effects of the higher order may lead to biased estimations of propensity for participation, and as a result biased estimation of the net effect of the undertaken intervention.

Selection of the control group

Various methods of matching the control group, based on previously estimated propensity scores were described in the statistical literature[27]. Rosenbaum and Rubin[28] very accessibly explain the implementation of the nearest neighbour matching method, based on the Mahalanobis distance metric, which includes the propensity for participation (Mahalanobis metric matching including the propensity score), and the implementation of a modified caliper method (nearest available Mahalanobis metric matching within calipers defined by the propensity score). Heckman and his co-workers[29] present the kernel matching method.

The nearest neighbour method is the most often used and also the most intuitive method for selecting the control group[30]. In the traditional version of this algorithm, a random arrangement of cases subjected to the intervention takes place prior to the selection of the control individuals, who have the most similar propensity, score for every person in the intervention group. This algorithm can be implemented with or without replacement. In a version with replacement, a selected control case if returned to the pool of potential control cases and can be matched again to another case subjected to the intervention. In a version without replacement, a selected control case is removed from the pool of the potential control group after it is matched to a case subjected to the intervention.

The nearest neighbour method allows for the selection of a single control unit ($k_i = 1$) or several control units (e.g. $k_i = 2$) for every case in the intervention group. Every person in the intervention group should be matched to the same number of individuals in the control group. Matching several control cases increases the numbers of the control group and at the same time reduces the standard error of the estimated intervention effect. However, it can also increase the bias of the estimated intervention effect, because it increases the probability of occurrence of weak (distant) matches. The nearest neighbour method effectively overcomes the common support problem because every person in the intervention group always receives a matched individual from the control group. The disadvantage of this method

[24] P.R. Rosenbaum, D.B. Rubin, *The central role...*, op. cit.

[25] See: D.W. Hosmer, S. Lemeshow, *Applied Logistic Regression...*, op. cit., chapter 4.5.

[26] Z. Zhao, *Sensitivity...*, op. cit.

[27] See: P.R. Rosenbaum, D.B. Rubin, *Constructing a control group...*, op. cit.; J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme*, Review of Economic Studies 1997, No 64, p. 605–654; J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator*, Review of Economic Studies 1998, No 65, p. 261–294.

[28] P.R. Rosenbaum, D.B. Rubin, *Constructing a control group...*, op. cit.

[29] J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme*, op. cit.; J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator*, op. cit.

[30] P.R. Rosenbaum, D.B. Rubin, *Constructing a control group...*, op. cit.

for selecting a control group is the possibility of occurrence weak matches mentioned above, at least for some cases in the intervention group.

The caliper method modifies to some extent the nearest neighbour method by allowing matches only within a predefined tolerance (caliper) *c*. The lower the tolerance is, the closer the match. Rosenbaum i Rubin[31] recommend determining the value of tolerance ***c*** as a fraction of the dispersion of the propensity score distribution, e.g. c = 0.25***s***, where $s = \sqrt{\left(s_T^2 + s_{PK}^2\right)/2}$ is an averaged standard deviation of propensity score distribution in the intervention group (s_T^2) and the control group (s_{PK}^2). On the other hand, Dehejia and Wahba[32] determine the value of tolerance as c = 0.0001 or c = 0.00001. The caliper method effectively prevents the occurrence of weak matches, but can also cause that certain individuals in the intervention group will not obtain a control unit, which leads to the occurrence of the common support problem. Just as in the case of the nearest neighbour method, the caliper algorithm may be applied with or without replacement, as well as a version with either one control unit ($k_i = 1$) or several control units (e.g. $k_i = 2$) selected for every case in the intervention group.

The kernel method which is a modification of the caliper method has been developed by Heckman and his coworkers[33]. In this method, every member of the intervention group obtains all units of the control group within a distance defined by the value of tolerance ***c***. Selected control units are weighted inversely proportional to the quality of match (distance) to their cases in the intervention group. Weights w_{ij} are defined as

$$w_{ij} = \frac{\left(\frac{P(X_i) - (P(X_j)}{c}\right)}{\Sigma_j K\left(\frac{P(X_i) - (P(X_j)}{c}\right)}$$

where $K(.)$ is the kernel function from the normal (Gaussian) distribution. The sum of all weights of the control units ***j*** within every unit in the intervention group ***i*** equals 1: $\sum_j w_{ij} = 1$.

Verification of the quality of matching the counterfactual group

The examination of the quality of matching the control group consists of mainly the assessment of the difference of the propensity score distribution and the distributions of participation predicators, utilized in the participation regression model, across the intervention group and the control (counterfactual) group. Lack of difference in distributions across these groups is an evidence of the quality of the selected control group.

Rosenbaum and Rubin[34] provide a method for defining the efficiency of the applied matching algorithm in selecting the control group. In this procedure, first we define the extent of the lack of balance between the intervention group and the control group for all participation correlates in the X.

---

[31] Ibidem.

[32] R.H. Dehejia, S. Wahba, *Propensity score-matching methods for non-experimental causal studies*, The Review of Economics and Statistics 2002, No 84(l), p. 151–161.

[33] J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme*, op. cit; J.J. Heckman, H. Ichimura, P. Todd, *Matching as an econometric evaluation estimator*, op. cit.

[34] P.R. Rosenbaum, D.B. Rubin, *Constructing a control group...*, op. cit.

To this end we calculate a standardized percentage difference between the averages of the predictor variables obtained in the intervention group and the group not participating in the intervention: $100\left(\bar{X}_{\mathrm{T}}-\bar{X}_{\mathrm{PK}}\right)\Big/\sqrt{\left(s_{\mathrm{T}}^2+s_{\mathrm{PK}}^2\right)\Big/2}$, where $\bar{X}_T$ and $\bar{X}_{PK}$ are the averages in the group subjected to the intervention and the group of potential control individuals (not subjected to the intervention), and (s_T^2) and (s_{PK}^2) are corresponding estimates of the variances. Similarly, we can calculate the standardized percentage difference between the averages of the predictor variables in the intervention group and the matched control group: $100\left(\bar{X}_{\mathrm{T}}-\bar{X}_{\mathrm{K}}\right)\Big/\sqrt{\left(s_{\mathrm{T}}^2+s_{\mathrm{PK}}^2\right)\Big/2}$, where $\bar{X}_T$ and $\bar{X}_K$ are the predictor means obtained in the group subjected to the intervention and the control group. The variances s_T^2 and s_{PK}^2 have been defined earlier. Comparison of these two standardized percentage differences, obtained before and after the application of the PSM method, shows the efficiency of the applied PSM procedure in the reduction of the lack of match between the intervention group and control units for every observable variable in the X.

However, it should be stressed that the above method for verifying the quality of the control group does not allow verifying the extent of meeting the SUTVA and the CIA assumption. Both assumptions cannot be verified empirically. The extent of the violation of these two assumptions can only be assessed with an appropriate theoretical knowledge, supported by strong empirical research of the process that leads to the selection and outcome of the intervention programme.

The analysis of the net effect in ex-post evaluation of the regional component of the Phare 2002 SEC HRD programme[35]

The purpose of the presented research was to estimate the size of the net effect for the Subproject 1 of the regional component of Phare 2002 Social and Economic Cohesion (SEC) – Human Resources Development (HRD). The intervention group in the Subproject 1 of the Phare 2002 SEC HRD programme comprised of unemployed persons or persons endangered by unemployment, who were provided with various labour market services. Subproject 1 had a goal to help its beneficiaries to find employment. In the presented analysis of the net effect we use data on beneficiaries, who were registered as unemployed in Poviat Employment Agencies (PEA). The analysed data comes from PULS - the register of unemployed persons[36]. The data in PULS is collected by Poviat Employment Agencies and includes information on the registered unemployed persons. It concerns the employment history of the unemployed, the utilization of various services on the labour market, and it also includes certain social and demographic characteristics of the registered individuals.

The presented evaluation of the Phare programme was conducted separately in each of 16 voivodeships. Details of selecting the sample of programme beneficiaries and the group of programme non-participants in every voivodeship can be found in the appropriate report[37]. For the purpose of this article we only concentrate on the research stages directly connected to the implementation of the PSM method. For clarity and presentation economy, presented results of the evaluation are aggregated at the country level.

---

[35] PBS DGA, *Ewaluacja ex-post komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza – Komponent Rozwój Zasobów Ludzkich*, Report from research conducted at the request of the Polish Agency for Enterprise Development, Polish Agency for Enterprise Development, Warsaw 2006, http://www.parp.gov.pl/index/index/110.

[36] SI PULS is an information system supporting the work of Poviat Employment Agencies.

[37] PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op. cit.

Selection and measurement of participation correlates

In the selection of available participation correlates we used two most important criteria. First, selected variables should have a putative effect on the decision to participate in the programme and the programme outcome (finding employment). Second, measurement of selected variables should be conducted before joining the programme. If the measurement of a given variable was conducted after the programme begun, then the values of this variable could not have changed in time. The list of variables selected as predicators of the estimated propensity for participation (propensity scores) is presented in Table 2. It also includes the list of averages and standardized percentage differences between averages for the group of N_T = 4818 programme participants[38] and the group of N_{PK} = 823476 individuals not participating in the programme. As can be seen in Table 2, the group participating in the programme is significantly different from the rest of the unemployed. For example, persons participating in the programme are younger, better educated, and remained unemployed for a shorter period of time before joining the programme. The purpose of applying the PSM procedure was to select a control group definitely more similar to the group of programme participants than the entire group of persons not participating in the programme.

Table 2

Variable	Description of the characteristic	Programme participants	Non-participants	% bias
V01	Sex (female)	50.17	52.35	-4.4
V02	Age (at the moment of programme commencement)	30.96	35.17	-40.3
V03	Marital status (married)	40.54	50.95	-21.0
V04	Single parent	2.89	3.00	-0.7
V05	Number of children	0.46	0.62	-15.7
V06	Disability (disabled)	3.13	2.52	3.7
V09 Education				
	1 None or incomplete primary education	0.12	0.40	-5.4
	2 Primary	15.01	27.81	-31.6
	3 Secondary school	0.66	0.31	5.1
	4 Basic vocational	25.16	31.47	-14.1
	5 High school	40.76	28.27	26.5
	6 Post-secondary	6.00	3.68	10.8
	7 University (including BA or BSc)	12.29	8.06	14.0
VII Occupation				
	0 Armed forces	0.48	0.19	5.0
	1 Representatives of public authorities, higher officials and managers	0.77	0.77	-0.1

[38] Total number of beneficiaries of the Subproject 1 of the regional component of Phare 2002 Social and Economic Cohesion (SEC) – Human Resources Development (HRD) was 13,005 persons. Due to practical and organisational reasons it was impossible to conduct an analysis on data from all PEAs. In relation to that a stratified selection of PAE sample was conducted. Details of this selection are described in the report: PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op. cit.

Table 2 - continuation

	2 Specialists	7.18	6.07	4.5
	3 Technicians and mid-level personnel	15.11	10.85	12.7
	4 Office employees	8.12	6.29	7.1
	5 Personal services employees and clerks	14.78	15.79	-2.8
	6 Farmers, gardeners, foresters and fishermen	0.95	2.06	-9.1
	7 Industrial workers and artisans	22.40	24.54	-5.1
	8 Operators and fitters of machines and devices	5.54	6.39	-3.6
	9 Workers at simple works	11.44	15.95	-13.2
	10 No occupation	8.47	7.75	2.6
	99 No data available	4.77	3.34	7.3
V14	Having driving licence cat. B	12.08	9.52	8.3
V16	Number of trainings in the year before Phare	0.04	0.01	14.7
V17	Number of training days	1.33	0.59	10.9
V13	Number of offers in the year before Phare	0.46	0.39	4.3
VI8 Participation in an internship before the programme commencement				
	0 Did not participate in an internship	85.08	83.69	3.8
	1 Participated later than a year and in the last year before the programme	1.54	1.73	-1.5
	2 Participated later than a year before the programme	8.01	8.30	-1.1
	3 Participated in an internship in the last year before the programme	5.38	6.28	-3.8
V07	Work experience at the moment of joining Phare	2035.79	2487.75	-15.1
V10	Number of days as unemployed before participation in Phare	1087.22	1438.69	-31.3
V12	Number of performed occupations	2.33	2.22	6.3
V15	Number of days worked on post-graduate internships before Phare, intervention works, public works	3.31	6.16	-10.1
V22	Number of days of the longest continuous period of long-term unemployment	287.31	337.32	-18.0
V21	Work or internship +/- 7 days from the programme	2.74	2.50	1.5
V19	Number of days on the unemployment benefit before participation in the programme	362.49	356.70	1.9
V08 Percentage of appearance at the office's call				
	0%	7.10	8.07	-3.7
	0-50%	5.77	5.25	2.3
	51-99%	9.36	8.02	4.8
	100%	34.45	30.01	9.5
	Not applicable	43.32	48.65	-10.7
V20 Right to unemployment benefit at the moment of joining Phare				
	0 No right for unemployment benefit	32.69	44.82	-25.1

Table 2 - continuation

	1 Had the right to unemployment benefits at the beginning of participation in Phare and in the <l following months	5.54	5.12	1.9
	2 Had the right to unemployment benefits at the beginning of participation in Phare and in the 1-3 following months	16.06	9.28	20.5
	3 Had the right to unemployment benefits at the beginning of participation in Phare and in the >3 following months	45.70	40.78	10.0

Estimation of the scores of propensity to participate

Estimation of the individual propensity to participate scores (propensity scores) for the $N_T = 4818$ programme participants and the group of $N_{PK} = 823476$ individuals not participating in the programme was conducted with the logistic regression model, independently (separately) in each voivodeship. In the estimated regression model the dependent variable was a dichotomous variable Phare (1 = participation in the programme, 0 = non-participation in the programme). The independent variables were the 22 characteristics defined in Table 2. The tested model was an additive model and did not include any interaction effects or higher order effects. Moreover, according to Rubin and Thomas'[39] recommendation, statistically insignificant variables were not removed from the model. The final effect of this stage of implementing the PSM method was the estimated individual propensity scores for all participants and non-participants of the Phare programme.

Selection of the control group

Selection of the control group from the group of programme non-participants was conducted separately in each voivodeship. In the selection of the control group we used the nearest neighbour method. It allowed for a relatively clear definition of the control units (person with the closest propensity scores) and a relatively simple implementation of the algorithm. At the same time, this method provides a control group with similar properties as these obtained by relatively more complex algorithms[40]. The relative simplicity of the nearest neighbour approach arises from the lack of the necessity to define and implement the tolerance level (like in the caliper method) and the lack of the necessity to implement a complex weighting of the control units (like in the kernel method).

The purpose of applying the selection algorithm was to select one control unit for every programme beneficiary. Because the number of persons not subjected to the programme, from which we selected the control group, is relatively large (several dozen thousand in each voivodeship), we used the method of selection without replacement. Despite its conceptual clearness and simplicity, implementation of the nearest neighbour method carries specific "complications" resulting from the fact that matching a control

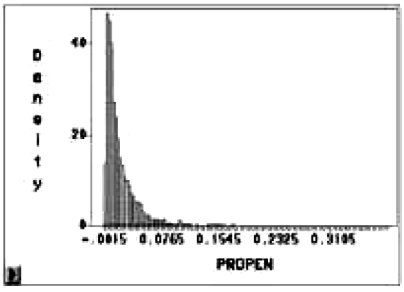
[39] D.B. Rubin, N. Thomas, *Matching using estimated propensity scores: relating theory to practice*, Biometrics 1996, No 52, p. 249–264.

[40] See e.g. P.R. Rosenbaum, D.B. Rubin, *Constructing a control group using multivariate matched sampling methods that incorporate the propensity score*, The American Statistician 1985, No 39(1), p. 33–38.

person to a given programme participant is not always one-to-one. However, for most cases matching one-to-one was possible. For cases when for one programme participant there were several potential individuals (with the same propensity score) in the group of non participants (matching one-to-many), the implemented algorithm selected one control individual at random. For cases when one person from the non participation group was the nearest neighbour for several individuals in the programme group (matching many-to-one), the algorithm selected subsequent nearest neighbours for every unit in the programme group with identical propensity score. This situation became numerically complex, when many persons in the programme group had several, the same, and equally good nearest neighbours (matching many-to-many). Then the selection of the exact number of the control group could be impossible. For that reason the size of the control group ($N_K = 4822$) slightly exceeds the size of the group of programme participants ($N_T = 4818$).

The quality of the applied selection algorithm can be verified through the comparison of distribution of the estimated propensity scores in the group subjected to the programme and the selected control group. Distributions of the individual propensity for participation scores and basic statistics describing features of these distributions in both groups are presented on Fig. 2. As we can see the distributions of estimated propensity scores in both groups are very similar with respect to the first statistical moments (averages) as well as the higher moments (variance/skewness/kurtosis).

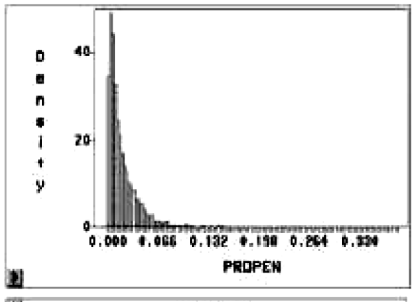
(A) Programme participants



Moments			
N	4818.0000	Sum Wgts	4818.0000
Mean	0.0824	Sum	187.9267
Std Dev	0.0316	Variance	0.0010
Skewness	4.6392	Kurtosis	31.3013
USS	7.2428	CSS	4.8252
CV	141.2887	Std Mean	0.0005

Quantiles			
100% Max	0.3848	99.0%	0.1635
75% Q3	0.0269	97.5%	0.1030
50% Med	0.0127	95.0%	0.0697
25% Q1	0.0060	90.0%	0.0478
0% Min	0.0001	10.0%	0.0029
Range	0.3847	5.0%	0.0018
Q3-Q1	0.0209	2.5%	0.0012
		1.0%	0.0007

(B) Control group



Moments			
N	4822.0000	Sum Wgts	4822.0000
Mean	0.0822	Sum	186.9271
Std Dev	0.0309	Variance	0.0010
Skewness	4.6154	Kurtosis	31.2928
USS	6.9713	CSS	4.8003
CV	139.3835	Std Mean	0.0004

Quantiles			
100% Max	0.3800	99.0%	0.1635
75% Q3	0.0269	97.5%	0.1023
50% Med	0.0127	95.0%	0.0676
25% Q1	0.0060	90.0%	0.0476
0% Min	0.0001	10.0%	0.0030
Range	0.3799	5.0%	0.0018
Q3-Q1	0.0209	2.5%	0.0012
		1.0%	0.0007

Fig. 2. Distribution of propensity scores in the group of programme participants (panel A) and the matched control group (panel B)

Verification of the quality of matching the control group

The efficiency of the applied PSM method to match a control group to the group of the Phare programme beneficiaries can be assessed through the analysis of the reduction of the measurable variable bias. Table 3 presents the mean levels and standardized percentage differences between the variable means obtained in the group of the Phare programme participants and the matched control group. Let us notice that the standardized differences in the fifth column of Table 3 have the same denominator as the standardized differences in the fifth column of Table 2. As a consequence, a comparison of the level of measurable variable bias (after matching) in Table 3 with the corresponding level of bias (before matching) in Table 2 shows the benefits resulting from the implementation of the PSM method. While before applying the PSM procedure almost half of standardized differences exceeded the level of 20% (see Table 2), after applying the matching procedure none of standardized differences exceeds the level of 9% (see Table 3). As we see in Table 3, the applied PSM method proved to be most efficient in equating the control group in terms of the characteristics for which the largest difference was observed (e.g. age, marital status, education, number of days as unemployed before the programme) between the beneficiaries and the group of individuals who did not participate in the Phare programme. However, we can also notice that although the reduction of bias for most characteristics is significant, the bias in certain educational categories (V07) remains at a relatively high level. As a consequence, further statistical corrections of the distribution of the education variable may be required for further reduction of the possible bias in the estimate of the programme net effect.

Table 3. Bias of observable variables: averages in the group of programme participants and in the fitted control group and the standard difference (in %)

Variable	Description of the characteristic	Programme participants	Fitted control group	% bias
V01	Sex (female)	50.17	50.37	-0.4
V02	Age (at the moment of programme commencement)	30.96	31.11	-1.5
V03	Marital status (married)	40.54	40.79	-0.5
V04	Single parent	2.89	2.68	1.2
V05	Number of children	0.46	0.46	0.6
V06	Disability (disabled)	3.13	3.30	-1.0
V09 Education				
	1 None or incomplete primary education	0.12	0.29	-3.2
	2 Primary	15.01	14.41	1.5
	3 Secondary school	0.66	0.41	3.6
	4 Basic vocational	25.16	28.37	-7.2
	5 High school	40.76	36.64	8.7
	6 Post-secondary	6.00	5.04	4.5
	7 University (including BA or BSc)	12.29	14.83	-8.4
VII Occupation				
	0 Armed forces	0.48	0.29	3.2
	1 Representatives of public authorities, higher officials and managers	0.77	0.73	0.5

Table 3 - continuation

	2 Specialists	7.18	7.01	0.7
	3 Technicians and mid-level personnel	15.11	15.41	-0.9
	4 Office employees	8.12	8.38	-1.0
	5 Personal services employees and clerks	14.78	14.66	0.3
	6 Farmers, gardeners, foresters and fishermen	0.95	0.89	0.5
	7 Industrial workers and artisans	22.40	22.48	-0.2
	8 Operators and fitters of machines and devices	5.54	5.23	1.3
	9 Workers at simple works	11.44	11.26	0.5
	10 No occupation	8.47	9.15	-2.5
	99 No data available	4.77	4.52	1.3
V14	Having driving licence cat. B	12.08	11.41	2.2
V16	Number of trainings in the year before Phare	0.04	0.03	1.3
V17	Number of training days	1.33	1.25	1.1
V13	Number of offers in the year before Phare	0.46	0.44	1.1
VI8 Participation in internship before the programme commencement				
	0 Did not participate in an internship	85.08	84.78	0.8
	1 Participated later than a year and in the last year before the programme	1.54	1.18	2.8
	2 Participated later than a year before the programme	8.01	8.52	-1.9
	3 Participated in an internship in the last year before the programme	5.38	5.52	-0.6
V07	Work experience at the moment of joining the Phare	2035.79	2043.63	-0.3
V10	Number of days as unemployed before participation in Phare	1087.22	1094.36	-0.6
V12	Number of performed occupations	2.33	2.34	-0.5
V15	Number of days worked on post-graduate internships before Phare, intervention works, public works	3.31	3.26	0.2
V22	Number of days of the longest continuous period of long-term unemployment	287.31	296.47	-3.3
V21	Work or internship +/- 7 days from the programme	2.74	3.17	-2.7
V19	Number of days on the unemployment benefit before participation in the programme	362.49	369.52	-2.3
V08 Percentage of appearance at the office's call				
	0%	7.10	7.84	-2.8
	0-50%	5.77	5.85	-0.3
	51-99%	9.36	9.42	-0.2

Table 3 - continuation

	100%	34.45	35.44	-2.1
	Not applicable	43.32	41.46	3.7
V20 Right to unemployment benefit at the moment of joining Phare				
	0 Not entitled for unemployment benefit	32.69	31.81	1.8
	1 Had the right for unemployment benefit at the beginning of participation and during next 1- months	5.54	5.89	-1.5
	2 Had the right for unemployment benefit at the beginning of participation and during next 1-3 months	16.06	16.01	0.2
	3 Had the right for unemployment benefit at the beginning of participation and during next >3 months	45.70	46.29	-1.2
Propensity score		2.24	2.22	1.0

The net effect of the regional component of Phare 2002 SEC HRD programme

The graph of the percentage of employed in the group of Subproject 1 Phare 2002 SEC HRD beneficiaries, the control group, and the group not participating in the programme is presented in Figure 3[41]. When comparing the percentage of employed among beneficiaries and in the control group, we can observe that already in the first month after completing participation in the programme, beneficiaries were employed statistically more often than the individuals from the control group. The percentage of employed among beneficiaries after one month since completing participation in the programme equalled 21.4%, whereas the corresponding percentage for the control group was 17.1%. In the following months, the difference between percentages of employed in both groups was increasing and statistically significant. After six months since completing participation in the programme, the percentage of employed in the group of beneficiaries was 40.8%, whereas in the control group only 31.0%. One year after completing the programme 53.0% of beneficiaries and 40.8% of the individuals in the control group were employed. After one year since completing participation in the programme, the percentage of the employed beneficiaries was therefore 1.29 times higher than the same percentage in the control group, which means that two years after completing the programme, participation in it increased beneficiary's chances for finding employment by 29%[42].

As we can see in Figure 3, the difference between the selected control (counterfactual) group and the group of persons not participating in the Phare programme is approximately 5%, regardless of time that elapsed since programme completion. This difference emphasizes two aspects of the Phare programme evaluation we undertook. First, it brings us closer the potential bias of the estimation of programme effect, if such estimation was based on a simple comparison of the programme beneficiaries with non-beneficiaries, and thus it shows the benefits resulting from the application of the PSM procedure.

[41] On the basis of: PBS DGA, *Ewaluacja ex-post komponentu regionalnego...*, op. cit.

[42] Ibidem.

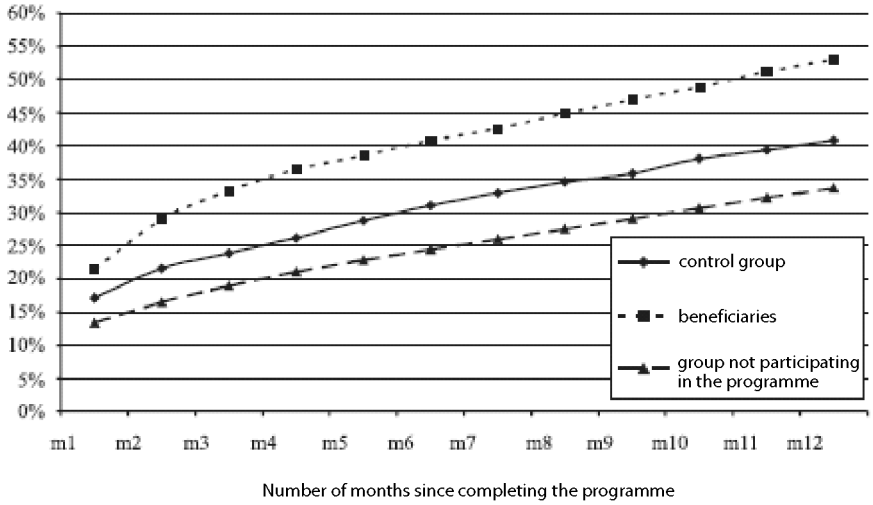


Fig. 3. Percentages of employed among beneficiaries and in the control group of Subproject 1 Phare 2002 SEC HRD

Source: Authors' own work.

Second, the difference between the control (counterfactual) group and the whole group of individuals not participating in the programme indicates, that most likely, the estimated net effect as the difference between the beneficiaries and the control group is only an upper limit of the programme effect. Unfortunately, the lower limit of this effect is unknown because we do not know what would be the net effect if the CIA assumption was completely fulfilled and the control group was fully matched with the beneficiary group with respect to variables (measurable and immeasurable) relevant for the selection process and the programme outcome.

Summary

In this chapter we presented the potential of the PSM method in ex-post evaluation. The PSM method allows for selecting a comparative group from a larger group of potential control units on the basis of estimated individual propensity to participate (propensity score) in a social programme. When the group of programme beneficiaries is significantly different from the group of individuals not participating in the programme, the PSM method offers an effective way of correcting bias in the estimates of the net programme effects. The quality of the corrected estimate of a programme net effect depends on the quality of matching the control group, and the extent of meeting the assumptions that constitute the foundations of the PSM method. Only full understanding of the process that leads to programme participation and programme outcome allows for adequate definition of important, measurable characteristics, which then can be used in models which estimate individual propensity scores. Of course, the data on these characteristics must be available for the evaluators of intervention programmes.

Roman Konarski – is a graduate of York University (Toronto, Canada), where he obtained a title of doctor Ph.D. of Psychology, specialisation in psychometrics and statistics. He lectured on statistics at York University, Queen's University, and University of New Brunswick. Currently he is the senior lecturer in the Institute of Psychology at the University of Gdańsk, and chief statistician in PBS DGA (formerly The Institute of Social Research in Sopot). His academic interest concentrates on applications structural equation models to the problems of psychological measurement. In PBS DGA he deals with methodology and statistical issues in market research. He is the author of many works on psychometrics and statistics.

Michał Kotnarowski – is a graduate of the University of Warsaw, where he obtained M.A. in Sociology. Presented analyses were conducted when he worked at PBS DGA (formerly The Institute of Social Research in Sopot) as a consultant – manager of research projects. Currently he is working in the Institute of Political Studies, Polish Academy of Sciences. He has many years of experience in conducting projects regarding public opinion research on current social and political issues (including measurement of parliament and presidential preferences), ethnographic research, evaluation of projects financed by the European Union funds.

Sources

Cox D.R., Snell E.J., *Analysis of Binary Data, Second Edition*, Chapman and Hall, New York 1989.

Dehejia R.H., Wahba S., *Propensity score-matching methods for nonexperimental causal studies*, The Review of Economics and Statistics 2002, nr 84{1), p. 151–161.

Heckman J.J., Hotz V.J., *Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training*, Journal of the American Statistical Association 1989, No 84, p. 862–880.

Heckman J.J., Ichimura H., Todd P., *Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme*, Review of Economic Studies 1997, No 64, p. 605–654.

Heckman J.J., Smith J., Clements N., *Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts*, Review of Economic Studies 1997, No 64, p. 487–537.

Heckman J.J., Ichimura H., Todd P., *Matching as an econometric evaluation estimator*, Review of Economic Studies 1998, No 65, p. 261–294.

Heckman J.J., LaLonde R.J., Smith J.A., *The economics and econometrics of active labor market programs*, in: O.C. Ashenfelter, D. Card (ed.), *Handbook of Labor Economics*, vol. 3, p. 1865–2097, Elsevier, Amsterdam 1999.

Holland P.W., *Statistics and causal inference*, Journal of the American Statistical Association 1986, No 81{396), p. 945–960.

Hosmer D.W., Lemeshow S., *Applied Logistic Regression*, Second Edition, Wiley, New York 2000.

LaLonde R.J., *Evaluating the econometric evaluations of training programs with experimental data*, American Economic Review 1986, No 76, p. 604–620. Menard S., Applied Logistic Regression Analysis, Second Edition, Sage Publications, Thousand Oaks 2002.

PBS DGA, *Ewaluacja ex-post komponentu regionalnego programu Phare 2002 Spójność Społeczna i Gospodarcza – Komponent Rozwój Zasobów Ludzki*ch, Report from research conducted at the request of the Polish Agency for Enterprise Development, Polish Agency for Enterprise Development, Warsaw 2006, http://www.parp.gov.pl/index/index/110.

Rosenbaum P.R., *The consequences of adjustment for a concomitant variable that has been affected by the treatment*, Journal of the Royal Statistical Society 1984, series A, No 147, p. 656–666.

Rosenbaum P.R., *Constructing matched of sets of strata*, in: *Observational Studies*, p. 200–224, Springer-Verlag, New York 1995.

Rosenbaum P.R., Rubin D.B., *The central role of the propensity score in observational studies for casual effects*, Biometrika 1983, No 70(1), p. 41–55.

Rosenbaum P.R., Rubin D.B., *Constructing a control group using multivariate matched sampling methods that incorporate the propensity score,* The American Statistician 1985, No 39(1), p. 33–38.

Rosenbaum P.R., Rubin D.B., *The bias due to incomplete matching,* Biometrics 1985, No 41, p. 103–116.

Rubin D.B., *Estimating causal effects of treatments in randomized and nonrandomized studies*, Journal of Educational Psychology 1974, No 66, p. 688–701.

Rubin D.B., *Assignment to treatment group on the basis of a covariate*, Journal of Educational Statistics 1977, No 2, p. 1–26.

Rubin D.B., *Bayesian inference for causal effects: The role of randomization*, Annals of Statistics 1978, No 6, p. 34–58.

Rubin D.B., Thomas N., *Matching using estimated propensity scores: relating theory to practice*, Biometrics 1996, No 52, p. 249–264.

Winship C., Morgan S.L., *The estimation of causal effects from observational data*, Annual review of Sociology 1999, No 25, p. 659–706.

Zhao Z., *Sensitivity of Propensity Score Methods to the Specifications*, Discussion Paper No. 1873, Institute for the Study of Labor, Bonn 2005.

Dominik Batorski

Network analysis: methods and application in evaluation

Evaluation research uses numerous data collection techniques. However, in practice the type of collected and used data depends - to a great extent - on the ability to analyse them. The purpose of this chapter is to present social network analysis methods, also referred to as network analysis. As opposed to classic analysis methods used in quantitative research focussing on attributes (variables) of individuals, their properties and interdependencies, network analysis requires the examination of relations between the individuals themselves. It is significant insofar as everyone, including companies and organisations, functions in a network of relations with other persons and organisations. These networks can be of key importance for the individuals' achievements, institutions' success and specific projects. The research indicates that the nature of networks in which individuals are involved exerts influence on virtually everything - ranging from health[1], through professional success[2], to identity. Similarly, the situation of organisations depends strongly on the environment in which they operate.

By using the data regarding relations, network analysis allows examining the structure of relations between individuals, but also the dependence of structure on individuals' attributes and the influence on the processes that take place through relations (transactions, information flow, and cooperation). These methods are used not only in sociology, but also in many other fields. In organisational research, the analysis of communication structure in an organisation might allow to identify informal relations, the significance of particular individuals for information and knowledge flow, as well as to identify informal leaders[3]. Network methods in criminology are applied in order to determine the relations between objects and to combat organised crime. In the financial sector (banks, insurance companies), the analysis of relational data allows for the detection of fraud. In epidemiology and pharmacology, it is used to examine disease and addictions spread and allows for efficient planning of eradication. These are only examples – there are much more fields of network analysis application since the study of relations and their structure is of great value in all areas where relations are of greatest importance.

This chapter constitutes an introduction into the network analysis methods. The first section contains basic definitions and presents the specific nature of the network approach comparing it to traditional studies. The second part is devoted to the research on networks, the methods of collecting data and the types thereof. The third part contains a description of basic analytical methods used in network analysis. Application of network approach in evaluation studies is presented in the final part of the paper.

---

[1] Persons that have close relations with other persons live longer. Those with more diverse social contacts are more resistant and fall ill less frequently.

[2] Relations and structure thereof affect the chances of finding a job and the frequency of gaining promotions.

[3] M.Cf. Kilduff, W. Tsai, *Social Networks and Organizations*, Sage, London 2003.

Networks and network approach

Basic definitions

In simplest terms, a network is a collection of objects that are interconnected by certain relations. A social network is usually a collection of persons interconnected by socially significant relations. However, in practice, network analysis is not limited to the research on social networks only and both objects and relations in a network may vary considerably. The individuals, i.e. network (graph) nodes, may comprise persons, ideas, publications, events, projects, organisations or even countries. As it is the case in many other research types, this depends primarily on the problem being examined. For example, a network exists in the case of communication between persons, as well as in trade and cooperation between companies or migration between countries. The selection of a population to be analysed, and thus the definition of the boundaries of an analysed network, can represent a serious problem. Before we proceed to the description of this problem, however, it may prove necessary to discuss the nature of relations in more detail.

Since a network is composed of individuals and relations between them, it is necessary to define not only the types of individuals, but also the relation types that are taken into consideration. Principally, network nodes may be interconnected by any type of relation. Relations analysed in a traditional social networks approach are frequently represented by: acquaintances, communication, cooperation, or perhaps advice-giving. Relational data may also include membership in the same organisation, participation in the same meetings and – between organisations – transactions, financial flows, cooperation etc.

From the point of view of social network analysis, several formal features of relations are of importance, as they are largely responsible for the selection of analysis methods. The principal issue is obviously the type of relations. Namely, which relation, what does it relate to, and how can it be determined as to whether such a relation occurs between two individuals. It will not suffice to say for instance "friendship relation" without giving the criteria of classifying a relation to such group[4].

Another problem is the possibility of defining the direction of a relation. A relation can be directed or undirected (symmetrical). In the former case, both the originator and the adresee are clearly defined. For instance, X is the parent of Y or X is the superior of Z. A symmetrical relation would e.g. be "being a family", X and Y are a family, or work relationship, X works with Z on the same project.

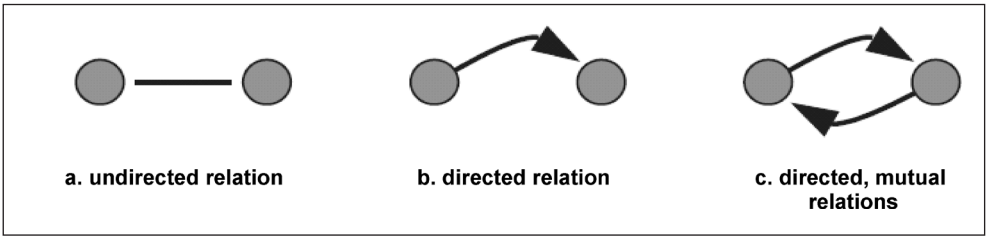


Fig. 1. Example relation types

Topological and metrical features of relations are also important. First of all, we can distinguish relations that are of yes-no nature – a connection either exists or does not exist, or relations for which it is

[4] It can be a bilateral statement and description of another person as a friend and it can be a category fulfilling defined conditions, such as spending leisure time together, helping each other, etc.

possible to determine the strength of connection (for instance the frequency of visits or communication, acquaintance duration, transaction value, etc.). Sometimes it can be a relation orientation, for instance positive-negative or like-dislike relation.

It should be pointed out here that the nodes are usually constituted by individuals of the same type – e.g. persons only or organisations only - thus a unimodal network. Studies relatively often cover also the networks in which the nodes belong to two or more types (bi- and multimodal relations). Bimodal networks are typically referred to as affiliation networks since the relations connecting the nodes are usually affiliation relations. The relations of membership of persons in organisations, participation in projects, and participation in events may serve as an example.

Sometimes it is possible to define a greater number of various relations for the community in question - multiple networks. These are the networks in which the same individuals are connected by means of various relations – for instance by means of cooperation under the same projects, spending leisure time together, discussing professional issues, etc. In the case of organisations these can be transactions, cooperation, personal ties (when the same persons belong to more two or more organisations), joint property. One option would be to consider each of these networks individually, but there are methods of analysing such multiple relations together. Multiple connections are usually stronger and more durable.

The methods that can and should be used to analyse a given network depend, to a great extent, on the formal network features (individuals and relation between them) that are described herein. Other measures can be used when the relations have a defined direction, value, and other when the relations do not have such a nature. All of the foregoing, as well as the necessity of applying separate computer programmes dedicated to network analysis, makes these methods seem very complicated to budding researchers.

Examination and importance of the structure

Social network analysis methods are widely used in numerous disciplines applying studies on various types of phenomena: study of the attitude shaping, social and economic mobility, diffusion of innovation and information, communication, community structures, companies, organisations, interpersonal relations and political behaviour. However, they typically serve to study the structure of a network itself. According to the dictionary definition, structure is "a collection, system of multilaterally related elements and/or a system of relations occurring between the parts or elements of a system, a certain entirety". Another definition of "structure" provides a slightly more comprehensive explanation, saying that it is constituted by regularities in the relation patterns between individuals[5]. Social network analysis aims primarily at the discovery of an interaction pattern between persons, organisations, countries, etc. It serves the purpose of indicating the pattern of social structures and the influence exerted by such structures on the variables and the changes of the structure itself.

Network analysis is based upon the conviction that patterns of relations between persons and organisations are not accidental and that the way persons behave depends largely on the relations between them and other persons (and the characteristics of those persons). Frequently, the success of organisations, communities and even whole societies depends on the pattern of their internal structure. The dependence of work and efficiency of work teams on their structure seems a good example here[6].

[5] D. Knoke, J.H. Kuklinski, *Network Analysis*, Sage 1982.

[6] J.N. Cummings, R. Cross, *Structural properties of work groups and their consequences for performance*, Social Networks 2003, No. 25(3), p. 197–210.

In addition, the functioning of social mechanisms – of social influence, information flow, diffusion of innovation for instance – is related to the quantity and quality of social contacts and communication structure.

The social network analysis method allows for the examination not only of the whole groups or isolated relations, but also various indirect dependencies. We can study the network density, isolate stronger subgroups, analyse whether it is a centralised network or a one diversified in its volume and cohesion, whether the relations have narrow specialisations or are multi-thematic, and how non-direct relations affect the behaviour of individuals. Social network analysis allows researchers to take more precisely into consideration the context in which the analysed relation and particular individuals are located.

Network approach

The origins of network analysis can be traced to the concept of sociometry created by Jacob Moreno[7]. The first significant development stage of network analysis methods took place in 1970s.[8.] Further development of this field took place in the course of the last decade or so. Firstly, methods emerged that allowed not only a network description[9], but also an analysis regarding the dependencies occurring in these networks[10] and the processes taking place within networks[11]. Secondly, mathematicians and physicists grew interested in networks and introduced a new perspective into social research[12].

Network analysis is, however, something more than just a set of measures and tools. It is the paradigm and approach to social reality and to studying social phenomena based primarily on the analysis of relations between objects[13]. In fact, network analysis is a separate research perspective within social sciences. It focuses much more on the relations between the individuals interacting with one another. There are also quite a lot of sociological and psychological theories related to the network concept and explaining the social phenomena with a reference to relational and structural factors[14].

---

7 J.L. Moreno, *Foundations of Sociometry: An introduction*, Sociometry 1941, No. 4(1), p. 15–35.

8 P.W. Holland, S. Leinhardt (ed.), *Perspectives on Social Network Research,* Academic Press 1979.

9 S. Wasserman, K. Faust, *Social Networks Analysis: Methods and Applications. Structural Analysis in the Social Sciences*, Cambridge University Press 1994; J. Scott, *Social Network Analysis: a handbook*, Sage Publications, London 2000.

10 T.A.B. Snijders, P.E. Pattison, G.L. Robins, M.S. Handcock, *New specifications for exponential random graph models*, Sociological Methodology 2006, p. 99–153.

11 G. Robins, P. Pattison, *Random graph models for temporal processes in social networks*, Journal of Mathematical Sociology 2000, No. 25, p. 5–41; G. Robins, P. Pattison, P. Elliott, *Network models for social influence processes*, Psychometrika 2001, No. 66, p. 161–190; T.A.B. Snijders, *Models for Longitudinal Network Data*, in: P. Carrington, J. Scott, S. Wasserman (ed.), *Models and Methods in Social Network Analysis*, Cambridge University Press, New York 2005, Chapter 11.

12 D.J. Watts, *Networks, dynamics, and the small-world phenomenon*, American Journal of Sociology 1999, No. 105(2), p. 493–527; D.J. Watts, *Six Degrees: the Science of a Connected Age*, W. W. Norton 2003; A.-L. Barabási, *Linked: How Everything is Connected to Everything Else and What it Means for Business, Science and Everyday Life*, A Plume Book 2003; M.E.J. Newman, *The structure and function of complex networks*, SJAM Review 2003, No. 45, p. 167–256.

13 D. Knoke, J.H. Kuklinski, op. cit.; B. Wellman, *Structural analysis: from method and metaphor to theory and substance*, in: B. Wellman, S.D. Berkowitz (ed.), *Social Structures: a Network Approach*, Cambridge University Press 1988, Vol. 15 series *Contemporary studies in sociology*, Chapter 2, p. 19–61; S. Wasserman, K. Faust, op. cit.; R.A. Hanneman, M. Riddle, *Introduction to social network methods*, University of California, Riverside 2005. Published electronically on: http://faculty.ucr. edu/~hanneman/.

14 P.R Monge, N.S. Contractor, *Theories of Communication Networks*, Oxford University Press, New York 2003.

By studying social phenomena, both the features of individuals and the relations between them can be observed. Let us consider an example: a social group is discussed in the categories of the similarity of individuals or in the categories of relations between the individuals. In the former case, e.g. in a school, students from the same class can be perceived as a group. In the latter case, students would be recognised as a group due to the relations existing between them, not alone, due to the fact that they belong to one class. In many situations, the attributes themselves that define the group affiliation will not suffice. The origin of such an approach can be found in the idea of Georg Simmel. While studying social phenomena, we must not analyse only "features of individuals separated from their multilateral relations"[15].

Social studies, including a substantial part of evaluation studies as well, require both possible approaches to be taken into consideration – both the attributes of individuals and the relations between these individuals. According to Barry Wellman "the structure of social relations is a much stronger source of sociological explanations than the attributes of persons belonging to a given community"[16]. Network approach provides that there is a range of features that differ from the traditional approach:

- Behaviour is interpreted in the categories of structural limitations imposed on the action and not on the volunteer basis – in terms of internal forces encouraging an individual to operate and achieve a goal.
- Analyses concentrate on the relations between individuals instead of grouping them into categories identified based on their attributes.
- The main objective consists in the method of influencing the behaviour of actors found in the network by patterns of relations between other individuals. Therefore, relations are treated in isolation (from other relations). Network approach is significant since the course of multiple social processes depends on the relations between individuals and structures created by the relations.
- The structure is perceived as a network of networks that can, but does not have to, be divided into separable groups. There is no need to assume that consolidated and precisely isolated groups are the integral components of the structure since not all networks need to have the structure of a group.
- Analysis methods are concerned directly with the patterns and relational nature of the social structure. Sometimes they replace and sometimes supplement the mainstream statistical methods, which require independent individuals to be analysed.

Organisation of network analysis

Selecting the community to be analysed

One of the first stages of social network analysis consists in defining the borders of the studied structure and thus the type of individuals comprising a network and relations between them, which will be analysed. This is not a problem in the situation when a community, group or organisation is small and well defined. The identification of studied individuals is much simpler also when network nodes are constituted by projects under a larger programme, or companies or organisations. The situation is much more difficult when the population under research is very large. The problem lies not only in the fact that it is difficult to reach every person, but also rather in identifying all existing relations. Asking - even

[15] J. Szacki, *Historia myśli socjologicznej*, Warsaw 1981, p. 512.

[16] B. Wellman, op. cit., 1988, p. 31.

a single person - about their relations to every other person of the examined population would be very time-consuming. Quite often, it is much easier to determine affiliation networks for large networks.

Network data and collection thereof

Apart from the abovementioned problem of outlining the borders of the studied structure, many other problems regarding the collection of network data may also occur. Specific difficulty may be encountered in the study on large communities. In the research on social networks, the entire network is subject to analysis, if possible. Information on the relations with all other persons is collected from all persons, which allows for a precise determination of the position of each person in the network. Since such a research is possible only when the network is relatively small and has clearly outlined borders, such a method of collecting data is the most adequate one for the study of a small population. An exception is the situation when previous data regarding all individuals belonging to a community that we want to study and relations between them may be used.

A partial solution to the problem of research on large populations is to analyse only some relations – the most important ones. For large communities, instead of questions about the relations with all persons, (the number of questions is proportional to the number of population members), respondents may be asked to list all friends or list the three most important ones. Similarly, in surveys conducted within large organisations, the respondents are asked not about all the other persons – members of the organisation, but rather about the ones with whom he/she works, contacts in professional matters, from whom they seek advice most and others depending on the range of issues under the study. However, this research method has disadvantages too, since it provides only the data on the most important, and thus the strongest, relations. The acquisition of information on the weaker links is in such a case impossible. This problem concerns particularly the studies in which weak relations are of huge importance – e.g. in the studies of social capital where weak relations are essential for the information flow[17].

Another possibility of avoiding difficulties concerning exhaustive research on a large population is to conduct it using a sample18. Examination of all relations between all persons is difficult in a large group, therefore the studies are carried out using samples. Such a solution is also better in networks with better-outlined borders. In sample-based analyses, the survey examines so-called egocentric networks. This type of research is closest to the traditional research. Data on particular respondents (ego) and persons to which the respondents are related (alter) are collected. Very often it is not only the definition of relations for the individuals under study, but also an estimation of relations between the contacts of a given person, made on the basis of questions asked to the respondent or on the basis of information acquired directly from his/her contacts. Such an approach is most often applied in research on very large networks when we have no previous data at our disposal and the research on the entire population and relations between all persons is not possible. In the case of research that covers egocentric networks and require reaching broader social groups (for instance acquaintances of acquaintances), we are dealing with analysis of partial networks. At the same time it is less labour-intensive than examining relations between all pairs of actors in a given community.

Good examples of network research carried out using samples are found in numerous studies devoted to social capital[19]. Here, a name generator technique is usually applied. An individual names the

[17] M. Granovetter, *The strength of week ties*, American Journal of Sociology 1973, No. 73, p. 1360–1380.

[18] B.H. Erickson, T.A. Nosanchuk, *Applied network sampling*, Social Networks 1983, No. 5, p. 367–382.

[19] E.g.: N. Lin, K. Cook, R.S. Burt (ed.)., *Social Capital: Theory and Research*, 2001; R.S. Burt, *Brokerage and Closure: An Introduction to Social Capital*, Oxford University Press 2005.

persons with whom he/she is in relation with and describes this relation. Another method is the position generator technique used in social capital research. Respondents are asked to state whether they know any persons occupying positions that provide access to resources valuable in a community[20].

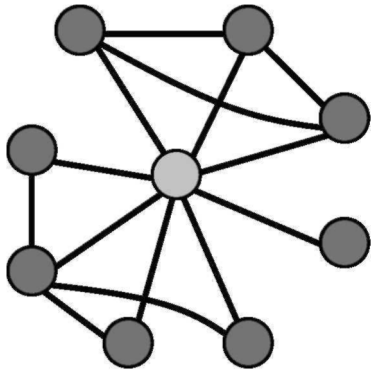


Fig. 2. An example of egocentric network

The type of collected and analysed data (the entire network vs. egocentric networks) is primarily related to the selection of the studied population and the type of studied network. On the one hand, the analyses that are possible depend, largely, on the type of data at one's disposal. On the other hand, many analyses possible for data concerning the entire network are not feasible when only the data on selected egocentric networks are available.

Due to the mentioned difficulties with the research on large populations, data on affiliation networks are collected relatively often therein. Such data are much more available as well. What is more, an affiliation network (and thus a bimodal network) can be reduced to two unimodal networks. For example, a network of affiliations of persons with projects can be replaced with a network of relations between projects by inclusion of the same persons and with a network of relations between persons where a relation is constituted by affiliation with the same projects.

Another particularly important problem of network research is the group of sensitive data, which often includes information on relations. It concerns both many interpersonal relations and relations between companies (for instance concluded transactions and the value thereof). The tendency to submit such data is small and additionally, before commencing the research, the question of ethics should be considered. In the case of network research, especially the ones in which a small community is under research, it is much harder to guarantee anonymity of the surveyed persons.

Firstly, when the presented results concern particular persons and when they are in graphical form, the persons belonging to the population under research will often be able to realise who is who. Therefore the researcher should exercise due diligence to prevent the research procedure and presented results from harming the individuals and relations between them.

It should be pointed out as well that the absence of data in network research is very dangerous. In comparison with regular quantitative research, each absence of data is of much greater importance.

[20] N. Lin, Y. Fu, R.-M. Hsung, *The position generator: measurement techniques for investigations of social capital*, in: N. Lin, K. Cook, R.S. Burt (ed.), op. cit., p. 57–81.

It causes the loss of data on relations and the consequence thereof is simultaneously distortion of information on a substantial number of persons that participated in the survey and that could be in a relation with individuals that have not been reached. Absence of data makes it impossible for instance to precisely determine the level of relation mutuality and many other, more complex analyses. More information on sample selection in the research on social networks is included in the works by Erickson and Nosanchuk[21] and Frank[22].

Social relations existing in various communities are often misperceived. The way of perceiving relations by individuals is influenced by their personality[23] and by the fact how close to the relations they are and how well they know them[24]. The differences between the perception and real communication have consequences for the quality of collected data and therefore should exert influence on the methods of collecting them[25]. It is a further argument, apart from the inconveniences concerning the sample selection and threats related to the absence of data that limits the research conducted by means of interviews.

Data collection methods in network analyses can be very diverse and do not have to be limited to interview or surveys. The other used methods include: observations, logs and previous data from various archives, records and registers. In practice every method of data collection can be used also in network research. Very often multiple, various methods are combined.

Network analysis methods

Social network analyses can be carried out in various ways. Firstly, various levels of network analysis are possible. The level of individuals, i.e. of egocentric networks, is concentrated on the relations of individuals and related persons. The next level is the level of subgroups isolated in a network and the final one – the level of the entire network. In network analysis, there are a lot of methods and possibilities of analysing the properties of relations in their structures. A significant research element is the visualisation of network data. Networks are presented by means of graphs composed of nodes and edges connecting them. Currently graphs are increasingly becoming a method of presentation and initial exploration of data rather than analyses[26].

Only the most basic issues analysed by means of network data will be described below. In order to explain the described methods in a better way, I will refer to simple examples of networks. Several applications of these methods presented below will refer to networks of cooperation between organisations whose operations concern similar issues. It would be possible to use them e.g. in evaluation of training projects aimed at the promotion of clustering as a cooperation form between enterprises and public partners with the third sector[27]. The goal of such projects is: to create durable cooperation forms

---

[21] B.H. Erickson, T.A. Nosanchuk, op. cit.;

[22] O. Frank, *Network sampling and model fitting*, in: P. J. Carrington, J. Scott, S. Wasserman (ed.), op. cit.

[23] T. Casciaro, *Seeing things clearly: social structure, personality, and accuracy in social net- work perception*, Social Networks 1998, 20(4), p. 331–351.

[24] D. Krackhardt, M. Kilduff, *Whether close or far: Social distance effects on perceived balance in friendship networks*, Journal of Personality and Social Psychology 1999, No. 76, p. 770–782.

[25] P. V. Marsden, *Network data and measurement*, Annual Review of Sociology 1990, No. 16, p. 435–463.

[26] W. de Nooy, V. Batagelj, A. Mrvar, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press 2005.

[27] An example of such project is *Turystyka wspólna sprawa*, cf. www.tws.org.pl, http://www.parp.gov.pl/index/more/24. However we can imagine any other example of organisations with other goals, e.g. organisations operating for the promotion of healthy lifestyle or institutions whose operations are of importance for the national informatisation.

oriented towards the creation of local communities developing a common brand and maintaining multi-dimensional, intensive contacts (e.g. The Aviation Valley in the Podkarpacie region of Poland).

Network characteristics of individuals

One of the main questions concerning social network analysis is the question about the position of an individual. The frequently formal structure of persons in an organisation or organisations in larger projects does not correspond to the real one and informal relations and hierarchies are of immense importance for the functioning of particular individuals and entire organisations. Network analysis allows for the description of individual's position in the network structure on the basis of their relations with the other individuals. It allows answering the question which individuals are more central and which are located at the peripheries – which are more important and which are less. What is the importance of the position in the network for the access to information and control of information flow (broker position)? What is the importance of the position in the network for the access to opportunities and possibilities? Which individuals gain competitive advantage in the structure due to their position[28]? Which individuals have more power than the other does? All these questions concern the positions of individuals in relation to other persons.

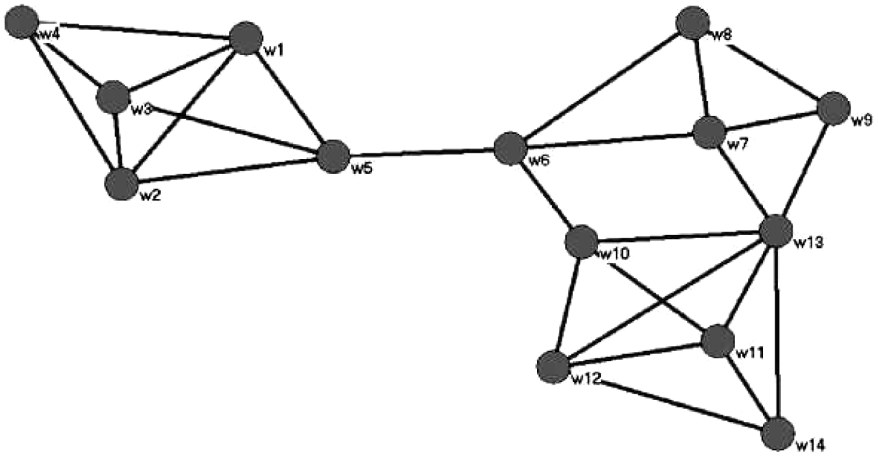


Fig. 3. An exemplary network composed of 14 nodes and with a diverse position

The simplest characteristic of position is the degree of relations of a given individual. The nodes in Figure 3 have 3-6 relations (node 13). Individuals with more relations generally are of greater importance and have more possibilities, although it is not a principle. In case when relations in the analysed network have a defined direction, it is also possible to make a differentiation according to the selections made by an individual and their partners. It is possible to determine the degree of relations initiated by an individual (outdegree) and the degree of incoming relations (indegree) whereas the latter is a good measure of popularity or prestige. In this case, while studying the number of indications received by a person (how often it is selected by others), we can talk about the level of prestige or popularity. It is also possible to take into consideration the fact how popular the selecting persons are. In such a case, the fact

[28] Por. R.S. Burt, *Structural holes: the social structure of competition*, Harvard University Press 1992; R.S. Burt, *Brokerage and Closure...*, op. cit.

of being selected by a popular person might be much more significant than being selected by several persons that have not been selected by anyone.

In the case of directed relations, there are four available direct connections between two nodes (often referred to as dyads). The nodes can be linked, a relation can be non-mutual (outgoing or incoming) or it can occur in both directions (mutual). In many networks, the level of relation mutuality is an important network characteristic. However, with the data on the entire network we have the possibility of analysing not only the dyads that are linked by means of direct relations, but also those that are linked by other individuals. As far as the network is connected (more information on this in the following part), the determination of distance for each node pair is possible by means of relation paths and other nodes. In order to measure the distance, generally the shortest paths, called geodesic distance, are used. The longest of the shortest paths in a network are called the "network diameter". In the case of graph presented in Figure 3, the nodes that are most distant from each other are nodes 4 and 14 while the diameter amounts to 6.

Positions in the network that relate to power, greater access to information and greater possibilities of exerting influence are characterised by means of the term "centrality". There are many measures of centrality[29], but in practice all of them are based on various concepts of individual's involvement in relations in a network. The greater the involvement in all network relations, the greater the centrality of a given actor. The relations themselves can be defined in various ways. The above-mentioned relation degree is the simplest measure of individual's centrality in a network, therefore it is sometimes referred to as "actor centrality". Sociometric star is a typical example of a central individual. However, this measure relates to direct involvement in relations only. Regard to indirect relations in the determination of node position in the network causes the isolation of more complex indices of the node position centrality, that is, however, of greater informational value. It is worth pointing out that most of these measures require regard to the entire network and not only the relations of an individual.

For example, "betweenness centrality" is the frequency of occurrence of a given individual in the shortest paths between various node pairs. These are potentially persons with greater control of information flow and influence on the others. In Figure 3 the individual of highest value of this index is node 6 that controls communication between many network fragments. Node 5 is almost equally important, followed by 10, 13 and 7. The differences between nodes 6 and 13 should be pointed out; the former has only four direct connections, while the second has six, but it node 6 that has a much more important position.

Another measure of centrality is "closeness centrality" defined by the distance of an individual from all the others in the network. Freeman recognised this as a measure of independence from the others and immunity to being influenced[30]. In this respect, the position of node 6 is best as well, but the position of nodes 7 and 10 is not much worse as their average distance from other nodes in the network is lesser than in the case of node 5 and the remaining individuals.

It should be also noted that some authors, *inter alia* Scott[31], differentiate local and global centrality. This differentiation is of use particularly for large social networks in which the number of nodes is much higher than the average relation degree. Local measures indicate the node position in its surrounding (neighbourhood) whereas the global ones – in the entire network.

---

[29] L.C. Freeman, *Centrality in social networks: I. conceptual clarification*, Social Networks 1979, No. 1, p. 215–239; S. Wasserman, K. Faust, op. cit.

[30] L.C. Freeman, op. cit.

[31] J. Scott, *Social Network Analysis: a Handbook*, Sage Publications, London 2000, p. 82.

Additionally, calculation of the corresponding measure of centralisation of the entire network is possible for every centrality measure. Centralisation measures assume the values ranging from zero to one and determine how a given network is centralised in relation to the maximally centralised network. Because of that, it is possible to compare various networks.

The importance of the node position in the network can also be based on characteristics other than centrality measures described herein. Both presented approaches can give similar results since the properties described herein translate into higher centrality of nodes. Some persons in the network can have essential meaning for information flow when they are links between various network fragments. The position of such a broker is occupied by nodes whose removal from the network would result in the disconnection of most network fragments (in graph theory they are referred to as cutpoints). In the network presented in Figure 3, nodes 5 and 6 represent such nodes. The connection between two groups can be a single relation between two individuals (relation between nodes 5 and 6 in Figure 3). Such relations whose removal from the network causes its disconnection are called bridges. To summarize, it can be said that the importance of particular nodes and relations for communication, stability and efficiency of the processes taking place in the network and other processes, depending on the network type, can be established by means of analysed effects of removal of a given node/relation from the existing structure.

Particular individuals can be described also by means of measures characterising their surrounding in the network. An example of such a measure is ego-network density[32], i.e. the proportion of connections between nodes linked with an individual. For example in Figure 3, almost all individuals that are in relation with node 11 have a relation between each other (ego-network density amounts to 5/6); it is different in the case of nodes 6 and 7.

Analysis of characteristics of individuals' position in the network could be successfully applied in the above-mentioned example of analysing the network of cooperation between organisations. This would enable answering the question of which organisations working on the implementation of a larger problem are more important for its implementation. Which of them cooperate with more partners, the position of which is more important for an efficient cooperation of multiple individuals, and which individuals are indispensable during the programme implementation? The length of paths between individuals allows to determine to what extent the flows of knowledge and experience are possible between particular organisations. Carrying out the research in time would additionally allow us to verify the durability of the network in time.

Group structures

Most measures applied to the description of the positions of individual persons can be applied to the description of subgroup properties in the network after small modifications. Such subgroups can be defined on the basis of attributes of individuals (e.g. the gender in the case of persons, branch in the case of companies, etc.). It is possible, however, to isolate subgroups and divide networks based on relation patterns. From this perspective, a group is a group due to the specific structure of relations between the individuals comprising such structure and not only due to having a certain feature or characteristic by an individual.

[32] Ego-network density is sometimes referred to using different terms. In the applied literature there are several terms, inter alia "cohesion" and "clustering".

In the mentioned example, we could isolate organisation groups both on the basis of their attributes (e.g. companies, administration institutions, non-governmental organisations) and on the basis of the fact who cooperates with whom, isolating closely cooperating groups in this manner. The fact to what extent an organisation cooperation network is of a group nature can be of essential importance for the success and efficiency of the entire programme being implemented by them. Sometimes the creation of durable cooperation forms between organisations is the goal in itself. In both cases the isolation of groups of cooperating organisations, possible due to network method, is necessary. The following questions are also important: are organisational activities independently undertaken by groups, each of which forms a closed clique that is uncooperative and does not exchange experiences with others? Does the cooperation include much more entities and is the isolation of organisation groups not possible?

However, before we proceed to the problem of isolating groups within the network, we should discuss some more basic concepts related to the division of nodes belonging to the network into fragments. First of all, one should ask themselves a question whether the network in question is connected, that is whether each pair of nodes is linked by means of relation path. If not, it means that the network is composed of more than one component. A component is such a maximal group of nodes for which it is possible to link any pair of nodes in this component by means of relations and other nodes belonging to the same component. There are various types of components that are distinguished according to the requirements against the type of relations connecting the nodes[33].

In network analysis, two basic approaches and arguments regarding communities and groups within a network are possible as structures are isolated on the basis of empirical data – connections between individuals. One is social cohesion: groups are isolated because of the links between particular players. The other is constituted by structural equivalents: when the groups are identified based on contact similarity with the others. Obviously, these approaches do not exclude the isolation of groups on the basis of other non-network attributes. At this point, we will discuss the isolation of cohesive subgroups on the basis of social cohesion whereas structural equivalents will be discussed in the subsequent part of this chapter.

At the local level, network cohesion can be considered by analysing relations in triads – i.e. between three individuals since only in the case of three individuals we deal with a group. In theory, 16 various relation combinations in triads are possible which asks what type of triads and how often they occur. Based on such information (so-called triad census) it is possible to precisely analyse the transitionality of relations, as well as the network cohesion and fragments thereof[34]. Higher transitionality of relations, i.e. higher probability that two individuals that are in relations with a third one are also in relation with each other, occurs in structures in which divisions into groups are stronger.

Cohesive subgroups are made up of actor collections between which there are relatively strong, direct, intensive, frequent or positive relations. However, due to the fact that the notion of social group can refer to many diverse properties of social networks, there are also numerous definitions of social network subgroups. The basic properties in conceptualisations of the notion of subgroups are the following:

- mutuality of relations,
- closeness and reachability of subgroup members,

[33] Weak components are concerned when the direction of a relation is not taken into consideration and the strong ones when for each pair of nodes A and B it is possible to go through the path of directed relations both from A to B and from B to A. We are talking about recurrent components when the paths go through mutual relations.

[34] F.Cf. Heider, *On balance and attribution*, in: P.W. Holland, S. Leinhardt (ed.), *Perspectives on Social Network Research*, Academic Press 1979, p. 11–23.

- frequency of relations between group members,
- relative frequency of relations between group members in comparison with the relations with other individuals.

The most important approaches to the isolation of cohesive subnetworks are derived from it:

- complete mutuality,
- reachability and distance,
- degree of relations (direct connections with sufficiently many subgroup members)[35].

There are many formal definitions of subgroups. The selection of one of them should be justified by the nature of the network and relations connecting the nodes. One of the basic definitions of a group with high cohesion within a network is a clique. It is the maximal[36] complete subgraph and thus a group of completely connected nodes. In other words, each individual has to have relations with all the other ones belonging to the same clique. Cliques in a network (especially in a larger one) can be convergent with each other, therefore the division into subgroups is not a separable division and one node can belong to more than one clique. In the network presented in Figure 3, the largest cliques contain four nodes each – e.g. nodes {1, 2, 3, 4} or {1, 2, 3, 5}. There are also many cliques containing three individuals.

Due to the requirement of completeness, the definition of cliques is particularly restrictive and it lacks immunity to measure errors. Absence of a single relation causes that a subgroup is not a clique and, what more, the volume of subgroups is limited by the degree of actors. There are practically no (structural) differences between the clique members and therefore the examination of an internal structure is not possible. Hence various other definitions assume that it is not necessary for a relation to exist between each pair of group members. Other definitions take also the distance between persons and proportions of relations within a group (connections are dense inside of a group and rare outside of it) into consideration.

Based on a reachability concept, we could define a group as a collection of actors that are closely related (on the basis of a given relation). The notion of n-clique derived from this concept is the maximal subgraph in which each clique element can be reached in n steps. If n = 1, we are dealing with regular cliques. The most commonly applied distance is n = 2. According to this concept, not all paths have to go through persons from a given n-clique and it can occur that the distance n is only through the nodes belonging to another subgroup. Therefore the subgroup diameter can be larger than n. It looks different in the case of n-clans, i.e. n-cliques in which all distances cross the paths within the subgraph. According to these definitions, it would be possible in graph in Figure 3 to isolate inter alia such node groups as {1, 2, 3, 4, 5} or {6, 7, 9, 10, 11, 12, 13}.

The distance itself is not the best indicator of cohesion (a good example in this case is the second above-mentioned subgroup). Therefore, sometimes other definitions are introduced. The definitions of k-plex and k-cores are used relatively often. In the former case, each node belonging to a subgroup containing g nodes has to be directly linked with at least g – k other members thereof. Thus, 1-plex is a group in which everyone has relations with everyone. However, the definition of k-cores requires each node in a subgraph to be connected with at least k other nodes from this subgraph. Instead of imposing requirements concerning the number of edges that can be absent, it is required to have a given number of existing relations. 3-cores in Figure 3 are only groups {1, 2, 3, 4, 5} and {10, 11, 12, 13, 14}.

Another approach to the isolation of cohesive subnetworks is possible when the nodes in the analysed network are connected by means of relations with diverse strength. In such cases, it is possible

[35] This definition is particularly important when direct contacts are significant.

[36] Meaning the largest available one.

to apply the definition of maximally strong components (often referred to as m-slices as well). It is the maximal subgraph that all nodes are connected by means of relations with strength m or higher.

Positions and roles within networks

A different approach – than the above-mentioned ones – to the isolation of subgroups within a network consists in distinguishing social positions and roles. Individuals occupying the same positions in a network are in the same or similar way connected with the entire social system, but they do not have to have relations with each other. Structural equivalents are the subgroups within a network that are defined by means of pattern of relations connecting actors with each other. Two nodes are structurally equivalent if they have precisely the same relations with all other nodes in the network. In other words, networks from which one of those nodes and its relations had been removed would be the same.

Figure 4 presents four groups of structurally equivalent individuals. It is worth pointing out that two positions can be identified in this situation, although nodes 1-5 form a clique since nodes 4 and 5 are additionally related with nodes 6 and 7. The two latter ones also have the same positions despite the fact that they are not related to each other.

In practice, in larger networks, such situations occur relatively rarely. Therefore a certain measure of probability of relation patterns is defined for each node pair. It can be for instance a number or percentage of individuals with which the given nodes have relations. By having the data on a matrix of distance for all node pairs, methods of hierarchical clustering are applied and the network is divided into subgroups.

A slightly different method, recently applied less frequently, is the CONCOR method[37]. It makes it possible to divide networks into two position subgroups, with each being capable of being further divided. The method consists in the creation of correlation matrix from the matrix of relations between individuals. Multiple repetitions of the same operation causes that only +1 values between the nodes remain in a matrix that are in the same group and –1 values for cross-group relations.

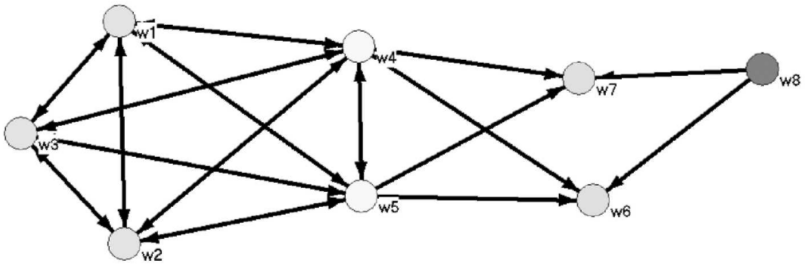


Fig. 4. Positions in a network – four structurally equivalent groups of individuals (directed relations)

Isolation of subgroups within a network is not a goal in itself. The network between individuals can be reduced and the dependencies between groups can be isolated in a separable manner. An example of such a reduction is presented in Figure 5, in which relations between positions from the previous example have been presented. The first group comprises nodes 1-3 that have relations between each other and with nodes 4 and 5. Nodes 6 and 7 do not initiate any relations, but they receive relations from nodes 4, 5 and 8. Apart from the reduction of the entire network, it is possible to analyse the functional context of

[37] D. Knoke, J.H. Kukliński, op. cit.

particular subgroups by means of reduction of the remaining subgroups, but without reducing a given group. Such methods are applied particularly in the analyses of large networks[38].

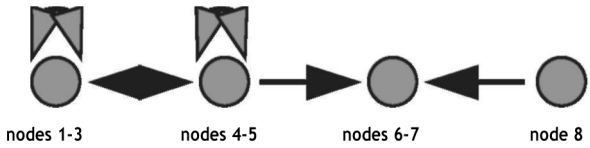


Fig. 5. Relations between positions identified in Figure 4

In the example of research on the network of cooperation between organisations, it would be possible by means of these methods to isolate organisations that perform the same functions in the project and their roles are interchangeable to a certain extent. It would be possible to identify the groups of individuals that are more or less involved in the implementation of a given programme, which may constitute the first step towards the recreation of informal hierarchies. The methods mentioned above could also serve the attempt to isolate the roles/types of involvement of various organisations – for instance such that undertake specific measures and those that finance the measures of others without getting involved directly.

Properties of the entire network

Many properties have already been mentioned at other analysis levels. One of the basic questions is whether the network is connected and, if not, how many and what type of components it is comprised of. If a network is composed of only one component, it is important what the diameter is and what the average distances between the nodes are.

A further simple, yet significant characteristic of network is relation density. Network density can be calculated based on the number of persons and degree of relations. It constitutes the proportion of existing relations to all possible ones (the degree of possible relations can be calculated based on the quantity of persons N and it amounts to N*(N.) for directed networks. Network density describes the extent of network's cohesion – its value is high in a situation when all nodes create one group in which the nodes are strongly related. On the other hand, the value of density is small when the nodes are weakly related to each other or when the network is divided into many groups that are strongly connected internally, but weakly with each other. Another measure, supplementing the information provided by the value of density for networks in which relations have a defined direction, is network cohesion. It is the proportion of mutual relations, i.e. the number of mutual selections, divided by the maximal number of such selections.

The entire network can be also characterised by its centralisation, calculated as a difference between centrality of the most central individuals and all the remaining ones. These differences are then referred to the maximal possible sum of differences for a network with the same number of nodes. This measure shows the extent of network's organisation around the most central nodes and the extent of its structure's diffusion.

An important characteristic of a network may be also the distribution of particular node features, the distribution of relation degree in the simplest case. It appears that this distribution in numerous networks

[38] W. de Nooy, V. Batagelj, A. Mrvar, op. cit.

can be described by power dependency, that is, there are many individuals that have very few relations and very few individuals that have very many relations[39].

Particular network measures may be of importance while comparing the structure of various networks between each other. However, many network characteristics are strongly dependent on the number of individuals composing it, therefore the comparison of networks that differ a lot in terms of volume is not advised.

Dependency of individuals' traits and relations in network

The network analysis methods described above were concentrated on the network structure itself, using almost exclusively the information on relations between individuals. However, in many studies, it makes more sense to take into consideration both the relations and the attributes of nodes. The only use of information on node characteristics mentioned so far was the possibility of isolating subgroups within a network and it constitutes the basis for the reduction of larger networks to networks of relations between groups. In practice, the possibilities of combining various data are much wider.

To put it very simply, two questions can be asked in this case. Firstly, to what extent do the relations in a network depend on the attributes of individuals? – this is the question about the phenomenon of social selection. Secondly, to what extent do the attributes of individuals depend on the relations between them? For instance, do the nodes that are related to each other have similar properties? This is a question about social influence. The phenomenon of homophilia, i.e. similarity of nodes related to each other, which occurs relatively often in networks, can be a result of both the selection process and social influence. Such issues were studied inter alia using the example of teenager groups and behaviours such as smoking cigarettes, using drugs, drinking alcohol. In the above example, it would be worth verifying at least how the cooperation between organisations depends on their type – do enterprises, administration bodies and non-governmental organisations cooperate more frequently with the same entities? Or, to what extent are durable groups of cooperating entities composed of various types of organisations?

A slightly different case of using the information on individuals' attributes is represented by the research on social capital, where one of the significant elements is the access to resources by means of social relations. Such analyses are concentrated on the type of resources that can be efficiently reached by an individual by mobilising their own social network.

Research on network effects is possible to a certain extent by using network variables, for instance the ones describing an individual's position as dependent variables in traditional statistical analyses. However, analysis of such dependencies is often very hard. The main problem is that traditional statistical methods require the individuals in the analysed population to be independent from each other. In the case of network analysis, it is possible only in the research on egocentric networks carried out using a sample. Consequently, in the recent years, many models have been developed that allow testing of hypotheses for network data – exponential random graph models ERGM[40] – and methods of analysing the effects of social influence and social selection[41].

[39] A.-L. Barabäsi, op. cit.

[40] T.A.B. Snijders, P.E. Pattison, G.L. Robins, M.S. Handcock, op. cit; G.L. Robins, T.A.B. Snijders, P. Wang, M. Handcock, P. Pattison, *Recent developments in exponential random graph (p*) models for social networks*, Social Networks 2007, No. 29, p. 192–215.

[41] T.A.B. Snijders, op. cit.

SNA in evaluation

As pointed out by Durland and Fredericks, network analysis methods have been introduced relatively recently into evaluation research and they are still not fully used yet[42]. It may be a result of the fact that there are many social network analysis methods and various types of network measures, the use thereof requires skills of using specialist programmes for network analysis[43] and consequently, network research conducted by budding researchers is very difficult[44].

Currently, evaluators in various countries use social network analysis methods. One of the examples is the application of network analysis in health and epidemiological programmes, in particular those aiming at counteracting the spread of diseases such as HIV/AIDS, SARS, but also in the promotion of health[45], counteracting alcoholism, drug addiction, combating smoking cigarettes and obesity. As a basis for constructing specific programmes, network approach is used in these cases... Research shows inter alia that these behaviours are of network nature, obese persons have more acquaintances among obese persons, smokers are inclined to interact with smokers, etc., and thus efficient counteracting and combating addictions should focus on relational mechanisms. Recent research by Christakis and Fowler indicated that giving up smoking is of group nature and hence anti-tobacco campaigns should not be oriented towards individuals, but towards groups of persons that know each other and smoke together[46]. Evaluation research of such programmes should be based on network analysis.

Network analysis methods were also used in the evaluation of school programmes aiming at the improvement of functioning of schools by means of increase in the cooperation between teachers[47]. Generally speaking, the application of network analysis in evaluation can be considered where the data of relational nature is being dealt with.

However, network analysis can be used not only for evaluation of impact results of various programmes, but also for evaluation the programmes themselves and organisation thereof. Especially larger programmes can be conceptualised as networks[48]. Practically every project, and all the larger programs, involves various persons, frequently from diverse organisations. Relations between these persons, their involvement at particular levels or in various programme parts can be described by means of affiliation networks. The knowledge resulting from network analysis can contribute to the facilitation of the analysed programme.

---

[42] M.M. Durland, K.A. Fredericks, *Social Network Analysis in Program Evaluation*, New Directions for Evaluation 2005, No. 107.

[43] The specific nature of network analysis causes that majority of popular computer programmes for statistical analysis does not allow analysing data by means of network methods. However, there are many programmes available in the internet free of charge that allow for the visualisation and analysis of network data. The most popular ones include UCINET and R programme (in particular SNA, Network and Statnet packages) and Pajek, NetDraw for visualisation, as well as SIENA, pnet for more advanced analyses.

[44] R.J. Davies, *SNA tools: their potential and limits*, 2007.

[45] T.W. Valente, *Evaluating Health Promotion Programs*, Oxford University Press 2002.

[46] N.A. Christakis, J.H. Fowler, *The Collective Dynamics of Smoking in a Large Social Network'*, New England Journal of Medicine 2008, No. 358(21), p. 2249–2258.

[47] W.R. Penuel, W. Sussex, Ch. Korbak, Ch. Hoadley, *Investigating the Potential of Using Social Network Analysis in Educational Evaluation*, American Journal of Evaluation 2006, Vol. 27(4), p. 437–451.

[48] R.J. Davies, *Network Perspectives on the Evaluation of Development Interventions*, Paper for the EDAIS Conference November 24-25, *New Directions in Impact Assessment for Development: Methods and Practice 2003*; R.J. Davies, *Scale, Complexity and the Representation of Theories of Change*, Evaluation 2004, Vol. 10(1), p. 101–121, Sage Publications, London; R.J. Davies, *Scale, Complexity and the Representation of Theories of Change Part II*. Evaluation, Vol. 11(2): 133–149, Sage Publications, London 2005.

An example of using network analysis in order to analyse information on a very large programme is the application of these methods by the Polish Agency for Enterprise Development in the analysis of Measure 2.3 of the Sectoral Operational Programme Human Resources Development (SOP HRD). Under the SOP HRD within Measure 2.3 "Development of personnel of modern economy", over 400,000 persons (from over 100,000 companies) participated in approximately 700 training programmes. On the basis of data collected during programme implementation, it is possible to analyse the participation relations of persons and companies in particular training programmes. It allows inter alia to examine the relations between trainings (by means of joint participation) and to draw conclusions concerning the most common training combinations, thematic supplementation of projects and their substitutability.

To sum up, network analysis can be useful at various evaluation levels. These methods allow researchers to analyse not only data regarding the attributes of individuals, but also the connections between them. It is also an approach to social research in which one of the main sources for explanations of actions and situations of individuals are relations and structures of mutual connections in which these individuals are involved. Therefore, network analysis can be useful in evaluation research both in assessing the results of undertaken operations whose significant element is constituted by relations, e.g. the above-mentioned addiction-combating programmes, as well as in analysing the way of organising the measures and projects and importance of this organisation for the efficiency thereof. Thus, it is worth considering, while carrying out evaluation tasks, the usefulness of applying the network approach and collecting the data of relational nature. Despite difficulties related to the multiplicity and complexity of network methods and usage of programmes for analysis of network data, faced by the persons that have just begun to work on such studies, it should be hoped that this approach will be applied more commonly in evaluation research as well.

Dominik Batorski, PhD – is Assistant Professor of Sociology at the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw. His scientific work encompasses primarily the issues of social networks, as well as dynamic of social processes occurring within networks such as diffusion of innovation or social impact. He also works on social network analysis methods and methodology of carrying out psychological and social research via the Internet. He is a member of the Council for Social Monitoring preparing the cycle of studies called Social Diagnosis. In this research, he is responsible for the module devoted to conditions, and consequences of using information-communication technologies. He is a member of several scientific associations; inter alia the International Network for Social Networks Analysis and the Polish Sociological Association.

Sources

Barabási A.-L., *Linked: How Everything is Connected to Everything Else and What it Means for Cusiness, Science and Everyday Life*, A Plume Book 2003.

Burt R.S., *Structural Holes: the Social Structure of Competition*, Harvard University Press 1992.

Burt R.S., *Brokerage and Closure: An introduction to Social Capital*, Oxford University Press 2005.

Casciaro T., *Seeing things clearly: social structure, personality, and accuracy in social network perception*, Social Networks 1998, 20(4).

Christakis N.A., Fowler J.H., *The Collective Dynamics of Smoking in a Large Social Network*, New England Journal of Medicine 2008, No. 358(21).

Cummings J.N., Cross R., *Structural properties of work groups and their consequences for performance*, Social Networks 2003, 25(3).

Davies R.J., *Network Perspectives on the Evaluation of Development Interventions*, Paper for the EDAIS Conference November 24–25, New Directions in Impact Assessment for Development: Methods and Practice 2003.

Davies R.J., *Scale, Complexity and the Representation of Theories of Change*, Evaluation 2004, Vol. 10(1), Sage Publications, London.

Davies R.J., *Scale, Complexity and the Representation of Theories of Change Part II*. Evaluation, Vol. 11(2), Sage Publications, London 2005.

Davies R.J., *SNA Tools: Their Potential and Limits*, 2007.

Durland M.M., Fredericks K.A., *Social Network Analysis in Program Evaluation*, New Directions for Evaluation 2005, No. 107.

Durland M.M., Kimberly A.F., *Social Network Analysis in Program Evaluation*, New Directions for Evaluation 2006, No. 107, Jossey-Bass.

Erickson B.H., Nosanchuk T.A., *Applied network sampling*, Social Networks 1983, No. 5.

Frank O., *Network sampling and model fitting*, in: P. J. Carrington, J. Scott, S. Wasserman (Ed.), *Models and Methods in Social Networks Analysis*, Cambridge University Press 2005.

Freeman L.C., *Centrality in social networks: I. conceptual clarification*, Social Networks 1979, No. 1.

Granovetter M., *The strength of week ties*, American Journal of Sociology 1973, No. 73.

Hanneman R.A., Riddle M., *Introduction to social network methods*, University of California, Riverside 2005. Published electronically on: http://faculty.ucr.edu/~hanneman/.

Heider F., *On balance and attribution*, in: P. W. Holland, S. Leinhardt (Ed.), *Perspectives on Social Network Research*, Academic Press 1979.

Kilduff M., Tsai W., *Social Networks and Organizations*, Sage, London 2003.

Knoke D., Kuklinski J.H., *Network Analysis*, Sage 1982.

Krackhardt D., Kilduff M., *Whether close or far: Social distance effects on perceived balance in friendship networks*, Journal of Personality and Social Psychology 1999, No. 76.

Lin N., Cook K., Burt R. S. (Ed.), *Social Capital: Theory and Research* 2001.

Lin N., Fu Y., Hsung R.-M., *The position generator: measurement techniques for investigations of social capital*, in: N. Lin, K. Cook, R. S. Burt (Ed.), *Social Capital: Theory and Research*, 2001.

Marsden P. V., *Network data and measurement*, Annual Review of Sociology 1990, No. 16.

Monge P. R, Contractor N. S., *Theories of communication networks*, Oxford University Press, New York 2003.

Moreno J.L., *Foundations of sociometry: An introduction*, Sociometry 1941, No. 4(1).

Newman M.E.J., *The structure and function of complex networks*, SIAM Review 2003, No. 45.

de Nooy W., Batagelj V., Mrvar A., *Exploratory Social Network Analysis with Pajek*, Cambridge University Press 2005.

Penuel W.R., Sussex W., Korbak Ch., Hoadley Ch., *Investigating the Potential of Using Social Network Analysis in Educational Evaluation*, American Journal of Evaluation 2006, Vol. 27(4).

Robins G., Pattison P., *Random graph models for temporal processes in social networks*, Journal of Mathematical Sociology 2000, No. 25.

Robins G., Pattison P., Elliott P., *Network models for social influence processes*, Psychometrika 2001, No. 66.

Robins G.L., Snijders T.A.B., Wang P., Handcock M., Pattison P., *Recent developments in exponential random graph (p*) models for social networks*, Social Networks 2007, No. 295.

Scott J., *Social Network Analysis: a handbook*, Sage Publications, London 2000.

Snijders T.A.B., *Models for Longitudinal Network Data*, in: P. Carrington, J. Scott, S. Wasserman (ed.), *Models and Methods in Social Network Analysis*, Cambridge University Press, New York 2005, Chapter 11.

Snijders T.A.B., Pattison P.E., Robins G.L., Handcock M.S., *New specifications for exponential random graph models*, Sociological Methodology 2006.

Valente T.W., *Evaluating Health Promotion Programs*, Oxford University Press 2002.

Wasserman S., Faust K., *Social Networks Analysis: Methods and Applications. Structural Analysis in the Social Science*s, Cambridge University Press 1994.

Watts D.J., *Networks, dynamics, and the small-world phenomenon*, American Journal of Sociology 1999, No. 105(2).

Watts D.J., *Six Degrees: the Science of a Connected Age*, W. W. Norton 2003.

Wellman B., *Structural analysis: from method and metaphor to theory and substance*, in: B. Wellman, S.D. Berkowitz (ed.) *Social Structures: a Network Approach*, Cambridge University Press 1988, Vol. 15 series Contemporary studies in sociology, Chapter 2.

Stanisław Bienias

Application of economic modelling results in the process of Cohesion Policy evaluation

What are the expectations of recipients concerning the results of evaluation when they reach for the final report? They want to see hard facts - and what is more solid than figures showing the effects of the examined intervention? Most of us would be almost entirely content if the figures concerned also the impacts not only also information relating to the level of outputs and outcomes (after all, these are provided by monitoring activities[1]). Since persons, occupied with research for a longer while would obviously like to see the figures relating to real results (so-called net effects) and not only the gross results of evaluated interventions. What is more, sometimes we want to estimate the scale of these effects (see ex ante evaluation) even before our programme/project is launched.

As you can see, it is difficult to please recipients of evaluation, there are tools and research techniques which make it possible to obtain the desired effect. Economic models are one of these tools and the ones which are perfectly suited for that purpose. Economic modelling enables precise quantification of the effects of our intervention which goes far beyond general statements (so very common in evaluation reports) saying that "it seems that the project/programme caused considerable positive effects both in the social and economic area, but due to the complexity of these interventions it is difficult to estimate the scale of this impact." Application of economic models allows for the presentation of net effects, and not only gross effects like it is the case in most traditional research methods. Future effects of our intervention may be estimated, and even the expected durability of these effects may be established – those which will be achieved after programme/project completion. Information on the impact of the intervention on the macro-economic variables, such as GDP, employment rate or labour productivity may be obtained – when the tool is properly constructed, the restrictions in this regard are very small. Moreover, the models make it possible to show the effects triggered by our intervention in the entire economy (national/regional/local) and to discuss them against the background of all changes taking place – we may see the effects of intervention in a dynamic context. Despite being quite complicated and requiring considerable workload in its development and further maintenance, the models have one more unquestionable advantage- they are relatively inexpensive[2] .

Does it imply that we should ground all evaluation exercises exclusively on economic modelling? Of course, not! Economic models have their limitations just like any other tool. The paper shall not touch

[1] In Poland in the programming period 2004-2006, although the monitoring of the financial progress is very efficient (the Prime Minister receives information who, under which measures fails to spent money within 20 days after completing each month), there is very little information available on the material progress. There are too many indicators, they are inconsistent and in majority of cases they make it impossible to aggregate information on higher levels of implementation. The problems are further strengthened by the defective information system SIMIK 2004–2006.

[2] The implementation of the research aimed at assessment of the net effect with the use of surveys on large control samples will in the majority of cases be more expensive that the usage of the economic model.

upon issues concerning the structure thereof; it will only and exclusively deal with the application of the sole results of the modelling.

Given the specific terminology and the need for the recipients to have at least some minimum macroeconomic knowledge, the first difficulty appears as early as the modelling results are presented. For most research recipients the difference between the "percentage point" and "percent" is slight. The differentiation between "the level of Gross Domestic Product" and "the growth rate of the Gross Domestic Product" is even more difficult. Quite often (the media are especially good at it) these terms are used interchangeably in respect to a specific figure! One can imagine the scale of errors arising this way! Another difficulty results from the fact that in the case of evaluation research the figures and tables demonstrate only the impact of the intervention on a specific macro-economic variable(i.e. the difference between two possible development scenarios, taking into account the intervention implementation and a hypothetical situation – what if a given intervention would not have been implemented). It often leads to additional misunderstanding of the research results. Sometimes the data presenting the cumulative effect of the researched intervention (total cumulative effect in several consecutive years) are added up, which in the end leads to a situation, where the result is considerably overestimated (exceeds the real result several times). The way in which one should understand the presented results of macro-economic impact of Cohesion Policy is presented below.

How should the results of economic modelling be interpreted?

Figure 1 presents two development scenarios. The development scenario considering the implementation of Cohesion Policy in Poland – grey line, and a hypothetical development scenario of the situation where Poland would not have been a beneficiary of this Policy. Dotted lines (difference between the two scenarios) illustrate the impact of Cohesion Policy implementation on the given variable. It is the net effect (since we are dealing with a reference to a development scenario in a counterfactual situation). Usually the figures presenting the results of a research have only one line showing the impact of the intervention on a defined economic variable – such presentation aims at ensuring better transparency. Let us take a look at a specific example of how the Cohesion Policy effects are presented.

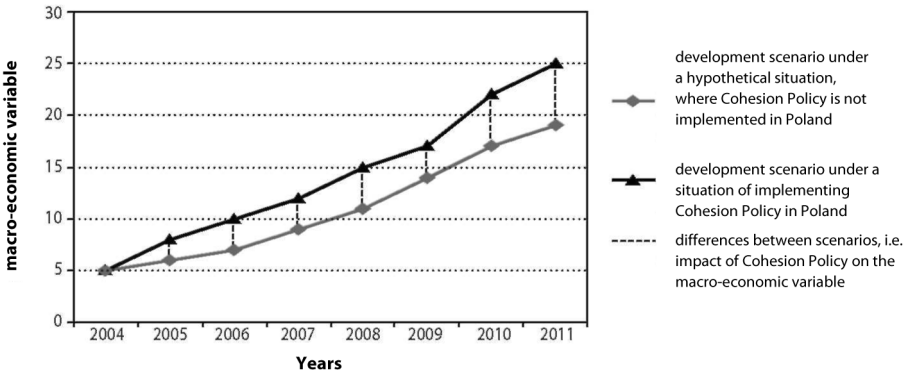


Fig. 1. Impact of Cohesion Policy on macro-economic variables

Source: Developed on the basis of materials obtained from the National Evaluation Unit of the Ministry of Regional Development.

Figure 2 demonstrates the impact of Cohesion Policy on the number of working persons aged 15-64, estimated using two macro-economic models (HERMIN model – grey line and MaMoR2 model – black line). According to the estimations carried out with the use of the HERMIN model in 2007, the number of working persons[3] was higher by 252 thousand persons than it would be in a hypothetical situation if Poland was not a beneficiary of Cohesion Policy and it had not enjoyed the financial support from the EU (note that the distance between the grey line and the horizontal axis on Figure 2 corresponds to the length of the dotted line – illustrating impact – on Figure 1). It should be remembered that the value of 252 thousand persons in 2007 is a cumulative effect. It would be incorrect to interpret that these jobs were created only in 2007. A great majority of persons – from among that 252 thousand – found employment (as a result of Cohesion Policy implementation) already in the previous years[4]! Following this trail, we can see that between 2008 and 2009 the effect of implementing Cohesion Policy will have a negative impact on the labour market[5] – although the cumulative effect will remain positive - 242 thousand persons. The results obtained with the use of the second model – MaMoR2 (black line on the figure), should be interpreted in a similar manner.

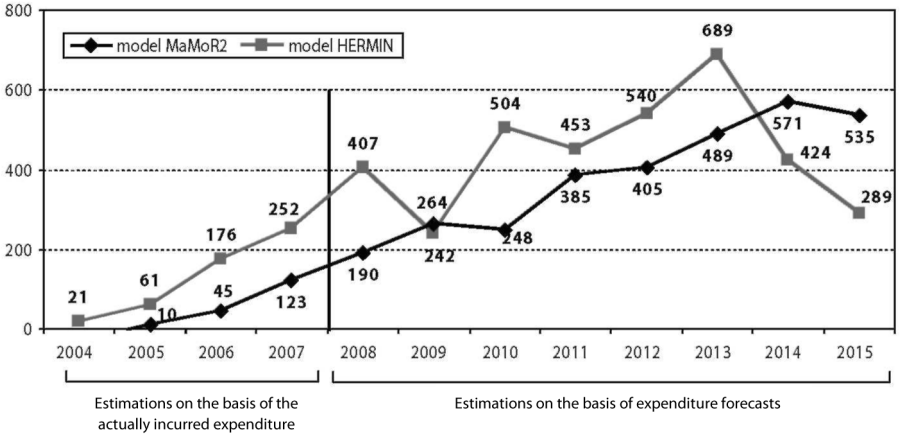


Fig. 2. Cumulative impact of Cohesion Policy on the number of working persons aged 15-64 (thousands of persons)

Source: *Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007* [*Assessment of Poland's progress in convergence with EU Member States and impact of EU funds on the economy in 2004–2007*], Department of Structural Policy Coordination, Ministry of Regional Development, July 2008

[3] Calculated on the basis of the results of the Labour Force Survey (LFS). According to the LFS methodology working persons cover persons, which in the given week:

- performed for at least one hour any work generating pay or income or assisted (without pay) in work on family agricultural farm or in conducting family economic activity outside agriculture,
- had work but did not perform it (e.g. due to sickness, vacation, break in the operation of an establishment, difficult weather conditions) but they were officially employed (average annual value).

[4] A statement that 510 thousand jobs were established in the years 2004–2007 (21 + 61 + 176 + 252) would be an even greater error. Such erroneous information was disseminated in acknowledged newspapers of nationwide range.

[5] A drop from 407 to 242 thousand persons. A major cause of such state of affairs will be the expected and forecasted decrease in the transfers to economy, which is related with completing the implementation of the 2004–2006 perspective, and the yet incomplete "regulation" of the expenditure of resources under the perspective 2007–2013.

It should be noted that each institution and each expert presents the obtained economic modelling results in a slightly different manner, which additionally hinders proper understanding of these results. The amount of information concerning different macro-economic variables presented in the evaluation report should rather be limited, devoting more space to provide a detailed description and interpretation of some key values seen as the most significant from the perspective of research implementation. Moreover, it is extremely important that the report itself be written in a non-professional language, understandable for a wider audience, not only for a narrow group of experts and researchers. Even more so, if the results of our report could prove interesting not only for the general public but the media as well then such an approach could prevent many misunderstandings and the need to write corrections to press articles[6].

Thus, we have identified the first problem related with the use of modelling results – they are relatively incomprehensible to a wider audience. Another problem occurs at the moment of their interpretation - explanation why the results are the way they are and not otherwise. Most persons (virtually everyone not directly engaged in modelling) see the model itself constitutes as a "puzzle box." After putting some values into the model – in the case of Cohesion Policy these cover actual payments (on the basis of historical data) and/or planned payments (forecasted payments in subsequent years) for beneficiaries of aid broken down into expenditure categories[7] – we obtain results which are of interest for us – namely the impact on GDP, on the number of working persons. The mechanism behind this conversion will always remain a mystery for most evaluation recipients. The problem most often occurs at the moment when the effects are different from what had been expected, e.g. there is a sudden change in the trend of impact, high fluctuation of forecasts or other unexpected results.

Controversial results of macro-economic modelling

The analysis conducted by the Office of the Marshal of the Dolnośląskie Voivodeship[8] with the aim to optimise the allocation for the regional programme, constitutes an interesting example. Several scenarios of resources allocation (different share of resources used for interventions, which may be defined as acting for achievement of the Lisbon Strategy objectives) were developed within the framework of the analysis along with their expected impact on the development of regional economy.

To everybody's surprise, it was discovered that the less "pro-Lisbon" allocation scenario ensures better development effects! Does it mean that we should not implement the Lisbon Strategy with the use of resources under Cohesion Policy? Of course not!

We need to remember that the model is only an instrument and the obtained results are burdened with some assumptions and restrictions of the instrument itself. Firstly, the differences between individual

[6] Moreover, when presenting the results on the press conferences it is also necessary to prepare a note for the press, which should be written in a simple and easy to understand language, as well as it should be placed on a relevant Internet website.

[7] The models used in Poland consider three major categories of intervention, i.e. directions of using resources: investments in the production sector, investments in infrastructure and expenditure on human resources development.

[8] J. Bradley, J. Zaleski, P. Tomaszewski, M. Zembaty, *Ocena efektu makroekonomicznego Regionalnego Programu Operacyjnego na lata 2007–2013 na gospodarkę Województwa Dolnośląskiego za pomocą modelu HERMIN [Evaluation of macro-economic effect of the Regional Operational Programme for 2007–2013 on the economy of Dolnośląskie Voivodeship with the use of HERMIN model]*, Wrocław 2006.

scenarios were relatively slight and they were within the limits of the modelling error. Secondly, while under „pro-Lisbon expenditure" the model perceived different expenditure scenarios in a definitely more simplified form than we understand them. The data delivered to the model were grouped under three intervention categories (interventions in the production sector, infrastructure investments and expenditure in human resources development). For example, the category of infrastructure covered both "pro-Lisbon" expenditures, i.e. construction of motorways, as well as expenditure which is very unlikely to generate high development stimuli, e.g. construction of swimming pools. Under the model, one euro allocated to construction of a motorway generates the same effect on GDP increase as one euro spent on construction of social infrastructure. In this case the institution, which implemented the analysis, did not leave the results unexplained and clearly indicated that the obtained estimations do not allow to make binding conclusions within the scope of effectiveness of „pro-Lisbon" resources. Had the authors of the model left the results without a relevant comment, they could have been used for the purpose of making wrong decisions on the political level.

The reasons behind atypical/unexpected results may be very different – starting from the assumptions implemented under the model, through to the economic theories on which modelling is based[9]. If the interpretation of controversial results is left only to the recipients of evaluation, it may be expected that several improper and unjustified explanations will be formulated. The authors of the model are the most appropriate persons to provide relevant interpretation each time the model is used, - they know the model the best and know which mechanisms contributed to the achievement of these and not some other results. However, the persons recommending the preparation of the evaluation should make sure that relevant explanations are included in the report[10].

Since - as pointed out above - for most people a model is a "puzzle box", the renown and authority of the authors of the instrument itself is of enormous significance when implementing research with the use of a given model. This is particularly important, since most models constitute the property of specific institutions and the full documentation presenting all elements of the model is unavailable [11]. In most cases, only general model descriptions are published together with a set of assumptions adopted by its authors. Therefore, the reliability of the results is to some extent "branded" with the authority and experience of the authors of the tool, and much of the information and data is assumed only on the basis of the opinion and information obtained directly from persons implementing the research. It is thus important to make sure that research results are backed with the knowledge and experience of the authors of the model, since the reliability of the modelling – among other things - will depend on it.

---

[9] At this place let us take a look at different patterns of behaviour of different models, e.g. in the case of sudden cessation of financing or a significant limitation of its scope. Under these circumstances some models forecast very rapid – negative "bounce" in the area of real economy (HERMIN), while other respond in a relatively calmer manner (MaMoR2).

[10] Practice shows that sometimes the authors themselves are unable to present a clear and unambiguous interpretation of "doubtful" results and they themselves do not know why the model shows these and not other forecasts.

[11] The documents on the Polish implementation of the HERMIN model are an exception here. As a result of financing the majority of works from public resources both the Ministry of Regional Development as well as the Wrocław Regional Development Agency (co-author of the Polish version of the HERMIN model) carries out a very open policy of providing and disseminating all information on the model construction. Such activities, undoubtedly correct from the perspective of financing the project from public resources, at the same time put the model at risk of quite intensive criticism of scientific and experts' environments. Because of the limited amount of available information it is impossible to thoroughly criticize other models.

Another problem encountered upon the application of economic modelling results, is the limited comparability of data. The difficulty lies in the fact that the results of forecasts made with the use of individual models vary considerably. These differences concern both the scope of impact, the moment when the impact is the highest, as well as response of the economy to limiting the scale of European resources inflow, etc.

Differences in forecast results

Results of analyses commissioned by the Ministry of Regional Development, carried out with the use of two models estimating the impact on the regional level, may be an example here [12]. The MaMoR2 model-based assessment of the entire impact of Cohesion Policy leads to forecasting results, indicating that the Policy will contribute to accelerating convergence between Polish regions. All poorest voivodeships will develop considerably faster than the national average. The HERMIN model-based forecast results produce unqualified reverse results – implementation of Cohesion Policy will contribute to strengthening the differentiation between regions of our country (four out of five of the poorest voivodeships will develop slower than the national average).

The impact of EU funds on the level of GDP per capita in relation to the national average for 5 of the poorest and 5 of the richest Polish voivodeships is presented in Table 1.

The differences in the modelling results upon the application of two different tools seem understandable – however intriguing – from the scientific perspective. The models are prepared by two different research teams, each tool is based on different economic theories and the research is carried out in line with slightly different assumptions. While using economic models, you should always keep in mind that although they are sophisticated and complicated, they are still only a tool and the obtained results need not always correspond to reality. As is in the case of implementing surveys – we can get different answers to an identical question depending on the type of the survey used. Similar situation (but slightly more complex) occurs if economic models are applied. The specific nature of the model and the adopted assumptions has a very strong impact on the obtained results. It is, therefore, necessary to be thoroughly familiarised with the assumptions of the research process, and in order to strengthen the reliability of the received results, it would be worthwhile to consider more than one economic model (triangulation). It provides the possibility to base the decision making process on better verified information.

Enormous differences in results obtained from the same models but for different time frames may be very confusing. In some cases, the modelling results differ by as much as 50% between the researches carried out year after year.

---

[12] T. Kaczor, R. Socha, *Badanie wpływu Narodowego Planu Rozwoju 2004-2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych [Study on the impact of the National Development Plan 2004–2006 and the National Strategic Reference Frameworks 2007–2013 on the selected indicators of the strategic documents]*, Gdańsk Institute for Market Economics, Prevision, April 2008.
Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin [The impact of the investment financed from the EU funds on shaping the major indicators of the strategic documents – NDP and NCS, as well as other macro-economic indicators on the national and regional level with the use of the national model and the regional Hermin models], Team of the Wrocław Regional Development Agency under the guidance of Professor J. Zaleski in cooperation with dr J. Bradley, May 2008

Table 1

Voivodeship	The impact of EU funds on the level of GDP per capita in relation to the national average for 5 of the poorest and 5 of the richest Polish voivodeships		
		MaMoR2	HERMIN
		2015	
Average change for the poorest voivodeships	The poorest voivodeships	+	-
Lubelskie		+ 1.8	-2.57
Podkarpackie		+2.3	+0.35
Podlaskie		+2.8	-2.59
Świętokrzyskie		+2.4	-2.7
Warmińsko-Mazurskie		+3.7	-0.13
Dolnośląskie	The richest voivodeships	+ 1.4	+4.72
Mazowieckie		-6.7	-3.73
Pomorskie		+ 1.7	+2.94
Śląskie		-0.1	+2.46
Wielkopolskie		-1.2	+2.99
Average change for the richest voivodeships		-	+

Source: *Ocena postępów Polski w konwergencji z krajami UE oraz wpływ funduszy unijnych na gospodarkę w latach 2004–2007 [Assessment of Poland's progress in convergence with EU Member States and impact of EU funds on the economy in 2004–2007*], Department of Structural Policy Coordination, Ministry of Regional Development, July 2008

In such circumstances, it is necessary for the authors of the model to present relevant explanations, which provide a background for the obtained results of forecasting. The most common reason for the differences in forecasts differ over time follows from extension of the time series by subsequent periods (model updates). It should be also noted that the "model makers" constantly perfect their skills; they compare the obtained results and consequently introduce some modifications to their tools, which in turn leads to different results[13].

Summary

As compared to other EU Member States, Poland pays relatively much attention to the value of economic model-based macroeconomic impact assessment in the implementation of Cohesion Policy. Models are used in Poland to prepare not only forecasts, which are subsequently considered by the evaluators (as was the case for ex ante evaluation of programme documents for 2007-2013) but also, on

[13] An interesting example of such action was the publication of the 4th Cohesion Report by the European Commission. The forecasting results, *inter alia*, for Poland presented in it consisted in a specific benchmark for results of forecasts prepared by Polish economists.

many occasions models constitute an integral part of the *on-going*[14] evaluation where they are used as one of the analysis methods and not only as a source of information.

Yet, the question still remains to be answered: whether or not - taking into account all the difficulties concerning the application of modelling results - any efforts should be made to further increase the results' role in Cohesion Policy evaluation process? Definitely yes! The benefits of modelling results application are incomparably greater than the problems that could be encountered. It is necessary, however, to try to limit the scale of potential difficulties:

- the modelling results should be presented in an easy to understand manner and the report itself should be written in a language which does not require specialist knowledge;
- the report should devote more attention to interpretation of key results than to "flooding" the recipient with a huge amount of different data, which are little but comprehensible;
- controversial, unexpected results should never be left without a relevant comment provided by the authors of the research who - due to their thorough knowledge of the tool - are the most appropriate persons to correctly interpret the obtained results;
- before starting to interpret the results, it is always necessary to familiarise oneself with the assumptions of the research process;
- the results of the research implemented with the use of economic models should be supported by the authors of the tool, whose knowledge and experience will increase the reliability of obtained estimations and forecasts;
- given the difficulty in comparing the results of the modelling, it is recommended to carry out – if possible – periodical research with the use of the same tool;
- simultaneous application of more than one economic model aims at increasing the accuracy and reliability of the research results [15];
- in order to improve the precision of the obtained results, it is important to base the assumption of economic models, which present results on the macro scale, on evaluation research carried out on micro-economic scale;
- it is worthwhile to use models in the evaluation research as one of the research tools, and not only as a data source – combination of different methods gives the possibility of triangulation and, thereby, increases the reliability of the obtained results.

In the end, it would also be worthwhile to look at the context under which we will have to use results of macro-economic modelling in Poland in the coming years. Poland, from 2007 onwards, became the greatest beneficiary of the EU structural funds. Taking into account the scale of the engaged resources, we became somewhat of a Cohesion Policy laboratory on a European scale[16]. Both the future interests

[14] For example, the study entitled *Wpływ interwencji finansowanych z funduszy strukturalnych UE na zatrudnienie [Impact of interventions financed from EU structural funds on employment]*, implemented by the PAG Uniconsult together with Pentor Research International for the MRD. The evaluation based both on the modelling results as well as on extensive survey studies. Results obtained via different research techniques were subsequently compared with each other, which made it possible to obtain very interesting and, at the same time, reliable, information. Another research for which the model was one of the research tools was the research entitled *Wpływ programów operacyjnych i NPR na realizację celów Strategii Lizbońskiej [Impact of the Operational Programmes and the NDP on the implementation of the Lisbon Strategy objectives]* implemented by the Ecorys for MRD.

[15] In the 4th Cohesion Report the European Commission used three different economic models. The implementation of periodic research of the macro-economic impact with the use of three different economic models is also planned by the Ministry of Regional Development.

[16] Further information on the issue in *Polityka spójności po 2013 r. Pożądane kierunki reformy* [*Cohesion Policy after 2013. Desired directions of reform*], MRD, January 2008

of Poland as well as the future of the entire Cohesion Policy will depend on our ability to use such large funds in a targeted, efficient and effective manner. Evaluation in Poland will be faced with a key task of providing arguments in the debate on the shape and scale of the Cohesion Policy after 2013. The results of the macro-economic modelling, which make it possible to show the overall effects of this Policy, will be one of the most important arguments in the starting debate.

Stanisław Bienias – is a graduate of the Warsaw School of Economics and the post-graduate studies within the field of project evaluation at the University of Warsaw. He has been working in the Ministry of Regional Development for four years, since 2006 on the post of the Head of the Evaluation Division (National Evaluation Unit). He is responsible for the implementation of the evaluation process at the level of the National Development Plan for 2004-2006 and the National Strategic Reference Framework for 2007-2013, as well as for the coordination of the evaluation process at the level of Operational Programmes implemented in the 2004-2006 and 2007-2013 programming periods. Bienias participated in the implementation of over 50 evaluation exercises, he organised a number of conferences and seminars on the establishment of the evaluation culture in our country and participated therein also as lecturer. He represents Poland in many meetings of the "Evaluation Network" under the Directorate General for Regional Policy in the European Commission; he cooperated with the OECD and other national and foreign institutions. He is one of the experts of the currently implemented German-Polish-Romanian Twinning. He conducts lectures on the post-graduate studies within the field of evaluation, monitoring and indicators. Bienias is co-author of some publications concerning evaluation and Structural Funds.

Sources

Bradley J., Zaleski J., Tomaszewski P., Zembaty M., *Ocena efektu makroekonomicznego Regionalnego Programu Operacyjnego na lata 2007–2013 na gospodarkę Województwa Dolnośląskiego za pomocą modelu HERMIN [Evaluation of macro-economic effect of the Regional Operational Programme for 2007––2013 on the economy of Dolnośląskie Voivodeship with the use of HERMIN model]*, Wrocław 2006.

Kaczor T., Socha R., *Badanie wpływu Narodowego Planu Rozwoju 2004–2006 i Narodowych Strategicznych Ram Odniesienia 2007–2013 na wybrane wskaźniki dokumentów strategicznych [Study on the impact of the National Development Plan 2004–2006 and the National Strategic Reference Framework 2007–2013 on the selected indicators of the strategic documents]*, Gdańsk Institute for Market Economics, Prevision, April 2008.

Polityka spójności po 2013 r. Pożądane kierunki reformy [Cohesion Policy after 2013. Desired directions of reform], MRD, January 2008.

Wpływ realizacji inwestycji finansowanych z funduszy unijnych na kształtowanie się głównych wskaźników dokumentów strategicznych – NPR i NSS oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modelu krajowego i modeli regionalnych Hermin [The impact of the investment financed from the EU funds on shaping the major indicators of the strategic documents – NDP and NCS, as well as other macro-economic indicators on the national and regional level with the use of the national model and the regional Hermin models], Team of the Wrocław Regional Development Agency under the guidance of Professor J. Zaleski in cooperation with dr J. Bradley, May 2008.

Jarosław Górniak, Karolina Keler

Indicators in *ex-post* evaluation of public programmes

Introduction

The presence of the following article in this volume makes it unnecessary to discuss the notion and functions of evaluation and differences in approaches to the method of conducting it. Irrespective of the position we hold in an ongoing discussion, one issue remains undisputable: in the course of an evaluation, it should be specified whether (and to what extent) intervention objectives were achieved. Without this, neither the assessment of effectiveness of a project nor identification of cause-effect relations between instruments and effects, nor facilitating activities would have proper foundations. It is a mistake to think that while performing evaluation one may abandon specifying the effects of undertaken actions: it contradicts the essence of evaluation. Introducing projects or programmes in practice aims at obtaining a planned change (in specific cases they may also be directed at restraining a change). The key issue of *ex-post* evaluation is measuring the size of a change caused by an undertaken intervention and, as a consequence, stating whether or not the accepted „programme theory" happened to be adequate[1].

Evaluation cannot be separated from the process of programming activities. Whether an action was programmed well or not may be verified for example through determining whether or not a programme may generally be evaluated or whether it is possible to specify the level to which it fulfilled the planned objectives. Objectives, state of affairs towards which a programme or project heads are expressed as notions. If we want to use these notions in a way not causing misunderstanding, they should be precisely defined. Objectives often have a complex nature: to specify them many terms are used and every one of them covers a particular dimension of a future situation, being the objective of the activity. In order to determine, whether an objective was achieved or to what extent we came closer to its achievement (when the distance to the objective is at least gradable) ought to be able to specify important dimensions of an objective in the categories of measurable "variables", i.e. to express it in the form of "indicators" (which is covered in greater detail later in this article). Therefore, it is not surprising that the European Commission demands from the project entities to set indicators for the objectives planned in programmes or projects. These indicators are to serve monitoring (if on-going observation of indicator values is possible) as well as periodical and final evaluations of planned undertakings. From the perspective of project management methodology it is natural to expect that effects of all actions involving organisational resources should be followed with the use of performance indicators among which those with crucial significance – key performance indicators - are usually distinguished.

---

[1] For more about evaluation based on the theory see J.D. Birckmayer, C.H. Weiss, *Theory-based evaluation in practice. What do we learn?*, Evaluation Review 2000, vol. 24, no 4, p. 407–431. This issue is also covered in the part devoted to objectives of a public intervention.

Indicators cause serious problems for programming entities. These problems result usually from the large complexity of objectives and low precision while formulating them or the absence of clear definitions of terms used in their description. Sometimes, indicators are not determined on the basis of precise explication of objectives and after consideration of the best method of measuring them but are selected from various types of catalogues or indicative methodological guides. As a consequence, they are alienated from the essence of what they should measure. Finally, the cause of project promoters' problems with indicators is sometimes misunderstanding of their essence and function.

Indicators (and absence thereof) cause problems also for evaluators while the programme objectives and indicators are not clearly defined and do not satisfy requirements of appropriate evaluation. In such a situation evaluators have to answer the question of whether to determine in detail and specify indicators on their own in order to be able to evaluate the efficiency and effectiveness of a programme or not. Moreover, an evaluator conducting an *ex-post* evaluation faces the need to specify the programme usefulness: to what extent did the programme solve significant social problems, whether it met beneficiaries' needs (irrespective of whether they were included in the programme or project objectives). Also, this aspect of evaluation requires applying indicators although they would rather be indicators describing a social context of the programme, not directly its objectives.

A basic task of this text is to specify functions of indicators in *ex-post* evaluation of public policies as well as assumptions and principles concerning their application for this purpose. The problems referred to above prove however that describing the role of indicators in *ex-post* evaluation requires referring to the entire logics of programming public policies and in particular to the process of defining objectives of an intervention as well as to the tasks of ex-ante and mid-term evaluation. Only in such broad context, it is possible to present indicators as a tool for *ex-post* evaluation. Determining indicators starts as early as in the phase of formulating objectives and plans of an activity, then - in the course of an *ex-ante* evaluation - they are verified. In case of programmes with longer duration, this verification should be conducted in the course of the first mid-term evaluation. Specifying the adequate set of indicators as a result of this process is crucial for the possibility of implementing the entire and correct *ex-post* evaluation and, as a consequence, for determining effects of the intervention.

The problems of indicators are present in a huge number of publications. They are often of „practical" nature: they provide principles for selecting indicators or their constructing, include sample or recommended lists of indicators, classify them and link with specific elements of the programme logics. As a result, we obtain a rich, although disordered, image in which the basic, universal matters are blurred. In order to properly construct or select indicators, it is worth asking a question about their essence and general features. Such approach will allow seeing common „typologies" or indicator characteristics as well as principles of application formulated in publications in a different light.

Indicator as a tool of measuring intervention effect

In order to capture the change caused by an intervention and, as a consequence, assess the intervention, we have to be able to register this change. Sometimes it can be done directly; sometimes, conclusions should be drawn on the basis of observing various phenomena connected with the one which really interests us. Indicators are measurable variables remaining in connection with relevant features of analysed objects or states of facts due to which they are used to specify whether a particular state of affairs take place or, when we deal with characteristics possible to be expressed quantitatively, in terms of its size. It is important to become aware that the word "indicator" is used to refer to a variable

or a feature of examined objects or states which can take various values (at least two, e.g. existence or non-existence of a particular feature) and not an object or numerical value. A number is treated as a value attributed to an indicator in a particular situation (e.g. before an intervention or after it) but in this approach it is not understood as a sheer „indicator". These issues are often confused an also one forgets that theoretical characteristics, abstracts also referred to as constructs which cannot be directly measured may not occur as indicators. Such constructs are for example terms often appearing in various programmes such as: competitiveness, social cohesion, poverty, innovativeness etc.

An indicator may be connected with the notion of characteristics measured with so-called *indicatum* – according to the definition connecting a more general notion (e.g. poverty) with measurable variables. For example, being poor may be defined as being a member of a family household with income per person (an equivalent adult) lower than 60% of the median of equivalent income in the analysed population. However, we want to underline that it is not the only possible relation[2]. An indicator may also be a variable which remains in statistical (correlational) relation with the indicated variable (*indicatum*). The indicated variable itself is potentially available empirically although from legal or financial reasons it is problematic to measure it.

Except for the abovementioned links, also so-called inference indicators are present in the methodology literature[3]. These indicators are treated as variables with values resulting from impact of latent variable (theoretical) which cannot be directly measured. Examples of applying inference indicators may be variables expressed through so-called scales, such as the scale of security feeling. Moreover, so-called indices, for example the UNDP Human Development Index have a slightly different status. Complex statistical aspects of measurement constructed in this way and detailed discussion concerning differences between scales and indices are beyond the scope of this article since the issue we would like to cover here is only to indicate the richness of potential relations between variables which are a subject of interest and indicators applied to empirically determine their level.

In the context of accepted here understanding of an indicator, let us consider the definition included in one of documents of the European Commission (EC). "An indicator can be defined as the measurement of an objective to be met, a resource mobilised, an effect obtained, a gauge of quality or a context variable"[4]. The word "measurement" means a scale with the use of which we may determine the value of an examined object. Various objects may obtain various values; therefore, here we deal with a variable, not a specific value. Such interpretation allows treating this definition as similar to the definition provided by us. However, it is visible that the indicator definition adopted by the EC is not appropriate since more focus is on what elements important for the programme may be measured with the use of indicators, than on how to understand the sheer term „indicator".

We have already mentioned that there are various types of relations between an indicator and *indicatum* (definitional, empirical and inferential)[5].

---

[2] Described relation may even bring various problems which are discussed in literature devoted to methodology of research and science philosophy under a term "operationalism", this discussion is omitted here.

[3] See for example S. Nowak, *Metodologia badań społecznych, (Methodology of social studies)* Państwowe Wydawnictwo Naukowe, Warsaw 1985.

[4] European Commission, *Indicators For Monitoring and Evaluation: A Practical Guide, The New Programming Period, 2007-2013: Methodological Working Papers, Draft Working Paper [X]*, DG Regional Policy 2006, http://ec.europa.eu/regional_policy/sources/docoffic/working/doc/indic_sf2007-2013_en.pdf.

[5] Definitional – according to the definition connecting the indicator with indicatum. Empirical (factual, correlational) - requiring empirical evidence of connecting two variables. Inferential – requiring a larger number of indicators and statistical algorithms needed for determining a hidden variable.

It is worth emphasising that various research situations require applying different types of relations and selection of a specific type of indicator. On the basis of what was said above, specifying indicators for objectives means selection of measurable variables representing important dimensions of the objective. How should this be understood? In specific cases, in a situation of a definitional link it may happen that the objective is possible to be expressed directly with the use of a measurable variable which exhausts the definition of a feature in relation to which this objective is set. It happens rarely and in fact only when we deal with a single-dimensional objective, encountered rather while defining objectives of the lowest level (the most detailed). In case of multidimensional (complex) objectives, it is better to use more indicators. It may happen that we come across objectives formulated as features which cannot be directly measured; since while specifying objectives of public intervention we often deal with theoretical variables called constructs or, according to another convention, latent variables which cannot be directly measured, for example, competitiveness, technical progress, employability and many others. In such situation, when we limit to indicate one definitional indicator, we narrow the objective to one dimension of a problem. Due to the fact that objectives usually cannot be simply transferred into single measures, applying only one indicator has to be an aware decision resulting from for example indication of the most important, significant from the point of view of accepted values, dimension of a problem. It results in a recommendation that in the course of specifying indicators for complex objectives also inference relations or sets of indicators should be applied.

It is worth noticing that indicators are not a perfect reflection of features of interest for us. Indicators bear a measurement error originating in limited reliability of our tools; however, methodology includes the ways of increasing the measurement reliability. An additional error, resulting from the fact that an indicator is not identical with the indicated variable and covers it only partially (we aim at covering it in the most possible part), refers to all indicators, except for single-used definitional indicators. Moreover, an assumption that it aptly reflects the essence of a measured objective may be a mistake. Therefore, the issue of assessing and improving the quality of indicators has become significant. What is more, connection of an indicator with an indicated variable requires theoretical foundations (definitional or inferential relation) or empirical evidences (empirical relation). Social sciences are a repository of knowledge about these relations.

To conclude, authors of programmes or projects within public policies are responsible for formulating (defining) objectives which would allow specifying the method of measuring intervention effects. In practice it is specification of indicators for characteristics which we want to change through an intervention and determining their planned values sets eventually the objectives to achieve. Let us consider this issue using an example. Let us assume that the goal of our intervention is to reduce the level of poverty in a particular society. In order to verify whether as a result of the programme implementation the poverty would decrease, we have to measure it which means to apply an indicator (or indicators) properly selected for the specific objective. However, before it happens, we have to, in the course of formulating objectives, precisely specify what we have on mind when we speak about the poverty we want to reduce. (1) It should be specified what kind of poverty we refer to: subjectively or objectively, absolutely or relatively and in what situation we would classify a person as poor. (2) It should be determined how to measure poverty according to the definition accepted by us and how to choose an indicator or indicators which would aptly reflect its essence. (3) Additional, supplementary indicators may be specified in order to complement the image obtained with the use of the main indicators.

The last remark concerns particularly the complex (multidimensional) objectives, within which specification of key indicators provides a special importance for particular dimensions of an objective.

When we take care about the precision of objectives, we would be sure that randomly determined indicators and evaluation based on them would not be inconsistent with our intentions.

Objectives of public interventions

"develop indicators within the discussion on the action in general, not to «add» indicators when the discussion has been finished"[6]

Notion of purpose – hierarchy of objectives

In the course of programming public policies and projects, it is particularly important to properly formulate the activity objectives. At the same time it is one of the most difficult tasks of the process of programming public policies. The objectives are often confused with means of an action and actions themselves. Moreover, they are often formulated in unclear and imprecise way. The objectives should be defined in the category of a state of affairs desired to be achieved. In the methodology of programming, while referring to objectives, the English language distinguishes the notions: *goal* and *objective*. In Polish, there are no good equivalents of these terms. Both of them are translated into Polish as *cel* which causes that in programme documents the authors often limit to use only ***goals***, **which are purposes formulated in the category of a general, qualitative vision**. Obviously, this vision is a necessary and significant starting point for an activity, is a basis for formulating programmes and motivates qualitative consideration of their course and results. However, these are **specific, calculable, reasonably formulated and achievable in particular circumstances as well as possessing clear time perspective usually called *objectives*** which should rather be the essence of planning an intervention and foundation for evaluation.

It is appropriate to treat them in relation to values of the same variables in the starting point which means in the categories of changes (from the initial state to the moment of completion of the activity). Correctly orchestrated intervention objectives should be easy to be expressed as indicators or variables whose values (and more precisely change of value) will allow us to specify whether the objective was achieved or to what extent we are closer to achieving it. Targets of indicators (which unfortunately in Polish are translated also as *cele*) are not the same as objectives although they make them so precise that while planning we immediately determine what target of an indicator would be satisfactory for us. While planning an intervention it is not only important to make an effort of formulating an image of what is desired to be achieved but also to sketch a specific, calculable image of a change which is planned to be performed, expressed as targets[7] . Objectives formulated in this way are subject to control of implementation conducted in the *ex-post* evaluation.

An objective is – let us emphasise it – a certain expected state of affairs towards which the undertaken actions should lead, not the sheer pursuit, intention or attempt. Obviously, project promoters will act in favour of some actions, however all persons involved in the programme should take into account the objectives of these actions and these are the objectives which should be treated as the main point of reference - the desired future state of reality. The reality is described through its dimensions/characteristics. An objective may be specified in relation to one or many dimensions: they may be simple or complex.

[6] European Commission, *Indicators for Monitoring and Evaluation...*, op. cit., p. 11.

[7] A.B. Atkinson, E. Marlier, B. Nolan, *Indicators and targets for social inclusion in the UE*, Journal of Common Market Studies 2004, vol. 42, no 1, p. 47–75.

A complex objective may actually be expressed as a combination of objectives set for its particular dimensions. What is more, achieving some objectives may require implementation of other activities, oriented towards instrumental goals in relation to a superior goal which leads to so-called goal tree.

In the case of programmes and projects financed from the structural funds, we often observe problems with proper determining indicators required by the methodology of the European Commission: output, outcome and impact. The difficulties are connected not only with providing features of indicators appropriate from the methodological point of view, but also with distinguishing outputs, outcomes and impact at the level of defining objectives of a programme or project. Errors and doubts in specifying indicators expressed as these three types of effects usually reveal the chaos in the process of planning and absence of transparency in terms of results which should be obtained by the planned public intervention. The basis for determining good indicators is mainly clear specification of the following:

- what should be provided by an intervention, what should the resources in the course of planned action be used for – which is defining outcomes,
- what changes in the situation or features of intervention beneficiaries are planned to be obtained directly as a result of undertaken actions – defining outputs,
- how such a change of situation or features of intervention direct beneficiaries will influence their situation in the longer perspective – specifying the planned local impact in the terminology approved by the European Commission which should rather be described as a permanent outcome, in comparison with a direct outcome;
- how a particular intervention would influence solving a specific social problem it is aimed at - whether a definition of global impact concerning issues such as for example solving the unemployment problem, limitation of spatial mobility of employees, deficiencies in investments in innovative production, small number of patented inventions, too long period of waiting for a court judgement or its execution, social exclusion of particular groups etc.

Public interventions are launched in order to solve social problems in a possibly permanent way, since it is crucial to specify their goals in the category of expected global impact. Public programmes should serve general purposes, treated in a form of expected changes in the field identified as a social problem, i.e. - as impact. Operational programmes are usually conglomerates of thematically similar programmes called priorities, out of which everyone is targeted at a specific social problem or problems. With this strategic nature of priorities even they should have specified goals in the category of global impact. Planning an action, which is a method to achieve goals, should most of all remain directed at providing particular products and causing direct changes with respect to beneficiaries, therefore the goals should be treated in these specific categories.

Various actions undertaken within a priority repeatedly lead to similar outcomes which can be, in relation to this, expressed in an aggregated way as detailed direct objectives. A target of such detailed objectives would be provided as a level of summed up outcomes at the level of priorities. Applying indicators of outcome would facilitate monitoring of implementing the entire priority, especially in a situation when we expect a gradual appearance of outcomes in the course of implementation.

In case of some programmes also outputs of particular measures (or their sum at the level of a programme or priority) may have a special significance for achieving programme objectives (e.g. kilometres of constructed roads of various categories) - it is worth to present the objectives also as output indicators. At the same time it is important not to lose contact with the objective hierarchy: intervention impact is crucial, implementation of indirect objectives in the form of outcomes should lead to this impact and an instrument for obtaining these outcomes should be providing a particular amount

of outputs of undertaken actions. In case of monitoring and current management of programmes, we take into account mainly output and outcome indicators (observing also the involvement of resources), while in the *ex-post* evaluation one should always attempt to examine the local impact of a programme, namely long-term outcomes of an intervention and, if possible, also the potential global impact.

This logic is applied also in case of projects, which are simpler than programmes within which they are planned. Projects should finally serve to solve social problems; however, units managing implementation of a programme should monitor it by selecting projects in such a way as to ensure that their outputs and direct outcomes give possibilities for the highest possible level of achieving planned programme impacts. Projects as operational units of a public intervention should be consistent with the key objectives specified in the programme and expressed as impact (which should ensure specification of criteria concerning selection of projects) but within the scope of its objectives they should concentrate on providing outputs and direct outcomes at the level of beneficiaries. Therefore, there is usually no need for the project documentation to determine impact indicators of a project, although outcome and output indicators should be specified.

Ex-post evaluation tasks in relation to formulating objectives

In evaluation not only the level of achieving targets is measured but also we should, if possible, determined the level a particular achievement is actually a result of an intervention impact and not an effect of spontaneous processes (not controlled by a decision maker). Such task is often very difficult to perform – if it is possible at all. Therefore, the minimum task of evaluation may be formulated as a survey whether a change is going in a desired direction. Evaluation has also other purposes, including performing assessment of implementation process in order to improve future actions; however discussing all tasks of *ex-post* evaluation is beyond the scope of this article. The current considerations show that in case of the *ex-post* evaluation the determination of a proper set of indicators which would operationalise the intervention objectives and enable measuring an obtained changes is a key issue.

For the purpose of *ex-post* evaluation, it is important to verify net effects of a programme or project, in other words what added value which would not appear without the intervention was brought by it. For example, whether conducted training programmes resulted in a permanent change of a situation on the labour market of beneficiaries which would not take place without a particular undertaking (or a surplus over what would be an improvement stemming from the independent change of the market situation)? Whether making a subsidy available for specific entrepreneurs operating on a local market did not cause elimination of competing entities and, as a result, a drop, not increase, of employment or level of economic activity within this market? In the *ex-post* evaluation the following question is asked: to what an extent it is appropriate to assume that a particular action through its outputs and outcomes allow solving a specific social problems? In the situation when an answer to such questions is very difficult or impossible, a minimum requirement for the *ex-post* evaluation should be determining a permanent net outcome (change on the side of beneficiaries/target situation minus the estimated value of the change which would take place in case of no intervention) named also as "local impact". Answering such questions allows to verify an accepted clearly or *implicite* theory of a programme – set of convictions about cause relation between undertaken interventions and objectives they serve. Therefore, within evaluation the question of programme purposes treated as impact should be asked. Obviously, an evaluator would also ask about mechanisms of implementation and thus he or she would be interested also in mechanisms which made it more difficult to obtain outputs and outcomes and which supported it. All these actions

are carried out in order to verify in the future the theory of programme implementation which is a set of assumptions concerning the most effective and efficient method applied to achieve planned effects.

Ensuring the transparency concerning objectives of programmes and projects is a starting point for specification of a method of measuring them which is determination of indicators. However, in programme documents we often deal with a problem of a proper treatment of objectives in the category of impact, in comparison with an outcome or output. What is more, the objectives are sometimes confused with activity, efforts or attempts which would lead to them (while specifying objectives there sometimes are expressions such as: "creating conditions", "pursuing efforts", "undertaking actions" etc.). All of these makes it very difficult to conduct *ex-post* evaluation and in the situation of absence of adequate indicators – their apt selection for evaluation purposes (such indicator selection after the project completion is only a substitution in relation to a proper settlement of indicators in the phase of designing).

Goals and objectives

The purposes in programmes should be formulated at various levels of universality:

- as purposes-challenges: generally single or sparse and often in a multi-faceted way expressed as quality images, general purposes *(goals)*,
- as purposes-tasks: treated in a specific way, measurable and particular time detailed specification of intervention purposes *(objectives)*.

Detailed specification of objectives and their measurability is ensured by specification of their indicators which are measurable variables representing key dimensions of an objective and accepting planned values of these indicators, targets, in determined time.

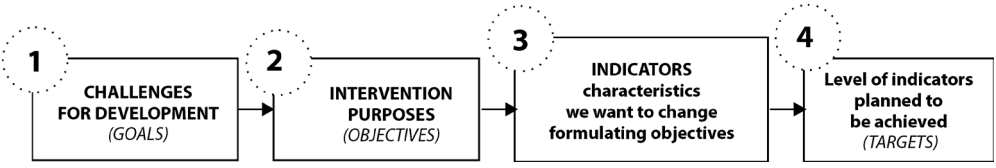


Fig. 1. Challenges, objectives of interventions and planned level of objectives implementation in relation to indicators

Source: Authors' own work.

Within this logics there is a danger of identifying the necessity of specifying indicators for intervention objectives (expressed through specified variables) with providing the planned level of achieving targets (value of a particular variable). Defining an indicator (variable) is a condition for the possibility of providing its target value; providing this value is however not necessary for correctness of formulating a net of objectives but rather desired from the point of view of the possibility of "settling" implementation of a programme or particular project.

It is required that objectives, purposes-tasks, meet the criteria described as SMART, which means:

- *specific* (what we want to achieve),
- *measurable* (one should be able to measure what they want to achieve),
- *achievable*,
- *realistic* – possible to achieve with existing resources,
- *time-bounded* – specified in time

Even at the first sight it is visible that in order for an objective to be SMART, it should be described with the use of measurable variables allowing determining the value of a particular state of affairs which are indicators. On the other side, there are no good indicators if the objectives are not specified in a way not only substantively proper but also formally incorrect.

Social impact in relation to "plan performance": division of intervention objectives

The second basic logics (let us call it "vertical") accepted while formulating programmes of public interventions is linking general purposes of an intervention at the level or general public with detailed purposes of projects at the lowest level relating to beneficiaries. Within the "vertical" logics (1) general purposes are formulated as impact and refer to the level of the general public. These are the proper points of interest of the social policy. In practice, programmes implemented within the social policy cover only a part of society, groups of so-called beneficiaries. (2) Detailed objectives formulated as outcomes and (3) operational objectives expressed as outputs relate exactly to programme beneficiaries.

Following the route of implementation logics: entities implementing an intervention undertake some (4) actions (which (5) are connected with costs). Performing these actions should eventually have a nature of (3) an output. An output may be for example conducting a particular training which would end with obtaining by beneficiaries a specific affirmation or certificate. A sheer output is an important element of designing activities and operational purposes (outputs) are only a tool for achieving detailed purposes (outcomes) including evoking a desired change on beneficiaries' side. (2) An outcome is defined as change concerning beneficiaries caused by an intervention. What constitutes this change, an outcome, should stem from an appropriately determined objective. If an objective is planned to lead to employment of beneficiaries conditioned by completion of training, the outcome indicator would be the number (proportion) of beneficiaries which managed to achieve this. An outcome may also be just obtaining a certificate, e.g. driving license if it would be considered a target change in relation to beneficiaries (an output in this case will be providing an appropriate course and driving lessons which may even be complemented with a criterion of passed internal exam) because the certificate means then an externally confirmed change of qualifications which, bringing significant abilities, itself changes a beneficiary's situation. Whether we limit to this understanding of an outcome depends on the method of formulating an objective. An outcome always means expressed at the level of an objective, expected change of a beneficiary's situation caused by an intervention.

Outcomes at the level of final beneficiaries should contribute to impact of a programme on some characteristics of the whole society. Let us assume that a project in which professional qualifications were improved was implemented within the Operational Programme Human Capital. The strategic objectives, which are a point of reference of the programme, were for example provisions of the Lisbon Strategy – increase of the employment level of persons in the working age. Then, we would like to know whether an output in the form of training contributed to implementation of (1) general objective (expressed as impact) through the change in beneficiaries' situation treated as (2) outcomes. In our simple example, an outcome may be the fact that a particular person was able to obtain the certificate and that the new ability contributed to finding by this person employment.

Distinguishing (2) outcomes from (1) impact allows (using the determined indicators) to trace the influence of changes referring to (2) beneficiaries on the change (1) of a general social situation. It may happen that our beneficiaries will be trained (output), as a result of an exam will obtain certificates and

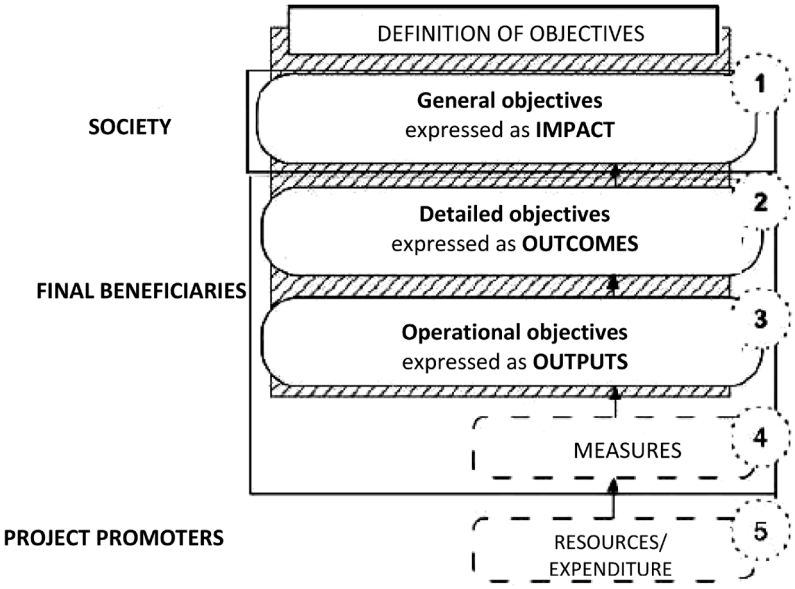


Fig. 2. Objectives: general, detailed and operational expressed as impact, outcomes and outputs

Source: Authors' own work.

as a consequence - employment (outcome) but their employment would cause that other persons will be made redundant or even that more persons would be made redundant than employed. In such a situation the general objectives (impact) were NOT achieved and in spite of successes at lowered levels (of output, outcome), the programme may not be considered successful from the point of view of general objectives. As a consequence of infringing the competitive market conditions, one group would obtain privileges (qualifications without incurring full costs and employment) at the expense of another group with no social effect. This type of action may sometimes be explained with some social consideration although it is not easy. What is more, using financial means for one objective always makes it impossible to use them for another. Obviously, the presented example is a very large simplification; however, it allows following the logics of the objectives hierarchy which is becoming a basis of *ex-post* evaluation. It is worth adding, that the first challenge in the present phase still remains the proper measuring of outcomes and determining the level at which they are an intervention effect.

Problem: should for all levels of intervention objectives specific targets be specified?

Let us refer to a principle question mentioned earlier: on which levels of an intervention programme is it required to determine objectives through values of indicators to be achieved (*targets*) and who should be responsible for this task? Determining the objective, we ask ourselves what is the current state of a particular problem (diagnosed situation) and what we want to achieve (future situation or objective). While formulating objectives for ourselves, we want to change the situation. Evaluation task is to suggest which actions worked, which objectives (outcomes and impact) were implemented, which objectives (outputs) contributed to achieving outcomes and impact. However, in order to do this, the objectives

should be quantified: we need the ability to measure them; we have to identify the variable (feature/property/characteristic which may have at least two different values for various objects) which informs us about the phase we are at in a particular time. A social change we want to perform through achieving an objective may be expressed only on directional basis; however, its scope should be specified in detail while specifying targets[8].

Figure 3 presents two approaches to determining objectives through targets. A target may be defined with relation to the fundamental values obtained on the basis of (1) extrapolation of existing trends or (2) accepting a benchmark value[9]. In the first case we use the concept of dynamic approach presented in Figure 3. The target is usually assumed as increase (in a normative sense) of the value in comparison with the basic forecast (whether it should be a larger or smaller value depends on a normative interpretation of an indicator). However, in case of specifying the target, various approaches are applied starting with those based on standard criteria and ending with standards taken from sample reference groups or determined statistical criteria (e.g. average for a broader social situation). Of course, the issue of determining closely quantified targets as a criterion of achieving a programme objective is an arguable question, in particular when it concerns impact indicators and even outcome indicators, in connection with more general matter of predictability of social phenomena. It is certain that outputs can be used for this and thus controls of implementation deal with output indicators particularly scrupulously. Outcomes are an effect of linking many circumstances and therefore their target value will always be only a kind of approximation of what would actually be achieved. Evaluation should take into account this fact and try to answer the question about causes of deviations in measurement and not only notice the fact that discrepancies (irrespective of their direction) occur. Various balance relations have a large significance for the level of targets of impact indicators, especially those macro-economic. It has to be taken into account as well. One also should not expect the possibility of correct predicting the future after an intervention.

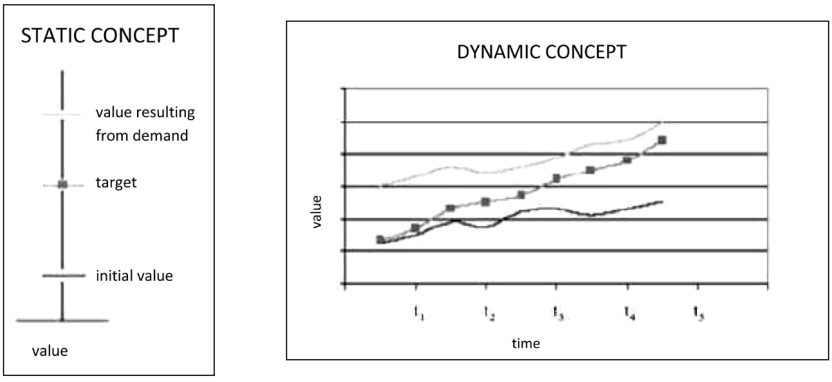


Fig. 3. Static and dynamic concept of formulating objectives through indicators

Source: Authors' own work, slight modification on the basis of: European Commission, Indicators For Monitoring And Evaluation..., op. cit., p. 15.

[8] For example, we want to reduce the level of poverty of children. This objective has not been quantified yet since we have to choose a poverty indicator reflecting the best our understanding of poverty. It may be for example so-called absolute poverty. Then achieving the objective would mean approaching the value of the absolute poverty indicator among children (which initially was higher) to the value of the whole society (benchmarking) or to a specified value (e.g. from 20% to 0%) when for example we assume that poverty of children should not appear at all.

[9] European Commission, I*ndicators For Monitoring And Evaluation*..., op. cit., p. 13.

Indicators in measuring programme utility – role of social diagnosis

A programme (project) utility is an important criterion in *ex-post* evaluation. It is the level to which a programme meets important needs of target groups or contributes to solving problems, irrespective of whether they were included in programme objectives. In order to be able to specify programme utility, it is necessary to relate to a broader social context. Such context should be provided as soon as in the social diagnosis which is an integral part of the programme. In the social diagnosis we apply social indicators for describing a situation and identify problems. The diagnosis should involve clear assessments of described phenomena: we expect that it will specify the points where there is too much, too little and in proper amount of things. The diagnosis should also include potential consequences of taking and not taking particular actions. A description of reality in the diagnostic part is performed also in the form of indicators since only with the use of them the analysed phenomena may be characterised in quantitative categories. These indicators do not directly refer to intervention objectives therefore we describe them as contextual indicators. A suitable selection of contextual indicators for description of social reality provides a possibility of noticing significant social problems which may be suggested by the programme as well as limitations which may be encountered during implementation.

Identification of problems serves – besides assessment in relation to benchmarking value theoretically or politically – comparison of values of contextual variables: (1) in a particular state (region) or in various states (regions) throughout several years which is a time series analysis, (2) in a particular state (region) to an average/mean for various states (regions) (3) and in a particular state (region) in division into separated segments (e.g. according to gender or age) which is used for example to identify risk groups.

In case of *ex-post* evaluation, determining values of contextual indicators after an intervention and comparing them with initial values gives a possibility to assess a programme utility. Also in the course of programme implementation, it is important to monitor some aspects of the social situation, which may influence the implementation and effects of the programme, such as: migrations, demographic transfers, etc.

While performing a diagnosis, focus on contextual indicators should not take our attention away from indicators directly connected with programme objectives, especially from impact indicators. The diagnosis has to be created in the course of an iterative process: on the basis of a discussion about the main social issues and initial vision of strategic goals of a particular policy we start designing the frames of the diagnosis (including indicators), perform the diagnosis, verify the definition of the situation and concept of goals, get back to the diagnosis in order to complement proper issues etc. Finally, public interventions are a kind of visions concerning importance of specific social problems and concepts of their effective solving. Their creative character causes that, on the one hand, their shape does never result autonomously from the diagnosis, on the other hand - due to uncertainty - require, at least in democracy conditions, to take responsibility in relation to public trust. The social diagnosis should support both responsibilities to society as well as public trust so that they could face challenges of uncertainty.

Therefore, besides analysis with the use of (1) "contextual indicators", after formulating intervention objectives and specification indicators of these objectives (2) a diagnosis should be conducted once again, this time with the use of the variables which would be applied in the proper evaluation. It helps verify indicators accepted for objectives, understand better the indicators and objective structure (expressed now in a measurable form with the use of indicators).

In the *ex-post* evaluation we return to contextual indicators and impact indicators used in the diagnosis phase.

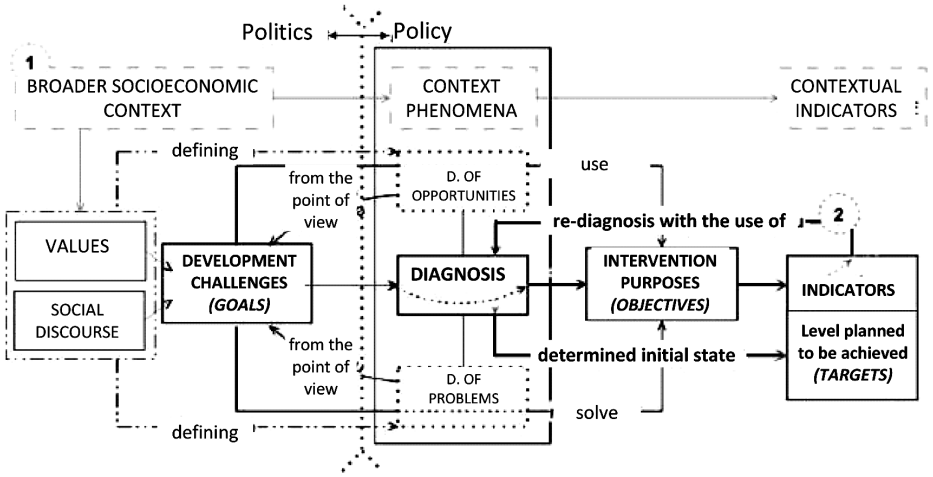


Fig. 4. Relation between the process of developing public policies and indicators

Source: Authors' own work.

Assessment of accepted indicators

In the course of selecting indicators and also after including them in a suggested set covering all key aspect of a programme, they have to be assessed. It is usually conducted in the phase of *ex-ante* evaluation and verified in the course of the first mid-term evaluation. While assessing, we apply a number of criteria which simultaneously function as a guide in the phase of establishing indicators. Due to the fact that indicators should be accepted in the phase of constructing a programme, an *ex-post* evaluator should basically have available a set of indicators, he or she would use – of course, this situation would take place in case of a well-designed programme. If an evaluator has a task of evaluating a programme which was not equipped in properly constructed set of indicators, he or she has to; in accordance with the unit commissioning evaluation – independently set indicators and in this way "reconstruct" the operational sense of a programme.

The key issue decisive in the value of an indicator is obviously the accuracy or ability of the indicator to provide competent knowledge about exactly this aspect of the matter for which it was established. However, effective capturing the essence does not cover the entire criterion of indicator adequacy. Even in case of determining intervention objectives, we mentioned the normative reference of indicators applied to illustrate changes which have to be monitored while conducting policies.

Explicite and *implicite* objectives of an intervention express assumptions concerning the value of a particular set and hierarchy of objectives. Indicators should reflect this normative vision due to the fact that they are the only a tool of measurement. Nevertheless, intervention objectives gain their final significance through specifying indicators allowing precisely defining these objectives and even quantifying them, hence the problem of taking values into account refers not only to formulating objectives but also to indicators.

The normative interpretation of indicators means whether it is "good" or "bad" and whether – in the time period – we aim at the "desired" direction. The social indicators have to be clear (can be understood in an intuitive way) and acceptable (by the majority of society) standard connotations[10]. They are not suspended in a vacuum, they usually refer to an entire group of assumptions and values expressed through ideologies or doctrines; moreover these groups may be of eclectic nature. It is advisable that axiological and normative assumptions are expressed directly while determining indicators since it facilitates correct choice of indicators, limits manipulation and increases objectivity and transparency of assessments.

These remarks refer especially to the impact indicators connected with the most general level of objectives; they condition the objectives (indicators) of a lower level. Indicators included in the whole set of social indicators should be equally distributed between all significant aspects important for the society (concerning conditions of development or social problems) and the set developed in this way should be understandable for members of the society the problem refers to[11].

Another important issue is also the number of indicators in the set: the more there are, the more reliably they reflect the examined phenomenon, however, on the other hand, the less there are, the clearer their results are. "But we cannot expect them [indicators] to be a complete representation of the state of society. [...] The nature [...] of that indication will depend on the choices made with regard to definitions and with regard to data. Different indicators highlight different features of social problems and suggest different priorities for policy intervention"[12]. We cannot neglect how ideologies, researcher's values, values of the general public or political system influence the survey outcomes; these values should be provided in the presentation of indicators, although they should not influence determination of the actual situation.

Assessments of individual indicators and of the composition of their full set are performed usually on the basis of a broader set of criteria. Indicators incorrectly assessed on the basis of such criteria should be verified or replaced with others. The assessment criteria are actually applied in the course of the whole work on indicators; therefore, the assessment is rather a continuous undertaking, not a single activity. In principle, it is advisable to consider criteria even in the course of establishing indicators. Coming to the characteristics of example sets of criteria or indicators quality, we underline once again that the priority here is always the accuracy criterion: theoretical and normative adequacy in relation to established objectives. The accuracy criterion is present in sets of criteria for indicators assessment, however sometimes in slightly concealed way, as for example: „effective indicators require a clear conceptual basis"[13].

The criteria of indicators quality are divided into those concerning a single indicator and those relating to the whole indicators system. An example of a set of assessment criteria, taken from the work of the Council of Europe devoted to the social cohesion is presented in Table 1 [14]. In the below scheme, four aspects of assessing quality of indicators of policies conducted in the analysed public domain are presented:

---

[10] See T. Atkinson, B. Cantillon, E. Marlier, B. Nolan, *Social Indicators. The EU Social Inclusion*, Oxford University Press, New York 2002, p. 19, 21.

[11] See ibidem, p. 24–25.

[12] Ibidem, p. 19–20.

[13] C.W. Cobb, C. Rixford, *Lessons Learned From The History Of Social Indicators, Redefining Progress*, San Francisco 1998, http://www.rprogress.org/newpubs/1998/SocIndHist.pdf, p. 16–17.

[14] Council of Europe, *Concerted Development of Social Cohesion Indicators: Methodological Guide*, Council of Europe Publishing, Strasbourg 2005, http://www.coe.int/t/dg3/socialpolicies/socialcohesiondev/source/GUIDE_en.pdf, p. 87.

- whether they provide what they are expected to (are they adequate for the planned objectives),
- what is their specific (detailed) contribution,
- whether they are easy to be used (from the point of view of a user),
- whether they are expensive/difficult to be obtained (costs of obtaining).

Table 1. Review of quality criteria

	Suited to the objective pursued	**Intrinsic contribution**	**User's point of view**	**Cost of obtaining data**
Individual questions	Relevant to knowledge needs		Transparent	Not applicable
Portfolio of questions	Complete and balanced	Coherent		
Indicators	Representative (of an indicator or a portfolio of indicators)	Unambiguous	Clear and accepted interpretation	Not too expensive
Data	Reliable	Significant	Not too expensive; comparable	Regularly available

Source: Authors' own work on the basis of: Council of Europe, *Concerted Development*..., op. cit.

The aspects mentioned above were contrasted with four components of the system of assessing the Cohesion Policy:

- particular questions which express the needs concerning knowledge,
- set of questions as an entity, covering with analysis of all significant aspects of an examined field depends on its composition,
- indicators which should provide answers to questions,
- data which are the foundation of calculating indicators values.

This scheme rightly pays attention to the fact that indicators derive from asked questions. This approach is particularly adequate in the conditions of constructing the social diagnosis. It will also be useful in the situation (still very often), when an *ex-post* evaluator does not possess a proper set of indicators included in the programme or for various reasons he or she has to go beyond this set at the request of the entity commissioning evaluation. In such situation, the evaluator's work would remain to a large degree a diagnostician's work. In the situation when a programme is equipped in a sensible set of indicators, criteria concerning questions may be applied for verifying whether while determining the indicators it was taken into consideration that they should reflect all important aspects of specified objectives and allow to perform assessment with relation to all approved criteria of evaluation. The criteria included in individual table cells are heuristic and may have a dychotomic scale of yes/no, informing about achieving the particular level of aspiration or they may be developed in an estimated scale of grades.

The example of criteria provided above is one among many. It often happens, by analogy with the above discussed *SMART* criteria, that in order to assess programme objectives the *SMART* criteria are applied for indicators. What is more, the acronym *SMART* is developed in various ways hence it is difficult

to treat it as a commonly approved methodology. For example, letter A may mean *attainable* (as below) or agreed (which indicates the necessity of recognising by the involved actors the same way to measure effects). What is more, the sheer term "measurable" is more suitable for assessment of whether we deal with an indicator at all, than for assessment of its quality: measurability is a feature which defines an indicator - there are not indicators which are not measurable. However, taking into consideration the fact that in various programmes, strategic documents or projects, there are still "pseudo-indicators" which may not entirely function as indicators, this criterion appears required. The whole set of criteria is heuristic and the acronym makes it easier to remember what should be taken into account while determining and also assessing indicators. *SMART* criteria are quite universal and may be applied for indicators from various levels of programming logics.

Below, there is an example of how the acronym SMART may be developed. According to it, a good indicator should be[15]:

1) *"Specific:*
   - is it clear exactly what is being measured? Has the appropriate level of disaggregation been specified?
   - does the indicator capture the essence of the desired result?
   - does it capture differences across areas and categories of people?
   - is the indicator specific enough to measure progress towards the result? For example, using the indicator "increase by 20 per cent in number of criminal complaints filed" may reflect a more effective justice system OR an increase in crime.
2) *Measurable:*
   - are changes objectively verifiable?
   - will the indicator show desirable change?
   - is it a reliable and clear measure of results?
   - is it sensitive to changes in policies and programmes?
   - do stakeholders agree on exactly what to measure?
3) *Attainable:*
   - what changes are anticipated as a result of the assistance?
   - are the result(s) realistic? For this, a credible link between outputs, contributions of partnerships and outcome is indispensable.
4) *Relevant:*
   - does the indicator capture the essence of the desired result?
   - is it relevant to the intended outputs and outcome? To judge the relevance of indicators, the CO [Country Offices] may have to identify the target groups and their needs, expectations and criteria.
   - is the indicator plausibly associated with the sphere of activity?
5) *Trackable:*
   - are data actually available at reasonable cost and effort?
   - are data sources known? CO should establish realistic principles, mechanisms and responsibilities for data collection.
   - does an indicator monitoring plan exist?".

[15] R. Sandhu-Rojon, *Selecting Indicators for Impact Evaluation*, UNDP, Paris 2003: http://www.npi-connect.net/c/document_library/get_file?p_l_id=18214&folderId=23505&name=DLFE-2004.pdf , p. 11–12.

In spite of the heuristic utility of the *SMART* criteria, we opt for slightly more precisely constructed scheme of assessing indicators, which more precisely and systematically refers to methodological and pragmatic aspects of their application. Such catalogue of criteria for assessing indicators was included in the report prepared by the team under headed by Professor Atkinson for the needs of the EU Social Protection Committee[16]. It comprises the set of features which good indicators and indicators sets should have for a particular domain of public policies. The authors emphasise that at the level of the strategies and programmes, capturing with the use of indicators the impact of a public intervention on the social situation and not expenditure or activities embodied in outputs is of a crucial importance. Features of good indicators of public policy impact on the area of social life, considered in detail by the authors, led them to the catalogue whose synthesis is presented below[17]:

1) "Indicators should identify the essence of a problem and have a clear and accepted normative interpretation":
   - indicators task is not to sum up the entire reality but to capture the essence of the problem,
   - a receiver should be certain how he or she should assess the direction of changes presented by the given indicator,
   - indicators should be considered sensible by the public opinion (suggested participation in setting indicators),
   - indicator should be characterised with intuitive accuracy,
   - indicators should correspond with specific objectives of policies and programmes,
   - indicators should have a form allowing to establishing planned targets and assessment of the level of their implementation.
2) "An indicator should be robust and statistically validated":
   - indicators should be based on credible data which were not subject to arbitrary corrections,
   - if the data come from polls, they have to be representative and meet the highest methodological standards in terms of selection of sample, construction of questionnaires, survey implementation, processing data and analysis of results,
   - minimising survey errors should be aimed at, especially avoiding loading the results with a systematic error,
   - indicators which are strongly influenced by unpredictable (e.g. weather) and inexplicable factors should be avoided,
   - indicators whose value to a large degree does not depend on public policies but on independent factors, e.g. business cycle should be treated with caution.
3) "An indicator should be responsive to effective policy but not subject to manipulation":
   - an indicator should change its value, if the intervention brings results,
   - it should be possible to connect indicators with actions,
   - indicators should not manifest apparent effects (e.g. proportion of entities covered with a benefit may be easily increased granting the minimum benefits on a large scale.
4) "An indicator should be measurable in a sufficiently comparable way across member states, and comparable as far as practicable with the standards applied internationally by the UN and the OECD". Measuring an indicator should not be too high burden for the state, state enterprises or citizens – therefore creativity in using administrative data and already available statistical data is required.

[16] The team consisted of: T. Atkinson, B. Cantillon, E. Marlier and B. Nolan.

[17] T. Atkinson, B. Cantillon, E. Marlier, B. Nolan, *Social Indicators...,* op. cit., 21–25.

5) "An indicator should be timely and susceptible to revision":
   - the basic macro-economic data are usually quite actual, although other data often do not satisfy the criterion of being sufficiently actual,
   - developing in the state appropriate statistical potential which would allow to provide on time proper information is very important,
   - if there is a change of definition of the policy objectives or a progress in understanding the sense of phenomena, it should be reflected in determining indicators.

What is more, Atkinson and his colleagues formulated principles concerning specifying a set of indicators:

6) "The portfolio of indicators should be balanced across different dimensions". A set of indicators should be reasonable in terms of the number of dimensions covered by it, although it has to include all possible key problems of a particular social issue.
7) "The indicators should be mutually consistent and that the weight of single indicators n the portfolio should be proportionate".
8) "The portfolio of indicators should be as transparent and accessible as possible to the citizens of the European Union".

This set of criteria is adjusted better to the domain of public policies than the presented earlier set of SMART criteria. However, we can easily observe that the key dimensions of assessment are still the classical, methodological criteria of accuracy and reliability of measurement. Indicators of objectives of public policies most of all have to be indicators correct in the general methodological sense of this word, which is often forgotten. As a consequence, it is also not remembered that indicators determined in the process of operationalisation of terms which is conceptual search for the method of expressing the examined notions (including objectives) in the form of their measurable manifestations. Becoming aware of this fact opens the door to comprehensive achievements of methodology of social research. One of the basic consequences of this indicator "enlightment" should be abandoning "choosing indicators for objectives from existing compilations in favour of conceptualising indicators in the process of determining clear and measurable objectives of an activity.

Conclusion

The above considerations show that indicators are only tools- direct or indirect - of determining the level of implementation of particular dimensions of an objective of a conducted undertaking. The problem with evaluation and even broader, with the whole process of designing and managing public programmes, is not a result of poor information qualities which indicators are accused of and the fact that they are not able to reflect well the quality of programmes. It is rather the effect of their separation from the programme essence. This problem concerns to a lower degree measuring expenditure since it is governed by accounting principles and capability to report undertaken actions. Moreover, to a larger degree it refers to measuring results, on which we have focused.

One may agree with many reservations concerning measurement of results, formulated by some evaluation experts and especially by representatives of the constructivist school treating evaluation as a social intervention and practical organisational change. In fact, it is often difficult to capture properly all important aspects of an evaluated programme with the use of quantitative indicators. Therefore, it is necessary to be open to multiplism – using many research methods in order to obtain a comprehensive

and deep view of evaluated programmes[18]. Nevertheless, the criticism of measuring effects with the use of indicators as an important component of evaluation, which may be encountered in literature and among practitioners, should not take the attention away from the essence of the problem – low quality of designed public policies and programmes. In Poland, it is necessary to place emphasis on measuring results, and in particular on formulating objectives of strategies, programmes and projects in a clear way which at the same time is possible to be expressed with the use of an indicator set, allowing specifying whether and to what degree a programme objective was achieved. This disciplines persons creating projects and decision makers, allows assessing the rationality of allocating resources for a programme and comparing effects with other projects.

Objectives of strategies, programmes and projects which are well defined and expressed through measurable indicators are in practice a condition for conducting a reasonable *ex-post* evaluation because one of its main tasks is assessing whether and how the programme objectives were implemented. Obviously, it should go further: specify relations between actions and results, verify whether theoretical (or quasi-theoretical) assumptions of a programme were adequate, and provide knowledge about causes of failures which may become a lesson learned for the future worth the money spent on it. Evaluation should stimulate changes in public policies, management and allocation of resources. However, the matter which may be brought by a decent evaluation (including *ex-ante*) is improvement of standards of designing public undertakings, especially concerning precision of formulating objectives. We know a lot about this, although in spite of his, serious programmes are created and implemented without a serious logic matrix which would be a technique organising consideration of objectives, indicators and monitoring and, as a consequence - evaluation. This state of affairs has to be changed.

Public programmes and projects, as well as measurement of results and evaluation, are usually connected with allocating activities, while the regulatory activities are not treated analogically - creating laws, regulations and other legal acts – although their results are usually calculable. It should become a rule to treat introduction of legal regulations as a project which should also have clearly defined objectives and system of measuring effects.

Measurement of achievements of public policies is not an easy task. It is the simplest task to determine expected outputs of a programme or project because they are a direct, tangible or at least possible to describe result of undertaken activities. It is more difficult to measure impact of an intervention on solving social problems to which the intervention is dedicated.

The ongoing debate with Laeken on indicators shows that applying these rules is not simple, and most of all, that selecting indicators is a function of the definition of objective of public policies, and therefore it is the choice of the strict political sense, however subject to methodological rigour. Therefore, selection of indicators, especially those eventually specifying the programme impact may not be separated from the political debate on the basic, strategic development goals. Shortly speaking - **separating indicators from the policy is a source of impotence of the measurement of results and evaluation**.

At various levels of state and self-governmental administration, diagnoses of the situation for the needs of development strategies are created. Their authors flood decision makers with statistical data which are values of indicators not at all connected with the finally approved, *expilicite* or *implicite*, objective of an activity. Therefore, the diagnoses do not often have large significance and are treated as a kind of a decoration. The effort to develop and approve a set of indicator for particular areas of

[18] Compare T.D. Cook, *Postpositivist critical multiplism*, in: R.L. Shotland, M.M. Mark (ed.), *Social Science and Social Policy*, p. 21–62, Sage Publications, Thousand Oaks 1985.

public policies should be taken (solutions prepared in the European Union will be partially accepted; however, they will not meet all needs). Also the involvement of the Central Statistical Office in gathering required data and transforming them into indicators approved for the needs of public policies should be increased. These indicators should be factual enough to have a practical significance for taking decisions. An alternative, although not always available, is commissioning required surveys to reputable scientific or research centres.

Emphasising the role of indicators of not only *ex-post* evaluation but in the whole process of designing and implementing public policies, it cannot be forgotten that although objectives have to be formulated in relation to indicators or even through them, decision makers should "keep in mind that the aim of the programme is to achieve objectives, not indicators"[19]. Determining the objectives, one should act according to the following principle: the aim is what we are able to achieve and have an ambition to achieve and not, accepting maintenance approach - what we know we will achieve anyway.

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.

Karolina Keler – is a graduate of Sociology of the Jagiellonian University (specialisations: social work and sociology of economy and market research), a doctoral student at the Centre of Economy and Education Sociology in the Institute of Sociology of the JU, lecturer at the Tischner European University, conducts classes concerning social research.

She participated for example in the pilot mid-term evaluation of Measure 1.2. and 1.3 of the Sectoral Operational Programme Human Resources Development 2004-2006 and in the research project Małopolski system of analyses and programming policies of labour market, devoted to constructing a system of indicators of social exclusion and labour market for Małopolska, implemented within the Integrated Regional Operational Programme. Moreover, in 2005-2007 she participated in three editions of the international course: „Social Exclusion in Europe". At present she coordinates research devoted to the system use of data collected by the Municipal Welfare Centre in Krakow for diagnostic and evaluation purposes.

[19] ILO, IPEC Guidelines. Identifying & Using Indicators Of Achievement, ILO 2003, p. 3, http://www.ilo.org/public/english/standards/ipec/themes/timebound/downloads/pp5_5en.pdf.

Sources

Atkinson T., Cantillon B., Marlier E., Nolan B., *Social Indicators. The EU Social Inclusion,* Oxford University Press, New York 2002.

Atkinson A.B., Marlier E., Nolan B., *Indicators and targets for social inclusion in the EU*, Journal of Common Market Studies 2004, vol. 42, no 1, p. 47–75.

Birckmayer J.D., Weiss C.H., *Theory-based evaluation in practice. What do we learn?* Evaluation Review 2000, vol. 24, no 4, p. 407–431.

Council of Europe, *Concerted Development of Social Cohesion Indicators: Methodological Guide*, Council of Europe Publishing, Strasbourg 2005, http://www.coe.int/t/dg3/socialpolicies/socialcohesiondev/source/GUIDE_en.pdf.

Cobb C.W., Rixford C., *Lessons Learned From The History Of Social Indicators*, *Redefining Progress*, San Francisco 1998, http://www.rprogress.org/newpubs/1998/SocIndHist.pdf.

Cook T.D., (1985) *Postpositivist critical multiplism*, in: R.L. Shotland, M.M. Mark (ed.), *Social Science and Social Policy*, p. 21–62, Sage Publications, Thousand Oaks 1985.

European Commission, *Indicators For Monitoring And Evaluation: A Practical Guide*, The New Programming Period, 2007–2013: Methodological Working Papers, Draft Working Paper [X], DG Regional Policy 2006, http://ec.europa.eu/regional_policy/sources/docoffic/working/doc/indic_sf2007-2013_en.pdf.

ILO, *IPEC Guidelines. Identifying & Using Indicators Of Achievement*, ILO 2003, s. 3, http://www.ilo.org/public/english/standards/ipec/themes/timebound/downloads/pp5_5en.pdf.

Nowak S., *Metodologia badań społecznych, (Methodology of social studies),* Państwowe Wydawnictwo Naukowe, Warsaw 1985.

Sandhu-Rojon R., *Selecting Indicators for Impact Evaluation*, UNDP, Paris 2003,: http://www.npi-connect.net/c/document_library/get_file?p_l_id=18214&folderId=23505&name=DLFE-2004.pdf.

Maciej Szałaj

Mechanisms of selecting assessment criteria for *ex-post* evaluation

Introduction

The evaluation within the meaning adopted in Poland is understood as the assessment of public interventions. This interpretation encounters many voices of criticism on the part of practitioners and theoreticians. They usually point to the strong connotations of the Polish word "assessment". In fact, evaluation should not be perceived only in terms of concepts associated with this term because of two reasons at least. First of all, this approach is wrong from the cognitive perspective, because the role of evaluation goes far beyond assessment or even control functions. Secondly, adopting such perspective makes it difficult to root evaluation in the local cultural context. The presence of evaluators in the institutions involved in the researched projects is often taken as a potential threat; whereas the evaluators as such are identified as representatives of the institutions entirely external to examined reality.

A feature distinguishing evaluation from academic studies or typical control activities is exactly its organic relation with society through a network of specific dependencies. The assessment component is naturally integrated in every evaluation, but in no case does it cover the whole evaluation process. Evaluation consists of defined elements, crowned only in the end by assessment and recommendations. One of the sensitive points of evaluation process is the assessment criteria which are the subject of this paper.

The first part of the present paper discusses the place of assessment criteria in the general logic of the evaluation process. It allows to indicate a specific "identity core" common to every evaluation, regardless of the specificity of the researched object or the adopted research method, as well as irrespective of the approach specific for a given case or the area of public intervention. The second part of the paper presents the classification of assessment criteria and the dimensions determining their selection. The last part in turn is the analysis of the specificity of the selection of criteria for the needs of *ex-post* evaluation.

By using the theoretical background outlined in the paper an attempt will be made to decide whether it is possible to indicate one uniform standard for selecting proper criteria and what features should these criteria have.

Assessment criteria under the logic of the evaluation process

The nature of every research is the pursuit of cognition, subordinated to defined canons of conduct, which are usually gathered in bulky textbooks on methodology or in somewhat less voluminous codes of good research practice. With regard to methodological reflection, evaluation has reached the stage of

self-awareness, matching the position of well-established scientific branches of a centuries-old tradition. Evaluators carry out research in many different areas, who profess different research approaches and often values that collide with each other, act in practice in line with a uniform model which consists of many stages, some of which are common to many research fields, and others are entirely specific to evaluation research. The latter category covers the definition of assessment criteria.

D.M. Fourier defines the above-mentioned model of the research process as the general evaluation logic model[1]. Apart from selecting criteria it consists of the following elements:

- identification of quality standards,
- measuring effects and comparing them with the standards,
- synthesis and assessment of obtained results.

This general logic is present in all evaluation research. Even the projects implemented within the framework of the trends perceived as standing in opposition to the so far commonly adopted research assumptions, are not an exception to this rule. The framework marked out by this logic defines the boundaries of the evaluation's "identity core." The author quoted above described the role of the general evaluation logic in a vivid manner by writing that "[the general logic of reasoning] determines the rules of the game which must be observed by all evaluators conducting research in any area. It also determines the peculiarities of the game called evaluation. Therefore, anyone who claims that he/she conducts evaluation must thereby indicate the criteria and quality standards, and then identify and measure the effects in line with the above, as well as carry out a synthesis of the obtained data in order to perform assessment[2]."

On the basis of the uniform set of rules of the game, the identity of the discipline directly translates into development of professional identity of the evaluators that come from very different environments. The well-established methodological identity also provides the raison d'être for professional associations and networks which, as a consequence, gain grounds to create legitimate canons of evaluation practice. Defining the assessment criteria constitutes the first stage of implementing projects which are subject to verification in the light of generally acknowledged standards. These enable to answer the question whether the evaluator makes any professional errors. However, the canons of practical action fail to determine how exactly a given project should be designed and implemented, because it is impossible to indicate one appropriate set of questions, research methods, assessment criteria or quality standards.

The limited scope of guidelines following from the methodological canon illustrates the differences between the general evaluation logic and the working logic of individual projects. The working logic, as opposed to the general one, is differentiated depending on the object of evaluation and the approach used towards it. It may be assumed that between the general logic and the working logic there is a relationship similar to that for strategy and tactics. Hence, the strategic element, i.e. assessment criteria, has its tactical counterpart in the form of specific sets of criteria indicated in the course of the evaluation practice. The literature on the subject distinguishes between two understandings of the working logic[3].

According to the first one, working logic is defined as a set of parameters concerning the research problem, phenomenon, evaluation questions and statements. The resultant of these two parameters defines the specific nature of individual research paradigms. They may also form a basis to indicate the

---

[1] D.M. Fourier, *Establishing evaluative conclusions: a distinction between general and working logic*, in: E. Stern (ed.), *Evaluation Research Methods*, vol. 1, Sage Publications, London 2005.

[2] Ibidem.

[3] Cf. ibidem, pp. 56–64.

dimensions according to which the theoreticians built their approach to evaluation practice. The selection of the assessment criteria for the needs of a given project is determined through the understanding of the phenomenon of the researched area that is typical of the defined approach. In this context the phenomenon should be perceived as the working concept of the nature of the research object defined on the basis of the adopted definition of the major research problem.

The relationship between the phenomenon and the assessment criteria is the key element of each evaluation, which determines the character of subsequent stages of project implementation, because the adopted criteria define the framework of the design of the research process until the moment of data synthesis and assessment formulation.

The second understanding of the working logic relates to the argumentation structure, which defines the reasoning scheme carried out according to defined parameters.

The place of assessment criteria in such understood working logic is not precisely defined. Their role consists rather in co-designation of the context of the project implementation. The impact of criteria selection is visible, for example, in the research methods and in the manner of collecting evidence because it defines the type of the data sought (e.g. assessment according to the criterion of effectiveness of achieving the assumed objectives implies a search for data on the effects of interventions which fit into the categories determined by the objectives). The perspective defined by the assessment criteria may constitute a source of legitimacy for the research, of which the methodological correctness cannot be assessed simply with the use of the usual scientific standards. The existence of assessment criteria in the research context is, in general terms, probably the clearest illustration of the values engaged in the project. It demonstrates well the differences between evaluation and the academic fields which provide theoretical inspirations, research tools and, finally, staff for the evaluation.

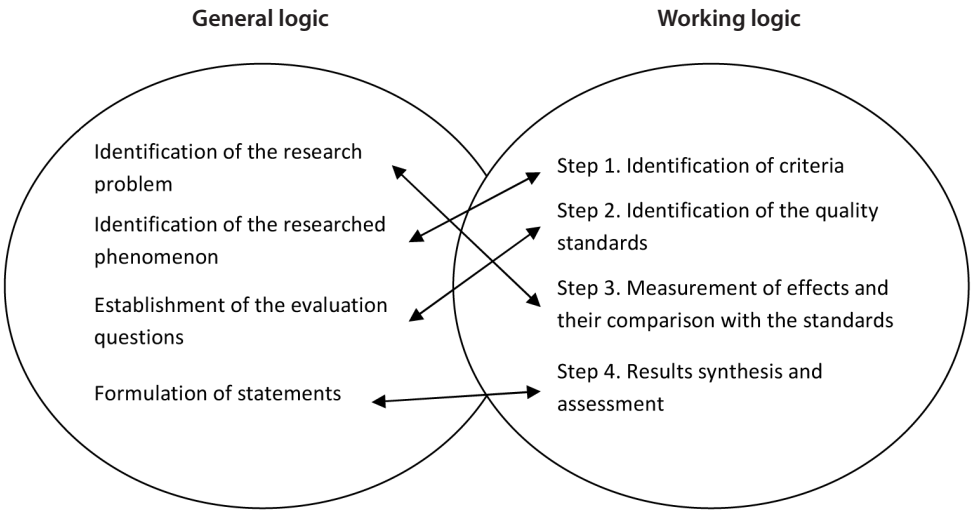


Fig. 1. Relationship between the general and working logic

Source: Own compilation on the basis of: D.M. Fourier, *Establishing evaluative conclusions...*, op. cit, p. 58.

Types of criteria

The issue of the types of criteria used in evaluation may seem an insignificant matter. According to some opinions the selection of additional criteria is determined only by specificity of projects, thereby the establishment of general typologies and uniform definitions defeats its own purpose. However, the issue is complicated by the fact that Polish sources, which enumerate the set of the criteria most frequently in practice, fail to describe fully the conditions determining their universal character[4]. Another problem is associated with the reception of English notions (efficiency, effectiveness, utility, relevance, etc.). In direct translation they do not seem natural which creates a kind of terminological chaos. Additional reason of misunderstanding is a widespread belief of evaluators that the names of criteria do not have a uniform interpretation and their meaning may be interpreted with great freedom.

Therefore, making an attempt to put order to terminology accompanying the assessment criteria will have a positive impact on the clarity of the subsequent part of this discussion. The classification presented below is based on the division into two most basic aspects of all public interventions – outlays and effects.

Adopting these aspects as the axis of the division between different types of assessment criteria we obtain two major categories and one intermediate category.

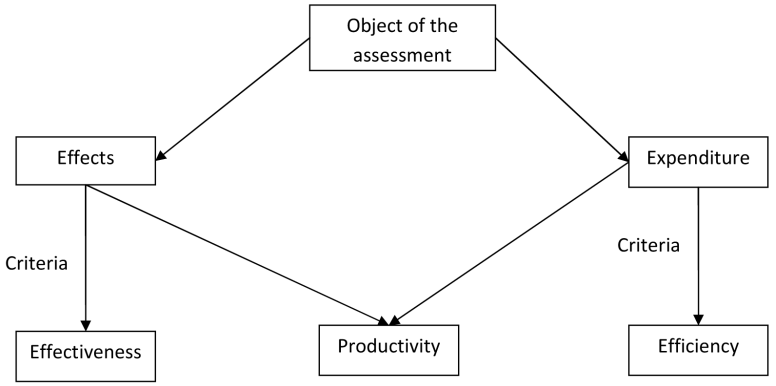


Fig. 2. Main categories of criteria

Source: Author's own compilation on the basis of: E. Vedung, *Merit criteria and performance standards*, in: E. Stern (ed.), *Evaluation Research Methods*, op. cit., pp. 183–200.

The first major category covers efficiency, i.e. broadly conceived criterion defining the relationship between the outlays and the effects achieved. This category is not the source of significant notional misunderstandings because its understanding is uniform and divergent approaches are rather uncommon. Assessment based on this criterion is carried out with the use of two major methods:

- cost-benefit analysis, which is based on calculation of effects expressed in monetary units in relation to outlays expressed in monetary units,

[4] Examples of sources: websites of the Implementing Authority for European Programmes: http:// www.wwpwp.gov.pl/index.php?params%5Bsection_id%5D=4; Polish Evaluation Society, *Ewaluacja funduszy strukturalnych [Evaluation of Structural Funds]*, developed for the Division of IROP Monitoring and Reporting, Department for Regional Programmes Implementation, Ministry of Economy, Labour and Social Policy, Warsaw 2003, pp. 37–38.

- cost-effectiveness analysis, which are a calculation of the effects expressed in physical values and costs expressed in monetary values.

An intermediate category of criteria in line with the discussed classification is productivity. This criterion takes into account both effects and costs. However, it is not completely adequate to the specific nature of evaluation research and thus it is not defined as a standard value assessment criterion. Productivity covers the analysis of the outlays incurred but it assumes their very narrow understanding – only as financial costs.

Another major category will be effectiveness, understood however in a broader than usual way when it is identified as the assessment of the degree of implementation of direct quantitative objectives of an intervention. This category includes many sub-criteria, whose common feature is the assessment range focused on differently understood effects. It should be noticed that this assumption is met both for the criteria of utility, impact, sustainability and even effectiveness understood in a narrower sense which was described above.

The criteria taking into account the effectiveness of achieving effects are probably in place in every evaluation. However, each sub-category of this criterion has its unique properties. The scope of their applicability is, therefore, subject to restrictions that are possible to determine. The subsequent part of the paper is devoted to the relationships between the selection of the assessment criteria and different dimensions determining the specific nature of a given project.

Criteria selection

Selection of a set of criteria is a very important stage of structuralisation of each evaluation research. Omitting this stage would prevent the evaluator from working out detailed methodological assumptions and research tools. Contrary to the popular opinion, the selection of criteria never consists in the direct use of guidelines or indications from textbooks. There are no canonical solutions mandatory in respect to specific commissioning institutions or areas of intervention. Each time, however, the criteria selection is determined by the characteristic nature of the research problem, which is defined by the evaluator at the first stage of his work.

The following aspects determine the scope of assessment to the greatest degree:

- function of evaluation;
- time of research;
- scope of research;
- research paradigm.

Each of the enumerated aspects has a specific impact on criteria selection. The final shape of the set of criteria results from these impacts and determines the peculiarities of individual evaluations.

The functions of evaluation have been modified to a large extent along with the changing cultural context. Thereby, similarly to social sciences, evaluation has undergone a revolution which revised the fundamental methodological assumptions. At the moment, four major functions of evaluation are enumerated: cognitive, conclusive, socio-political and administrative[5].

[5] Cf. Polish Agency for Enterprise Development, *Ocena (ewaluacja) programów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej [Assessment (evaluation) of socio-economic programmes against the background of Polish accession to the European Union]*, published by the Ministry of Finance from the grant of the World Bank allocated to support the system of monitoring and assessment of the use of the European Union assistance funds and other assistance funds, Ministry of Finance, Warsaw 2002, http://archiwum.parp.gov.pl/raport11.html, pp. 18–21.

In the case of the first of the listed functions, the emphasis is put on assessment criteria, which give raise to questions concerning the cause and effect (questions focused on the relationship between activities and the obtained material effects) or predicative questions (predications of material effects taking into account the implementation system and the context). The criterion of relevance corresponds to the research needs defined by the cognitive function of evaluation[6]. Its most popular form is used in *ex-ante* evaluations aimed at the assessment of the implementation system and intervention project. In such case the criterion of relevance has predicative questions assigned to it, which concern the most probable expected effects of the assessed project. The cognitive function may be, however, exercised also through the relevance assessment carried out at later stages of the intervention cycle. Then, the assignment of cognitive functions to the assessment targets the research process at empirical verification of the assumptions adopted during the programming stage.

The conclusive and formative function have an instrumental nature and hence they assume that the results will be used during the decision making process. Evaluations, in which they are inscribed, are focused on the quantitative effects. They imply the use of the effectiveness criterion understood as the assessment which takes into account the degree of achievement of the assumed material objectives.

Administrative function has a similar nature, because it addresses the needs of institutions related with public management. Inscribing this function into the process of assessment is associated with the expectation that evaluation will be directly translated into administrative practice. Then the effectiveness and efficiency criteria are applied. The application of the former is justified with similar reasons as for evaluations carrying out formative and conclusive functions. The efficiency criterion, on the other hand, forms a basis to assess the management of public resources. As a result of such an assessment, funds may be reallocated within the budget of a given intervention. Reallocations also often result from evaluations that perform a conclusive function; hence, also in their case the criterion of efficiency is applied. However, at this point, efficiency is understood usually within the cost-benefit categories, while the administrative function better corresponds to efficiency understood as cost-effectiveness calculation. The use of the effectiveness criterion in this shape is usually an answer to the demand for social accountability, which is the major premise for the conclusive function of evaluation.

The last of the listed functions, i.e. socio-political function, is rarely used by the public institutions, while it is more often used by non-governmental organizations (NGOs) or independent think tanks. This function excludes complete objectivity because its implementation foresees promotion of a defined idea or a systemic solution. Of course, socio-political evaluations cover identical methodological regimes as any other evaluation research. The research concepts, and thus also the criteria selection process are, however, in this case subject to defined objectives because they aim at providing results and recommendations which have impact on the decision-making process in a predefined manner.

The time of the research is a dimension which constitutes the basis for probably the most popular classification of evaluation. This division was established in the documents determining the conditions of implementing the structural funds. In the 2004-2006 programming period, three types of evaluation were defined, these are: *ex-ante*, *mid-term* and *ex-post* evaluation[7].

It is assumed that the *ex-ante* evaluation is the most appropriate **criterion of relevance**. Grounds for such statement are obvious, because relevance is the only general criterion that does not refer to

[6] Relevance is understood in this case as an assessment criterion taking into account the adjustments of the objectives assigned to a given intervention to the context of implementation and in particular to the problems identified at the programming stage.

[7] Council Regulation (EC) No 1260/1999, Articles 40-43.

actual effects but only to the expected ones. In this case, assumptions are assessed and the assessment itself is especially deeply rooted in the specialist knowledge of the evaluators or their consultants. As a result *ex-ante* evaluation consists, exclusively, in formulating forecasts. At this point, however, one should consider an idea, which was already signalled before, about the possibility to carry out an assessment of the relevance at different stages of the intervention cycle.

The assessment of the adopted assumptions constitutes often an element of the *mid-term* or *ex-post* evaluation. At that time, the validity of the project logic adopted at the beginning can be verified, as well as its adequacy in the light of the changing context. Also, the assumed direct and general objectives may be assessed, which, on many occasions, significantly enriches the utility of recommendations and hence the entire evaluation.

As regards, evaluations carried out after launching the operational stage, **the criterion of utility** may constitute a supplement to the assessment of material effects. For this criterion the assessment is based on the opinions of final beneficiaries which refer to the satisfaction from the received support. The criterion of utility is sometimes called "the mirror reflection of the criterion of relevance." According to the author of this paper this comparison is only partly justified, because it suggests that we have to deal with basically the same criterion which is used at different stages of the intervention cycle. However, these are not identical criteria, despite the fact that they have a common distinguishing feature, i.e. the range of assessment is not targeted at direct effects. The assessment of utility and relevance may thus, undoubtedly, constitute separate elements of the *mid-term* and *ex-post* evaluation and jointly supplement the assessment based on material effects, at the same time fulfilling different complementary functions. The application of the criterion of relevance in such case would be, for example, a continuation of the experts' assessment initiated during the *ex-ante* evaluation. This criterion would have some cognitive functions assigned to it, whose implementation requires the involvement of objective expert knowledge that can have a dual use, at the same time strengthening the formative functions; whereas the assessment of utility by its very nature strengthens the conclusive functions which thereby open the road to a more democratic control of the public management process. Adding utility to the catalogue of criteria is also an instrument of the reconstruction of the entire paradigm of administration according to the model of the *New Public Management*, which calls for commercialisation of public services aimed at maximization of the satisfaction of their recipients. This digression is important because the popularity of evaluation in Europe should be considered as the result of the reformist zeal of the eighties when the demands of the *New Public Management* gained broad acceptance in a large part of this continent.

Despite their strong impact on the shape of many evaluations, as well as their role played in the shaping of the entire evaluation culture, the assessment criteria basing on the programme assumptions or opinions usually perform a supplementary function. The assessment conducted after starting the implementation of the intervention is dominated by the criteria referring to the material effects. Both in the *mid-term* and *ex-post* evaluation the usually used criteria are **the criteria of effectiveness and efficiency** (assessment of the effects at the level of product and result). The *ex-post* evaluation may be extended with the assessment of effects at the level of impact. Therefore, textbooks name **sustainability**[8] as the most typical **criterion** for this type of evaluation. Its application enables to decide on this sensitive issue, which cannot be determined at the earlier stages of the cycle. Assessment of sustainability makes it possible to answer the question if the public policy in the given area favours the real solving of the problems. Low

[8] Sustainability is understood in this case as ability of the intervention effects to survive after the external financing is completed.

sustainability may point, for example, to the occurrence of the phenomenon of learned helplessness (social policy) or lack of the expected institutional development (pre-accession programmes, technical assistance). The need to obtain this type of knowledge was one of the key conditions determining the systemic placement of evaluation in the mechanisms of implementation of structural funds.

Scope of the research is another factor heavily determining the selection of the assessment criteria. Evaluations carried out at the level of central institutions or at the transnational level are in most cases focused on a very broad range of issues and they assess the entirety of effects generated by many interventions. Therefore, the evaluations of the sector programmes carried out by the EU Member States apply the criterion of impact. Assessment based on this criterion gives grounds to draw conclusions on the scope of synergy and, thereby, on the manner of structural policy management. Reliable and systematic use of the criterion of impact in large-scale evaluations is particularly recommended in administrative cultures marked with the "syndrome of sectoral division", i.e. lack of the coordination ability.

In a slightly different form, the impact assessment is also used for smaller-scale evaluations carried out, for example, at the level of individual implementing institutions or regional authority institutions. Then the evaluation is dedicated to the relationship between the impact of the effects generated at the level of individual projects and the general objectives of the given area of the public policy. Applying the criterion of impact in this type of evaluations verifies the relevance of the systemic placement of institutions carrying out the evaluation. Moreover, the assessment may also cover the intervention logic at its junction which links the neighbouring implementation levels.

The other criteria relevant for large-scale evaluation are also specific **horizontal** criteria. Their use aims at assessing the implementation of the defined objectives that are assigned a meaning common to all public policy. In the case of large-scale evaluations of the EU structural policy, the assessed elements cover, for example, the scope of the added value, impact on giving equal opportunities to women and men or the impact of the intervention on the natural environment.

The scope of evaluation is also determined by the use of different type of **specific criteria**. However, in this case the differences do not result from the level of placement of the implementing institution but from the range of the research spectrum. The above-mentioned classification of evaluation, which is based on the time of the research, points to an evaluation of cross-cutting nature. However, the thematic scope of the assessment tends to be limited to strictly defined issues which usually constitute the areas identified as especially problematic during implementation. These types of evaluations are *on-going* in their nature and the criteria used in them refer to strictly defined aspects or they constitute a detailed form of general criteria.

The selection of a given **research paradigm** is in practice of evaluation research identical with defining the researched phenomenon. Therefore the specific nature of the assessment criteria is determined by an assumption resulting directly from the adopted paradigm. For example, using the paradigm of goal-free evaluation or a paradigm of multi-goal evaluation is a very strong reason to focus the assessment on the criterion of utility; whereas, the cause and effect paradigm assumes an assessment based, primarily on the criteria of effectiveness of achievement of the objectives and on relevance.

Until now only the substantial conditions for selecting criteria have been discussed. The **organisational conditions** should also be indicated, which in evaluation practice have a very definite impact on the shape of the research. The selection of criteria must be carried out according to professional standards requiring each evaluation to fulfil, i*nter alia*, the conditions of utility and feasibility. Thus the set of criteria must be adjusted to the time-related and financial restrictions. Failure to consider the organisational conditions has a direct impact on the shape of the final product of evaluation. It is necessary to reach

a compromise in the area of methodology if there are significant restrictions. Then the final scope of assessment determined by the criteria should be defined only after the compromise is accepted. In such case it is indispensable to fulfil this assumption to be able to talk about "good evaluation" – one which will not constitute a disappointment to the contracting party and a reason for contractors to lower their professional standards.

Selection of criteria in the *ex-post* evaluation

The peculiarities of *ex-post* evaluation in every aspect, falls also within the scope of the selection of criteria, which are certainly a result of the specific time of implementation. As indicated in the previous part of the paper, evaluations carried out after the completion of the intervention naturally focus on the assessment of actual effects at all levels, including also the level of impact.

Although the above-mentioned feature determines the specific nature of the selection of criteria in *ex-post* evaluation, it is obviously not the only dimension targeting the selection of criteria on the needs of individual projects. The selection of criteria is determined by all of the above-mentioned dimensions (functions of evaluation, scope, paradigm, organisational conditions).

The European culture of evaluation research presents a very deeply rooted association of *ex-post* evaluation with the complex assessment carried out by the European Commission after the completion of the programming periods. It is a manifestation of the dominant role, which is played by the structural policy of the Community in the current history of evaluation in Europe, with its obligation to carry out an assessment inscribed in the implementation system. However, although the importance of structural policy for the development of evaluation and for the modernisation of public affairs management in Member States is undeniable, the "organic" fusion of *ex-post* evaluation with the activity strictly defined by legislation should be considered undesirable. Because in such circumstances the practical reasons, which should be the most important premise defining the relevance of the performed evaluation research and its nature, recede into the background

The subject and scope of *ex-post* evaluation may, therefore, become quite different in form than it was provided for in the regulations of the EU Council. This remark refers to all elements and stages of the assessment process, hence also to the selection of criteria. In order to illustrate the differences following from a different scope and function of evaluation, an analysis of criteria selection is carried out for evaluations performed at different levels of the EU funds implementation system.

It has already been mentioned that in the continental research tradition, *ex-post* evaluation is identified with the assessments carried out by the European Commission. The specific nature of this type of evaluation is defined by the regulatory framework of the structural funds implementation. The regulations also provide for the scope of evaluation and they define the entities responsible for preparation and performance of the assessment. The Council Regulation (EC) No 1260/1999 also defines the criteria which should be considered in the assessment, including effectiveness of implementing the effects at all levels (products, "utilisation of resources", results, impact), efficiency ("efficiency of the assistance") as well as horizontal criteria applied to assess the impact of the EU Structural Policy on strengthening the socio-economic cohesion within the Community[9].

The provisions of the Regulation define the nature of the *ex-post* evaluation established by the Commission not only by indicating the list of criteria but also by determining the scope and time of

[9] Council Regulation (EC) No 1260/1999, Article 43.

research. The object of the assessment in this case is defined very broadly as the entirety of activities covered with financing from the structural funds.

However, specific functions of this assessment were not established. They follow however from other provisions and contexts of carrying out evaluation. The expectation that evaluation will provide knowledge on "the factors contributing to the success or failure of performance[10]" points to the cognitive function. If the assessment component is to be inscribed into it, then the catalogue of criteria will have to be extended to cover relevance. At this point we go back to the above-mentioned issues of applicability of this criterion in the evaluation conducted after completing implementation. Its application in the discussed *ex-post* evaluation is not only possible but it even constitutes a necessary condition to fulfil all of the expected functions of evaluation.

Apart from the cognitive function the evaluation established by the Commission is characterised by, above all, instrumental functions, especially the conclusive function. Given the time of carrying out research, the formative function is inevitably restricted. Within the criteria, the dominant role of instrumental functions is expressed in a manner typical of them, i.e. by displaying the effectiveness criterion which is indicated both in the detailed criteria of *ex-post* evaluation[11], as well as in general provisions devoted to the assessment process[12].

The above conditions specific for *ex-post* evaluation prescribed by regulations defining the structural policy implementation framework, imply also a specific understanding of the researched phenomenon, hence they have a clear impact on the working logic of evaluation. As a result, the assessment criteria will be used at the stage of conducting research in a manner following from the working nature of "the Commission's" *ex-post* evaluation. Outside the research area marked out by Community regulations, remains a vast space, which is the space for the *ex-post* evaluation projects that are complementary with the activities of the European Commission. In case of these projects the selection of criteria will be, at least, partially similar to that described above, which follows directly from the features common to every *ex-post* evaluation. However, a different working logic will cause that these criteria will become a starting point for different evaluation questions and research problems. This phenomenon is reflected in the cases recorded in Polish evaluation research. Domestic public institutions started to carry out *ex-post* assessment a few years before the accession in response to their information needs at that time that arose during the implementation process. As a result Polish *ex-post* evaluation had, in general, a less cross-cutting character than the evaluation carried out at the EU-wide level. The scope of evaluation was restricted to defined programmes, so the criterion of impact was rarely used; whereas, the assessment performed based on this criterion focused on estimating impact of programmes on local or regional communities, not relying on macro-structural analyses.

The Polish Agency for Enterprise Development (institution implementing the PHARE Socio-Economic Cohesion Programme) was one of the institutions which carried out *ex-post* evaluation in pre-accession period[13]. The selection of criteria in *ex-post* evaluation implemented by the Polish Agency for Enterprise Development was determined by the conclusive function[14]. Therefore, the assessment

[10] Ibidem

[11] Ibidem.

[12] Ibidem, Article 40.

[13] Evaluation reports of the Polish Agency for Enterprise Development are available at the Agency's website at: http://www.parp.gov.pl/index/index/110.

[14] It covers orientation towards results, taking into account mainly the instrumental/control objective and the time-point perspective (surveying beneficiaries one year after completion of the intervention).

was predominated by the broadly-conceived criteria of effectiveness, including the sustainability that is typical of ex-post evaluation. Taking into account the large significance ascribed to the conclusive function, the evaluations carried out by the Polish Agency for Enterprise Development used in general the criterion of utility. The research approach reflected the elements of the multi-goal paradigm and it showed some features of the *New Public Management*. Adopting this perspective in this case followed to a large extent from organisational circumstances, because the role of institution maintaining direct contacts with final beneficiaries implied assigning greater importance to the issue of quality of products supplied to beneficiaries and the beneficiaries' satisfaction with these products.

The Polish Agency for Enterprise Development also used the criterion of relevance in its final evaluations. However, its application was determined by slightly different cognitive needs than in European-level evaluations. Since relevance did not relate primarily to a broad implementation context but to the programme assumptions what usually constitutes the object of assessment in *ex-ante* evaluation.

Different conditions are in place for *ex-post* evaluation of the pre-accession aid implemented by the Office of the Committee for European Integration[15]. The reason behind is that their scope was defined by the regulations concerning the PHARE implementation system. The applied approach was similar to the above-described approach of the Commission. Hence, it adopted a cross-cutting nature demonstrated by the emphasis put on in-depth assessment of effectiveness of the implementation of operational and direct objectives, as well as impact. At the level of criteria, it was manifested e.g. in the application of thematic criteria which refer to the implementation of the major programme objectives.

As was illustrated in the above-discussed cases, the selection of criteria for the needs of *ex-post* evaluation has its specific dynamics, which is determined by different factors. The very moment of carrying out assessment, which defines *ex-post* evaluation, has a very strong impact on it, determining the specific nature of the selection criteria in final evaluations. However, it is only one of many factors determining the selection of defined criteria of *ex-post* assessment. The reason for it is that in this type of evaluation all selection mechanisms, based on the previously mentioned key dimensions,, are applicable, while the statement that the boundaries of final assessment are precisely marked out is flawed.

Conclusion

The criteria constitute the specific evaluation "identity core." It is because they define the most important characteristics of every assessment. Thus, the methods of their selection determine to a very large extent the success of the entire research process. Therefore, an assumption may be made that the accurate selection of criteria preconditions "good evaluation." However, the practical problem consists in the fact that both the accurate scheme of criteria selection and good evaluation go beyond precise definitions.

A significant number of evaluation practitioners are of the opinion that searching for universal rules in evaluation is an activity as pointless as the attempts to develop a general definition of a good work of art; because the specificity of evaluation is as complex as the aesthetic or literary trends. Therefore, bringing all evaluations to a common denominator would require using a kind of a **meta-criterion** disregarding the specific assumptions adopted in specific cases. Referring to the already outlined general logic of evaluation, it may be assumed that the **purpose-fit** criterion could act as the meta-criterion.

[15] Evaluation reports are available at the website: http://www1.ukie.gov.pl/www/serce.nsf/DocByWykazZmian/5142FF4BF2348620C1256ED9004B0299?open.

The assessment of fitness for purpose is applied directly after completing the process of criteria selection. The relevance of the selection may then be verified with regard to the objectives set out after completing the stage of problem identification. It is, at the same time, the second of the critical points designating the place of the criteria in the evaluation process. Filling these sensitive points with the appropriate "tissue" requires refraining from applying schematic solutions and replacing this approach with extensive work performed by the evaluator, basing on a deliberate and conscious concept of the final product.

While addressing the issue of the evaluator's awareness as regards the final product of evaluation, it should be noted that a properly functioning system of evaluation is based on skilful coordination of activities carried out by individual evaluation bodies. The concept of assessment should, therefore, take into account in all cases the research plans of the partners. As a result there arises a need for intensive informal dialogue, within the framework of which the discussion on the selection of criteria will constitute a starting point for the common planning of the evaluation strategy. The following paper is a contribution to this discussion.

Maciej Szałaj graduated from the University of Warsaw (majored in sociology) and completed a postgraduate programme in project management at the Warsaw School of Economics. He works as a senior expert in the Evaluation Section of the Polish Agency for Enterprise Development. As a manager of evaluation research projects, he specializes in interventions supporting the SME sector and in measures aimed at developing the adaptability of enterprises. He participated in creating one of the first evaluation units in Polish public administration. He was a member Steering Group for Evaluation of the SOP/HRD (Sectoral Operational Programme Human Resources Development). Currently, he is involved in developing evolution system at the level the OP/HC (Operational Programme Human Capital). He is an author of articles exploring various issues linked with evaluation practice and theory. He also serves as an academic teacher and public administration trainer. As a panelist, he has participated in several seminars and conferences. Since 2008, he is a board member of the Polish Evaluation Society.

Sources

Fourier D.M., *Establishing evaluative conclusions: a distinction between general and working logic*, in: E. Stern (ed.), *Evaluation Research Methods*, vol. 1, Sage Publications, London 2005.

Polish Agency for Enterprise Development, *Ocena (ewaluacja) programów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej [Assessment (evaluation) of socio-economic programmes against the background of Polish accession to the European Union]*, published by the Ministry of Finance from the grant of the World Bank allocated to support the system of monitoring and assessment of the use of the European Union assistance funds and other assistance funds, Ministry of Finance, Warsaw 2002, http://archi-wum.parp.gov.pl/raport11.html

Polish Evaluation Society, *Ewaluacja funduszy strukturalnych [Evaluation of Structural Funds]*, developed for the Division of IROP Monitoring and Reporting, Department for Regional Programmes Implementation, Ministry of Economy, Labour and Social Policy, Warsaw 2003.

Vedung E., *Merit criteria and performance standards*, in: E. Stern (ed.), *Evaluation Research Methods*, op. cit., pp. 183–200.

Jacek Pokorski

Application of the CAWI technique in evaluation research

Introduction

The changes taking place all around us, which follow from the IT revolution, development of the mobile telephones and the Internet, turn out to be so extensive and dynamic that on many occasions they are referred to as culture shock. Until recently, it was impossible, unattainable and remote – within the lifespan of one generation became possible, attainable and familiar[1]. A radical change in the social communication, which occurred in the last years as a consequence of obtaining access to the global network, opens unparalleled opportunities as regards on-line gathering and processing of information concerning human activity in the virtual reality. In order to illustrate the scale of the phenomena in Poland it is enough to refer to some basic statistical data.

The Net Index data shows that approx. 17.3 million persons have access to Internet in Poland, however, it is worthwhile to highlight that within the last two years the increase in the number of Internet users amounted to 1.7 million persons per annum. Among persons aged 15-75 years the Internet users constitute 58%, whereas the Internet is used by 96% of persons in the age group 15-19 and only 14% of persons aged 60-75 years. Moreover, this media is daily used by 71% of persons having access to the net[2]. In the subsequent years the application of Internet to handle everyday matters increased gradually – in 2008 two thirds of Internet users stated that they do shopping in the Internet[3], and 90% of the surveyed users of the network aged 18-54 years also used the Internet banking services (personal account with on-line access, insurance, investment funds, loans)[4]. In the case of 60%, of users the Internet is also used to seek job (establishment of one's own professional profile, searching for job offers)[5]. It cannot also be overlooked that for Internet users the medium constitutes one of the basic sources of entertainment (watching video materials, downloading films and audio files, tracking the life of celebrities). The medium is more and more commonly perceived as an everyday good without which it is already difficult to imagine

---

[1] See the report entitled *Badanie zapotrzebowania na działania wspierające rozwój usług świadczonych elektronicznie (e-usług) przez przedsiębiorstwa mikro i małe [Research of the demand for measures supporting development of services provided via electronic means (e-services) by small and micro enterprises]*, Pentor RI for PARP, Warsaw 2009, p. 6

[2] See GfK report, *Net Index* of March 2009 (http://www.gfk.pl/podstrona.php?page=/page.php?id=235).

[3] See Gemius S.A report, *Polski Internet 2008/2009* [*Polish Internet 2008/2009*] of 2008 (http://pliki.gemius.pl/Raporty/2009/02_2009_Polski_Internet_2008_2009.pdf).

[4] Gemius S.A. report for Google Poland, *Google Poland Finance Study 2008* of September 2008. (http://www.Internetstats.pl/index.php/2009/01/internauci-o-uslugach-finansowych-google/#more-469).

[5] Millward Brown SMG/KRC report for the pracuj.pl portal, „Oferty pracy i poszukiwanie pracy w Internecie" *[Job offers and seeking job in the Internet]* of December 2008 (http://www.Internetstats.pl/index.php/2009/03/oferty-pracy-i-poszukiwanie-pracy-w-Internecie-pracuj.pl)

our life[6]. It seems that now no other mass socio-economic phenomena develops more dynamically on global scale than the virtualisation of society which thereby moves all possible human activity onto the Internet[7].

The dynamic growth of the significance of Internet within the last few years has a strong impact on the directions of development of the social and market research sector. In 2008, according to ESOMAR 1/5 of all expenditure for research were expenditure incurred on quantitative on-line research[8]. The market and social research implemented on the basis of the CAWI technique (Computer Assisted Web Interviewing) also gradually gains in significance within the last years, which is showed by market analyses conducted in Poland[9]. The increased demand for Internet research, undoubtedly, goes hand in hand with the progressive development of information society in Poland (including the dynamic development of the information and communication technologies, increased access to the broadband Internet, more and more numerous social groups use the network in the everyday works and to handle private affairs). It is a well-known fact that human activity, also in Poland, extensively enters and develops in the virtual reality and different manifestations of this activity on the field of social, economic, political and cultural actions constitute an ocean of countless research questions, which more and more often may be effectively asked only in the e-dimension. As the application of Internet becomes more and more popular among the growing number of Poles (both private persons, as well as different types of entities), increases also the popularity of research carried out in the net, via the net and for the purposes of different types of Internet projects[10].

Many prejudice have already appeared as regards applying Internet research to evaluation, despite they have a relatively short history in social studies, while in Poland also because evaluation is a relatively young field of research. After some first, not very successful trials of the Contracting Parties to use the evaluation results based on the CAWI research, the entities commissioning evaluation are especially careful in respect to this type of research offered by the Contractors in their technical (methodological) proposals of conducting evaluation. However, the CAWI technique can, in many cases, ensure the most efficient manner to measure the programme effects, gather the assessments and opinions of the beneficiaries or even fulfil the broadly-conceived reporting functions.

---

[6] See Millward Brown SMG/KRC report for ING, *Internetowy Barometr ING [Internet ING barometer]* of March 2009 (http://media.ingbank.pl/PressOffice/get/PressKit.PressKit.926.po?field=125125).

[7] See A. Mattelart, *Społeczeństwo informacji. Wprowadzenie [Information society. An introduction]*, Universitats, Kraków 2004.

[8] See ESOMAR, *Global Market Research 2009. Esomar Industry Report*, Amsterdam 2009.

[9] See *Katalog PTBRiO 2008. Rynek badań. Badacze. Firmy Badawcze. XIII edycja. [PTBRiO 2008 catalogue. Research market. Researchers. Research companies. 8th edition]* (http://www.ptbrio.pl/files/Katalog_PTBRiO_2008_na_strone.pdf of 11 October 2009). Apart from the data provided in the annual reports of the Polskie Towarzystwo Badaczy Rynku i Opinii (PTBRiO), which concerned the size of the market of Internet research (and the research of the Internet itself), this significance was confirmed in the recently submitted chapter on the CAWI research in the *Polskie Standardy Jakości Realizacji Badań Rynku i Opinii Społecznej w Terenie [Polish quality standards of implementing market and social opinion research in the field]* published by the Polish Association of Public Opinion and Marketing Research Firms, and the first PKJPA certificates (certificates concerning the programme for controlling the quality of interviewers work) obtained in 2008 by research institutes within the scope of CAWI research (last year the PKJPA certificates within the scope of CAWI research were granted to 8 research companies – see http://www.ofbor.pl/public/File/Wyniki_VII_edycji_audytu_PKJPA.pdf).

[10] See J. Sielicki, *Badania w sieci, poprzez sieć i dla projektów internetowych [Research in the net, via the net and for the Internet projects]*, PTBRiO training materials (21-22 April 2009 Warsaw).

It is worthwhile to emphasise, at the very beginning, that CAWI, as all other techniques of conduction surveys, has its strong and weak points. Although this paper discusses the possibility to use this technique to evaluation it should be stressed that it is impossible to recommend its use under all types of this kind research. As regards evaluation research within the frameworks of which the criterion of relevant targeting of the support programme is very significant, it needs to be emphasised that the selection of the technique under the projected evaluation research should also be relevant, i.e. is should result from a previously performed proper identification of the researched area and the respondents' profile. CAWI is worth using for this evaluation, which make it possible to use the strong points of the technique and under which the impact of the weaker points can be minimized or made completely insignificant. The following remain the significant factors during the selection process: types of evaluation, types of programmes covered by the evaluation, respondents' profile, but also a number of other variables. This paper identifies the areas of evaluation, under which it is really worth to choose CAWI taking into account the favourable relationship of the obtained effects and incurred expenditure. Observance of only few significant recommendations concerning the use of CAWI under evaluation research, may benefit in a high quality of the implemented research, which will be achieved by means of low project costs not burdening the budget.

Internet research in evaluation

At the moment, two major types of applied Internet techniques (CAWI) may be distinguished in evaluation research on the Polish market. The first of them – the most common one – is the CAWI researches that uses databases (of the Contracting Party or external databases), second are the pop-up surveys displayed in the Internet. The Internet technique of panel research[11], as yet, has a marginal significance for evaluation research.

Starting from the end, it may be relatively difficult to imagine how is it possible to run an active panel; especially, a one which would consist of selected beneficiaries of some programme, that would express their opinions, in a form established in advance (e.g. an on-line survey) concerning a defined theme from the area of the evaluated programme. Taking into account the fact that currently the majority of programmes, especially the EU programmes, is implemented under closed programming periods (a few years financing perspectives), for the beneficiary subsequent evaluation processes (especially evaluations of topic research) have a one-off character – participation in an interview, filling in a survey each time on different issue, despite the fact that, basically, they all concern the same project. The situation is different in the case of the on-going evaluations, under which in theory the beneficiaries may be asked to express their opinion on the same type of topics (e.g. processes, effects) at different stages of the project implementation (e.g. by means of CAWI surveys, Internet reporting forms or under more interactive forms – on-line focus groups / in-depth interview, panel discussions via different types of communicators, forums or specially designed thematic portals). Therefore, the beneficiaries may be treated as participants of a panel, which within the period of the project implementation and sustainability voluntarily agree by accessing the programme to participate in this type of research (devoting their time and sharing their knowledge on the project).

[11] See CAWI research classification included in the *Polskie Standardy Jakości Realizacji Badań Rynku i Opinii Społecznej w Terenie [Polish quality standards of implementing market and social opinion research in the field]* (www.ofbor.pl of 14 October 2009).

However, in reality this arrangement usually takes on a different form. Beneficiaries are informed that they are expected to take part in an interview or fill in a survey at the moment when they receive a letter of notification (via electronic means or traditional mail) concerning a specific evaluation research. The process of recruiting persons to such a "panel" adopts a completely different form than it is the case for ordinary market researches. By signing the project financing agreement the beneficiaries become the potential respondents (informers within the scope of their projects, a wider programme or elements of the entire system of public support). However, on many occasions this recruitment is not entirely intentional on the part of the respondent – it is one of many consequences of taking part in a public programme.

The use of Internet panel research is very popular in the case when the evaluation research foresees a research method in the form of a panel of experts, especially, implemented by means of the specific Delphi method. To put it in very simple terms the "Delphi method is based on the structured process of collecting knowledge from a group of experts by means of a series of questionnaires combined with a controlled collection of feedback opinions. The questionnaires are presented in the form of anonymous consultation procedure, which is many times repeated during the research (via mail or e-mail). The technique (...) is built upon (…) a series of questionnaires sent to a selected group of experts. Questionnaires are designed so as to obtain individual answers to the questioned themes, and then allow the experts to redefine their views along with the development of the work of the group.[12]" If we assume that the research is participated by experts from different geographical and language regions (e.g. transnational research of horizontal issues, sector research or during evaluation of cross-border cooperation programmes) the CAWI technique works perfectly. It is able to ensure a relatively short field time of research implementation and significantly decreases its costs (as opposed to the situation when it would be necessary to organise a meeting for all members of the panel in a single place and time). On many occasions the Internet form of the research is the only possibility for foreign experts from narrow specialties to participate in the panel. The strong point of Internet research, i.e. "geographical independence", which is often highlighted in different studies concerning such research, creates great opportunities also in the case of evaluation research.

The second types of CAWI research, i.e. the popular pop-up surveys, implemented by means of pop-up windows, are among relatively popular techniques also under the evaluation research. Research in the pop-up form may be used, for example, in the evaluation of projects, which consist in construction of thematic portals (e.g. European Funds Portal - www.funduszeeuropejskie.gov.pl; E-business Support Portal – www.web.gov.pl; Innovation Portal – www.pi.gov.pl), evaluation of promotion systems of defined public programmes implemented, *inter alia*, through www pages (e.g. www.poig.gov.pl) or Internet forums addressed to persons/entities potentially interested in the programmes. Short evaluation surveys in the pop-up form can cover randomly selected persons that currently use the information placed on the website, as well as more narrow groups of users (visiting a specific division / tab; downloading a selected scope of published materials; expressing opinions). Analysis of even a few simple socio-demographic features of the portal users enables to assess the effectiveness of communicators and the quality of the published information for a defined target group.

The use of this type of CAWI research is very often linked to the fact that there are no databases of expected respondents or the use of the existing databases does not exhaust the information needs

[12] *Ewaluacja Narodowego Planu Rozwoju i programów operacyjnych w Polsce. Poradnik [Evaluation of the National Development Plan and the Operational Programmes. A manual]*, National Evaluation Unit, Department of Structural Policy Coordination, Ministry of Economy and Labour, Warsaw 2005.

within a given scope. However, the pop-up surveys (just like other Internet techniques) should constitute only one element of evaluation, an additional source of knowledge within a given field under the entire process. It is impossible to perform a reliable evaluation of the undertaken research problem, which is carried out in line with the "rules of the game", by referring to only one source of information (especially, a one which verification in respect to respondents' selection is as difficult as it is the case with the pop-up surveys). It is worth to remember that such an Internet technique is characterised by a relatively large number of features significantly jeopardizing the research performance. Just like the majority of Internet users do not like the "automatically" popping-up windows, which disturb the current focus or search for some important information (hence, some users switch off the pop-up function in their browsers), so all kinds of pop-up surveys are generally "disliked." Therefore, in the case of such research one must expect that there will be a great number of surveys which were not even started or which were dropped out during their completion. Thus in order to increase the response rate it is very important to use all possible techniques relating to encouraging the respondents to begin participation in the research as well as raising the motivation for the survey completion (highlighting the social usability of the evaluation results, role of the respondent, providing additional forms of expressing gratitude for participation, etc.).

Of course, the evaluation surveys appearing in the pop-up windows cannot be long. Moreover, as with other CAWI type research there applies a very strict discipline within the scope of the application functionality, visual attractiveness, efficiency with different types of browsers, operational systems or worse transmission. Certainly, the authority of the researcher or the sponsor of the research is of great significance as regards encouraging respondents to participate in it (a recognised institution, whose logo is visible on the website of the research, sponsor of the programme, a respected person or a research institution). Nonetheless, taking into account the size of the programmes, funds or projects, which are currently implemented in Poland (to a large extent thanks to the European Union) it would be difficult to find a person or an organisation that have not met with surveys (including Internet surveys) which were related to participation in a co-financed training or different types of support schemes – in respect to recruitment to the programme or evaluation after its completion. Therefore, it is more and more difficult to implement surveys in Poland regarding evaluation, this difficulty will particularly relate to the CAWI research implemented through the pop-up surveys. Although it may be considered if in the case of complex CAWI tools applied for the purposes of evaluation, reporting or registration, which have to be handled by the respondents (applicants, beneficiaries, potential addressees of the intervention), it would be worthwhile to additionally use the pop-up windows to verify the quality of the prepared data collection tools in the opinion of the target group.

The third and most important from the perspective of evaluation Internet research technique – CAWI, that uses respondents' databases – has been used for some years now, with varying degrees of success, by the external performers of the evaluation (research institutions), as well as by evaluation units established under public structures (governmental, self-governmental, non-governmental) competent for individual programmes. Nonetheless, it should be emphasised that its popularity in evaluation progressively increases (as in the entire sector of public opinion and market research). On the one hand, it results from the availability of CAWI in the offer of a growing number of research agencies because of low costs and short time of the data collection process, and on the other hand, because in many cases it is the most efficient form of carrying out surveys among applicants and programme beneficiaries – both natural persons, as well as institutional entities.

The possibility to apply the CAWI technique (implemented with the use of databases) in evaluation, to the greatest extent, determines the profiles of beneficiaries (or other respondents) that have the

knowledge within the scope of the evaluated programme and can share it. The programmes supporting the agriculture and fisheries sector or different types of programmes counteracting social marginalization (exclusion) and, for similar reasons, many other programmes will, certainly, not constitute an appropriate research material that allows using any efficient Internet techniques. Of course, some specific exceptions can be indicated, but Internet penetration among our observation units remains the key issue. Maybe, under some circumstances, mixed methods may turn out efficient as regards reaching respondents and gathering data from them (e.g. transferred via traditional mail, invitation for the respondent to fill in a survey depending on respondent's preferences – 1. directly on the website under the indicated WWW address or – 2. in a paper form including a request to sent it back via traditions mail to the institute). Although by using the so-called mixed-mode data collection it is necessary to take into account the fact that research carried out under this form may "provide" different types of additional errors, which should be minimized as much as it is only possible[13]. Some researches fail to accept such an approach at all, though, under many circumstances its application could have been functional (e.g. by contributing to limiting the number of data gaps in an evaluation survey or by increasing the response rate).

The database, which will be used as the sampling frame, remains another issue for consideration in the context of using this type of CAWI technique in evaluation; because, a respondent is not anonymous at the stage of collecting data. Individually addressed invitations (bearing such data as the name and first name or name of the entity, contact address, project number, etc.) are directed to the respondent but, at the same time, in the very letter of notification and in the survey the researched person is ensured on the future "anonymity" of the results and collective analyses without the possibility to indentify the informer. A substantial database is necessary to carry out the sampling procedures, sent invitations, control the research implementation or later analyses considering significant profiles (e.g. taking into account the size of the grant localisation or the type of the project). As for the public projects (especially, EU programmes, starting with the pre-accession ones) the databases most commonly cover a number of variables defining the profile of the applicant / beneficiary and the characteristic of his/her project(s). As a result of application generators, which operate form some time now, this data is already collected in an electronic manner; whereas, in the case of the CAWI research the e-mail address of the beneficiary remains a particularly significant variable. When there is the e-mail address the invitations with the link to the survey, the reminder letter, contact with the implementation department or thanks for taking part in the research can be sent via the electronic mail and, as a result, the costs of the research are automatically reduced by several times[14]. However, it would be worth to highlight that the e-mail addresses, especially those of private persons, become outdated more often than other contact data. Taking into account the fact that evaluation research concern the effects of support, they are most commonly implemented at least 6 months after completion of project implementation (participation in the project) and, in the case of large investment projects, 24 months after their completion or even later, and thereby, the e-mail addresses (provided at the stage of application – in the project financing application) can be outdated even in the case of several percents of beneficiaries. During the project implementation and within the period of its sustainability the beneficiaries (but, unfortunately, not other applicants that may also become

[13] See H. Domański, K. Lutyńska, A. Rostocki (ed.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, IFiS PAN, Warsaw 1999.

[14] Although in the case of the programmes from the 2007–2013 perspective that were targeted, *inter alia*, at entrepreneurs a proposal was considered during the consultation of the models of the project financing applications under some measures for the applicants to give the variable "e-mail address" only optionally (which would make it impossible to reach many beneficiaries via e-mail).

a valuable source of "evaluation knowledge") are obliged to inform the implementing authority for the programme on all changes in the identification data. However, in the case of the postal address, names of organisations or data of the legal representatives of a given entity such changes are recorded, whereas the change of the e-mail addresses in not always registered (the more in the electronic database). The availability of "some" e-mail addresses of respondents can, therefore, constitute a significant problem under the evaluation research carried out on the basis of the CAWI technique. This can, above all, relate to the evaluation programmes directed at natural persons and small enterprises, as well as organisations; however, it occurs less often in relation to programmes for which the direct beneficiaries cover institutions that have a longer history and scope of activity (which in their case, in practice, also signifies a relative stability of electronic addresses).

Moreover, the databases (both secondary, as well as those "produced" during the Internet research implementation) are also closely associated with strictly formalized registration procedures and protection of their content (including guarantees of safe storage on the servers and safe processing), especially, if personal data or other legally protected data is concerned. However, these procedures are not very different from those applied in the case of other research techniques. It is also true, in the case of the CAWI research, that the encrypted connection proves correct as a tool securing the data transfer and, simultaneously, as a specific guarantee for respondents. The Secure Socket Layer Certificate, issued for the domain on which the research is implemented, guarantees that the information flow (data transfer) occurs only and exclusively between two points – the respondent's computer and the server including the database. Despite the fact that all over Poland Internet users are commonly concerned as regards the issue of giving any important (personal, corporate, official, etc.) information on the Internet, respondents of the CAWI research (especially, if they own a bank account with on-line access, or if they use internet shops or participate in on-line auctions) should feel definitely more secure as a result of announcements on secure connections, website address starting with "https" and a "lock" visible on the screen. If the CAWI research foresees collection of "sensitive" data from the perspective of the respondents the encrypted connection with the research page should, certainly, decrease the percentage of item-non response.

It is worthwhile to consider what type of evaluation research may reach the programme beneficiaries via the network if minimum requirements are adopted in respect to them (active Internet users, which made their e-mail address available in the contact details for the purposes of the programme). It seems that relevant application of the CAWI technique with the usage of the programme databases may be identified on a very extensive area of different types of evaluation research – related with collecting information within the scope of undertaking the research problem (including assessments, opinions or other type of data of quantitative nature) starting with the beneficiaries, applicants and other entities engaged or interested[15] in the programme. These would, above all, cover different types of mid-term thematic evaluations concerning relevance and utility of the projected support offer aimed at the target groups or the evaluation of functionality of the system of awarding aid at different stages of this process. CAWI may be also successfully applied in the evaluation research implemented under ad-hoc manner in response to urgent information needs of public services within the scope of a given programme, i.e. following from an observable procedural problem, sudden change in the socio-economic situation or a controversial political decision which is planned to be implemented. Under such circumstances rapid and cheap Internet research works perfectly. Experienced evaluation units are able to design and perform (e.g. among the programme participants) such a research on their own leaving aside time consuming

[15] In the case of some evaluation research the newsletter databases will have a great significance, which included data of the persons potentially interested in the programme offer (including, *inter alia*, their e-mail addresses).

public procurement procedures. Given the possibilities created by the semi-professional software (very often available under the open source form) for on-line surveys or more advanced commercial platforms of Internet research it comes as no surprise that such solutions start to be more and more willingly used to conduct research (of evaluation nature) by public bodies (self-governments, governmental agendas).

The CAWI research also proves very successful in the case of on-going and *ex-post* evaluations, which are targeted at effects measurement and assessment. The current measurement and assessment of the achieved result or impact indicators within the frameworks of implemented support instruments constitutes an invaluable tool both to watch (monitor) progress, as well as to verify (evaluate) their effectiveness and efficiency within the planned scope. As for the effects of the examination, in the majority of cases, the horizon of measuring the evaluation indicator (result or direct impact) should be coupled together with the moment of completing the project by a beneficiary and closure of a relevant reporting period for the indicator. The techniques of computer or Internet assisted researches give the possibility to select at an appropriate moment a respondent, which meets the assumed conditions of taking part in the research (selection criteria) and invite the respondent via e-mail to fill in the survey (e.g. 12 months after completing participation in the programme). The aggregated results gathered in this manner constitute data of far higher quality than the results obtained under a one-off ex-post research, which at one point in time addresses all beneficiaries regardless of the time that elapsed since their participation in the project (completion of an undertaking). The coordination of traditional field research implemented with the participation of an interviewer (e.g. PAPI, CAPI techniques) which, at the same time, controls respective dates concerning the project and optimum moments of an indicator measurement, would exceed the costs of the possible application of the CAWI technique in the case by several times. When the on-going evaluations are not implemented at all during the programme implementation or if they are implemented in an improper manner, the ex-post evaluation will inevitably attempt to catch up on the shortcoming, which unfortunately has a negative impact on the costs and quality of the collected empirical material. At that stage, problems with obtaining a relevant level of returns is also significantly strengthened. Beneficiaries, which were subject to control and monitoring on several occasions or who took part in different types of mid-term evaluations of the programme are less and less willing to cooperate with the evaluator during the ex-post evaluation.

However, in general, the ex-post evaluations may also apply the CAWI technique with great success, but in their case it must be remembered that the respondents are not that motivated, there are more outdated e-mail addresses (unavailability of valuable informers) and memory fails everyone when some time elapses from the event of interest to research. Moreover, despite the fact that the CAWI technique gives the possibility to conduct a comprehensive research (on the entire population) with slightly higher costs and longer implementation time, it is rather worthwhile to apply representative research on samples since, it is rather a pity to "exhaust" a sampling frame that may have been of great value in future toward researches of ex-ante nature[16]. The experience of the current beneficiaries can often be very valuable in the process of projecting a new generation of support programmes addressed to a similar target group.

[16] See evaluation entitled *Plany i potrzeby przedsiębiorców sektora MSP w zakresie ochrony własności przemysłowej, w kontekście uruchomienia działania 5.4 PO IG Zarządzanie własnością intelektualną* [*Plans and needs of entrepreneurs from the SME sector within the scope of industrial property protection in the context of launching Measure 5.4 of the OP IE Intellectual property management*], IBC GROUP Central Europe Holding Ltd. for PARP, Warsaw 2008 and evaluation entitled *Plany i potrzeby przedsiębiorców sektora MSP w zakresie rozwoju eksportu i powiązań z zagranicznymi rynkami i partnerami handlowymi, w kontekście uruchomienia działania 6.1 PO IG Paszport do eksportu [Plans and needs of entrepreneurs from the SME sector within the scope of developing export and relationships with foreign markets and commercial partners in the context of launching Measure 6.1 of the OP IE Passport to export]*, Policy & Action Group Uniconsult Sp. z o.o. for PARP, Warsaw 2009.

At the stage of the ex-post evaluation attempts to reach the final recipients of aid (final beneficiaries) are also very common. In the case of many support instruments the recipients of direct grants are institutional beneficiaries (e.g. business support organisations, self-governments, training companies, public benefit organisations, etc.) but the final recipient of aid transferred by these entities is somebody else. These can be both, recipients of the so-called *de minimis* aid (e.g. entrepreneurs – "customers "of scientific and technology parks or members of clusters), natural persons (e.g. participants of the co-financed trainings), and also intermediate users or recipients of the generated goods and services (e.g. self-governmental community that uses the created socio-economic infrastructure). These are usually very large populations, which are ideal for testing by quantitative techniques in line with the relevant information needs. Also, in this case the possibilities of applying the surveys in the form of CAWI research are rather extensive. Everything depends on the profile of respondents and the availability of a relevant database for the purpose of sending e-mails. Unfortunately, at the level of final beneficiaries the evaluators very often encounter the problem of unavailability of data in an electronic form. Shortcomings at this stage, which concern electronic resources of information on the programme, sometimes follow from the negative experience of institutions as regards central IT systems, "uncontrolled decentralisation" of database systems or fear of processing personal data. In many instances, it eliminates the possibility of applying the CAWI research to evaluation or entirely excludes the possibility of covering target groups (final recipients of the support) with research because of the costs.

The specific relationship: researcher – respondent – beneficiary

The already relatively large number of theoretical works within the scope of using Internet research – of both domestic and foreign authors – in many places emphasize that CAWI (using databases of respondents) is a very efficient research tool every time when the target group has a closed character and is clearly defined. Moreover, CAWI works especially well when the members of the group are aware that they belong to this group (e.g. beneficiaries are repeatedly informed that they participate in a programme co-financed by the European Union). The use of the Internet research techniques under the evaluation of programmes is certainly justified by, *inter alia*, these characteristics of the beneficiaries or the persons unsuccessfully applying to participate in the programme, as well as the previously mentioned features.

The experience gained during the face to face research confirms that the contact between the interviewer and the beneficiary, which relates to the beneficiary's participation in the programme, is not a surprising aspect for the latter. Beneficiaries voluntarily declared their willingness to participate in this type of research (i.e. giving interviews, filling in evaluation surveys, etc.) many months before the evaluation meeting; hence, their approach to the carried out evaluation research is entirely different than the approach of other respondents of social research (not beneficiaries). It may be stated that the beneficiaries expect that there will come a time when someone (e.g. an interviewer representing a research institute or a representative of an institution responsible for the programme) will ask them on the effects of the programme. Despite the fact that as for CAWI, the person of an interviewer does not exist, similarly, the invitations to take part in a research related with the programme evaluation are treated by the respondents from an entirely different perspective in respect to motivation than it is the case in the invitations concerning the participation in other social or market research. It is worthwhile to note that the provisions of project financing agreements concluded between the beneficiaries and the implementing authority almost always refer also to providing information for the purposes of the programme evaluation. The general terms of the agreements used, *inter alia*, by the Sectoral Operational

Programme – Improvement of the Competitiveness of Enterprises, 2004-2006, use the following wording "The Final Recipient (beneficiary) undertakes to (...) provide information on the Project effects for a period of 5 years from the date of the Project completion to the Managing Authority, the Final Beneficiary (Implementing Authority) or entities designated by the above."[17] This provision undoubtedly facilitates data collection and implementation of surveys, even on the basis of Internet techniques, if only it is an adequate communication channel for a given beneficiary. Regardless of the provisions of the agreement (that are, certainly, not precisely remembered by a large part of the programme participants), subconsciously, there is an experienced commitment towards the donator of the financial resources, the beneficiaries are, in general, more willing to take part in the evaluation research, than they are to participate in different types of research[18] as the "statistical Poles." Therefore, in the case of evaluation research it is possible to refer to the existence of a specific relationship between the researcher and the respondent (beneficiary), which favours increase in the indicators of the sample implementation (regardless of the used technique). At the same time, in the case of significant restrictions of the Internet methods within the scope of the level of the response rate the good relationship with the respondent that exists in evaluation research, certainly, increases the potential efficiency of the CAWI technique and justifies its application, exactly, in the case of research implemented for the needs of evaluation. Moreover, using a number of functional and methodological solutions available under the CAWI technique that are described in literature, there is made possible to achieve relatively highly satisfying implementation results in evaluation research for the self-completing survey. These solutions relate to the stage of arranging CAWI "interview" in letter of notification (e-mail and traditional ones), work of the respondent with a tool (e.g. quality of the questionnaire and functionality of the CAWI application; access of contact persons of the evaluation unit, FAQ tab, reference aid, etc.), the formula of reminding (e.g. mail or telephone reminders), as well as gratifications for participating in the research.

In the case of evaluation research implemented on-line, the problem of dropout surveys is also smaller. To a large extent the motivation which led the respondent (beneficiary) to participate in the research is so strong that it disciplines him/her to complete the survey to the very end (especially when the CAWI tool contains the login panel and the possibility of re-saving entered data, i.e. it enables to the respondent to fill in the survey in parts)[19]. Of course, there are some exceptions. Because, the researches implemented via the Internet are relatively easy to perform and they do not generate large costs it sometimes happens that some CAWI projects are launched in an incomplete form (e.g. incorrect selection of the sample[20];

[17] See http://www.parp.gov.pl/index/index/197 of 11 October 2009.

[18] This tendency will increase or decrease depending on many factors (e.g. number of evaluation research, in which the beneficiary participated, the amount of the obtained financing, time that elapsed from the moment of finishing the participation in the programme or from the project settlement).

[19] This solution works in evaluation research implemented with the application of the CAWI technique, which use the indicator forms together with the survey. Most commonly in order to fill in the survey the respondent should seek relevant data. The CAWI survey addressed, for example, to entrepreneurs (beneficiaries of the subsidy) does not have to be filled in by only one person in the company. It may often require the engagement of several persons (e.g. president of the management board, proxy for the project implementation, the chief accountant), which are the best informers within a given scope of the project.

[20] The databases to CAWI research obtained from the central information systems concerning EU programmes (SIMIK, KSI) in a relatively small extent allow to make an initial analysis at any stage of sample selection, thus it sometimes happens that this activity is mindlessly "transferred" to the filtering questions in the CAWI survey. Then it turns out that there is more specific cases that it was initially expected and the research was addressed to very differentiated groups, in which only some persons can refer to abstract, from their perspective (and sometimes even absurd), questions or tasks in the survey.

collection of data that was already given by the beneficiaries at the stage of application or reporting; lack of survey logic and errors in the transition rules, etc.). Despite, the sincere desire of the respondents, the level of implementing on-line surveys significantly decreases and the dropout rate increases.

Taking into account the decreasing impact of the discussed relationship between the researcher and the respondent, CAWI research will be less efficient when carried out on a group of beneficiaries, whose relationship (a commitment following from the received support or the provisions of agreement) with the representative of the programme's sponsor is not that strong. It is, especially related to the entities applying for financing, which finally were not awarded the resources (e.g. their projects did not receive a sufficient number of points at the selection stage) or final beneficiaries that use the benefits of the programme not entirely conscious of them. The former may be used as the sampling frame to construct a control sample for comparative evaluation methods because their contact data is easily available and possesses relevant socio-economic characteristics. The latter, on the one hand, form the basis for informers within the scope of effects at the stage of programme impact; on the other hand,, both the former and the latter group will not be as willing to fill in evaluation surveys (including on-line surveys) as direct grant recipients are.

However, despite this specific relationship existing under evaluation research between the researcher and the respondent, who commissions the research, the standards of the social research still apply. Highlighting to the beneficiaries that the participation in the research is obligatory or threatening that a possible refusal will be understood as non-compliance with the provisions of the agreement (which, in theory, implies return of the grant) will, certainly, lead to distortions in the results and completely undermines the purpose of their implementation. Only voluntary participation guarantees that the collected opinions are sincere and the assessments reliable. The provisions set out in the project financing agreement are, primarily, informative and it is difficult to image that they are executed in case of evaluation (the case is different, if the beneficiary refuses to be controlled or monitored). However, it would be worthwhile to note that often the beneficiaries mix up different types of researches concerning the programme that they used, failing to understand whether they are for reporting, control or evaluation purposes. This may be beneficial from the perspective of the level of the surveys implementation in evaluation researches, unfortunately, from the perspective of the quality of the collected material (independence of the opinion, reliability of issued assessments it may be unfavourable. The most important motivator for the beneficiaries (respondents) should be the sense of social usefulness of the research results, although there is no denying that also other factors have a great motivational significance.

In summary, the very commonly cited objection towards the CAWI technique that the level of implementing the assumed sample is low although the minimum methodological requirements are met (including, *inter alia*, the area of respondents selection criteria, construction of tools, organisation of the data collection process), to a large extent, should not refer to the evaluation research. Of course, along with the multiplication of different types of evaluation, which concern overlapping programmes, which on many occasions cover the same respondents, the response rate gradually decrease. However, most certainly in many evaluation researches a number of defined CAWI technique restrictions may be easily reduced to the minimum, at the same time, using its strong points.

Evaluation of the results of Measures 2.1 and 2.3 of the SOP ICE after 18 months (case study)[21]

The Polish Agency for Enterprise Development (PARP) presents an interesting example of applying the CAWI technique in the evaluation research[22]. Starting from 2004 onwards, within the frameworks of implemented evaluation projects, the Agency carried out research of entrepreneurs (beneficiaries of the public support) with the application of the CAWI technique. It is worth emphasizing that this technique works very well in the PARP research, primarily, because of the availability of Internet in this group of respondents, as well as the previously mentioned specific nature of the relationship between the supported entrepreneurs and the programme Implementing Authority.

According to the Central Statistical Office, Internet access in Poland is distributed in the following manner: over 91% of small enterprises (10-49 employed persons) have Internet access, 98.6% of medium-sized enterprises (50-249) and almost 100% of large enterprises (250 and more employees). In general, 93% of all entrepreneurs (excluding micro companies) have Internet access[23]. It thus creates enormous possibilities to implement research within the use of this medium- both sectoral research on small and medium-sized enterprises, as well as evaluation research covering the entrepreneurs taking part in the programmes.

The evaluation of the results of to subsidised measures addressed to entrepreneurs – Measure 2.1 and 2.3 of the SOP ICE, implemented by PARP demonstrates a good example of using the extensive possibilities of the CAWI technique in the evaluation research. The research was launched in 2007 and it will finally (until the middle of 2010) cover over 4,000 of supported small and medium-sized enterprises (SMEs). The selection of the research technique resulted, first of all, from the current good experiences of the Agency as regards implementation of the on-line research among entrepreneurs (among others beneficiaries of the Phare pre-accession programmes), but also from the broadly conceived objectives and functions (information needs) that were to be fulfilled by the research for individual recipients of the research results to which the CAWI technique fully responded. The project "Evaluation of the results of

---

[21] The following sub-chapter used fragments of the dissertation of the author entitled *Specyfika zarządzania w projektach ewaluacyjnych, na przykładzie Oceny Rezultatów wybranych działań Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw* [*The specific nature of management under evaluation projects on the example of the Evaluation of the results of some selected Measures of the Sectoral Operational Programme– Improvement of the Competitiveness of Enterprises]* written under the post-graduate study on Project Management of the Warsaw School of Economics.

[22] The Polish Agency for Enterprise Development is a governmental agency subordinate to the Minister competent for economy, which was set up by way of an act of 9 November 2000. The Agency is responsible for management of the funds originating in the State budget and the European Union, as well as implementation of economic development programmes supporting: entrepreneurs, innovative and research activity (especially of small and medium-sized enterprises), regional development, increase in export, human resources development or the use of new technologies. In the 2007-2013 financing perspective the Agency is responsible for the implementation of measures under three operational programmes: Innovative Economy (OP IE), Human Capital (OP HC) and Development of Eastern Poland (OP DEP). In the case of the 2004-2006 programming period PARP acted as the Implementing Authority, *inter alia*, for the measures implemented under the following programmes: Human Resources Development (SOP HRD) and Improvement of the Competitiveness of Enterprises (SOP ICE) (see http://bip.parp.gov.pl/index/index/505 of 11 October 2009).

[23] Central Statistical Office, *Wykorzystanie technologii informacyjno-komunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*, Warsaw 2008.

Measures 2.1 and 2.3 of the SOP ICE after 18 months" is an internal on-going evaluation[24], which assumes systematic examination of the beneficiaries of both measures[25] targeted, primarily, at the measurement of long-term effects of the awarded support.

In the further course of SOP ICE implementation, there appeared an "urgent need" to know the results of the projects implemented so far within the frameworks of the two key programme instruments. This need, on the one hand, followed from the challenges undertaken by the administrations of different levels associated with the programming of the detailed procedural solutions and guidelines for the new support programmes for entrepreneurs under the 2007-2013 perspective (OP IE, 16 Regional Operational Programmes). On the other hand, these arrangements could not have been made without the knowledge on the efficiency of similar activities undertaken so far. It is clear that reporting obligations imposed on the Implementing Authority forced it to carry out systematic research related with the measurement of the defined indicators for performing the objectives of a given measure of individual projects.

The facts that stood behind the adopted project assumptions were, sometimes, contradictory and required a compromise in respect to the undertaken decisions or measures. For example, in order to demonstrate some key result and impact indicators[26] - general, which are defined in the SOP ICE Programme Complement at the level of the entire measure it would be possible to carry out a field research on a representative group of beneficiaries within an established time horizon. In this case the project providers established the moment of measurement at 18 month from the moment of completing a given project (not the measure, which is worth emphasising). Whereas, at the level of the projects there was a total of tens of thousands of different indicators concerning several thousands of signed project financing agreements. They were measured with annual frequency in a perspective of few year period of the adopted project sustainability. What is more, the project indicators were not aggregated at the level of the measure to the form of general indicators (of higher order) – each project had its unique portfolio of indicators, which assumed the verification of material effects at the level of individual project, but did not allow to show the same effects in respect to overall of the financed undertakings[27].

In the case of general indicators (at the level of a measure) or "soft" assessment and beneficiaries' opinions regarding the effects of the undertaken measures the research of all projects was not necessary

[24] From the projects' perspective it was an ex-post evaluation carried out 18 months after their completion, while from the perspective of the measure, which implementation was not yet finalised (in 2007 and 2008) it was an on-going evaluation.

[25] The beneficiaries of the concerned measures of the covered enterprises from the SME sector, which implemented advisory (measure 2.1) or investment projects (measure 2.3) and obtained from the programme a refund of a part of incurred costs (i.e. purchase of advisory services – 2.1 or purchase of fixed assets – 2.3). In general terms, the projects were to strengthen the competitiveness of companies – beneficiaries. For measure 2.1 these were mainly projects foreseeing advisory within the scope of quality, innovation and new technologies, principles of running a company on the territory of the EU or obtaining external financing for development. Whereas, Measure 2.3. supported SMEs investments related with modernization (leading to a significant change in the product or the production process), implementation and commercialization of innovative technologies and products or application of the IT/ITC technologies in management.

[26] These were the so-called evaluation indicators, which go beyond the final report on project implementation and require to carry out research and not only to aggregate the data in place.

[27] For example, the project indicators on employment were formulated in various manners under individual projects, i.e. the creation of X number of jobs; increase in employment by X% as compared to the baseline year, keeping the current state of employment, increase in the number of employees, etc. If we know one of the indicators on employment from all projects it is still impossible to calculate (aggregate) the "new jobs – broken down into women and men – created in the supported enterprises." Thus, the knowledge on efficiency of the projects within this area did not allow confirming the efficiency as regards measure indicators.

(contrary to the detailed indicators from individual project financing agreements). At some point, it would disproportionally increase the costs of the research as compared to the small benefits in the form of the research accuracy. The use of representative methods would, certainly, be enough.

In the course of making arrangements between the parties, common objectives[28] of the initiated project were formulated.

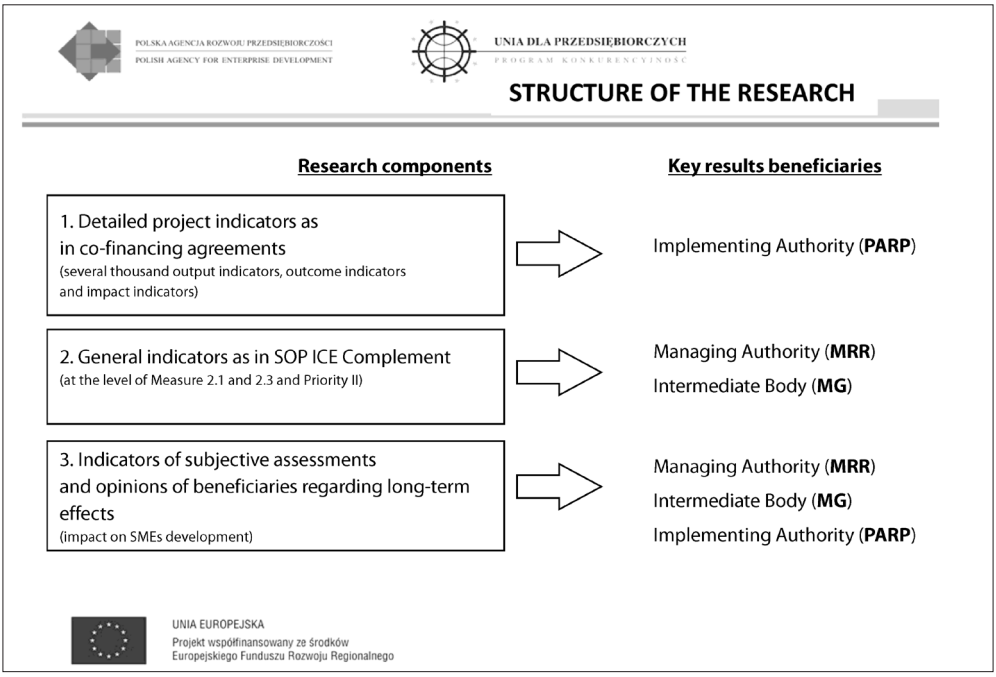


Fig. 1. Structure of the research

Source: A. Haber, J. Pokorski, *Ocena rezultatów SPO WKP. Systemowe podejście do mierzenia długofalowych efektów programu [Evaluation of the SOP ICE results. System approach to measure the long-term effects of the programme]* A lecture at the 3rd Evaluation Conference entitled *Rola Ewaluacji w kreowaniu polityki spójności [The role of evaluation in creating the cohesion policy]* (Warsaw, 25 October 2007) organized by the Ministry of Regional Development and the Polish Agency for Enterprise Development (por. www.konferencja-ewaluacja.pl).

The general concept foresees a research performed with the use of the CAWI technique targeted at all settled projects. The research was to be cyclical (periodical), implemented under quarterly rounds.

[28] The general objective of the project covered evaluation of the results of the projects implemented under measure 2.1 and 2.3 of the SOP ICE 18 months after completing their implementation. While detailed objectives corresponded to the above-described needs and were formulated in the following manner: 1) measurement of the effects of Measures 2.1 and 2.3 of the SOP ICE in respect to the indicator from the Programme Complement (jobs women/men, income, level of export, investment expenditure, certificates, innovations, strategic documents, etc.); 2) demonstrating the project effects within the scope of indicators foreseen in the project financing agreement (examining individual project indicators, which cannot be aggregated and were used for the purpose of the monitoring processes); 3) knowing the assessments and opinions of the beneficiaries on the long-term effects of the projects (above all, in respect to the criteria of efficiency and effectiveness of the support).

Beneficiaries that made up evaluated populations completed their projects in 16.5 to 19.5 months prior to the research[29].

In the case of technical resources, it was important that internal servers of the Agency were used to carry out the research and that the SSL certificate was applied to secure the data entered by the respondents. Success in respect to this issue, certainly, increased the confidence to the research among entrepreneurs whom, on many occasions, had to give data sensitive from the perspective of the company's secret (e.g. financial data). The reliability of the research was probably increased because of the use of the Agency's servers and, thereby, governmental address – parp.gov.pl.

It was assumed that the tool will consist of 3 major elements: the form of detailed indicators[30], the form of general indicators[31] and the evaluation survey.

The layout of the form of detailed indicators under the CAWI tool corresponded to the layout of the indicators in the applications generator for measures of the Priority II of the SOP ICE, in which the entrepreneurs declared the indicators of the project implementation and their values. It had the undeniable advantage that the beneficiaries were able to more freely move around a tool with a familiar interface. The beneficiaries in the detailed form could see their indicators, which they declared in the project financing application broken down into three subsequent years of the assumed project sustainability.

The general form displayed tables with indicators, which provided enough space to enter the values achieved in a given period (the value of the indicator had to be given in line with the suggested measurement unit and according to the indicated time period – as of the given year or moment of the research) and closed and open questions, for which it was necessary either to select or write in an answer. As for the narrow catalogue of indicators adopted in the general form the beneficiaries could use

[29] To increase the effectiveness of the implementation stage a compromise was reached regarding the horizon of the measurement. Instead of the one-off approach, 18 months after each project the invitations to take part in the research were sent each quarter to the beneficiaries who fell into the time frame of 16.5-19.5 months from the project settlement. The already examined research was excluded from the subsequent rounds of the research. Each round of the research in a given measure on average amounted to approx. 200 cases. The data collected in subsequent quarters were aggregated and analysed both as broken down into rounds and jointly. The research also used the independent significant variables, which are characteristic of a profile of beneficiaries or type of projects, in order to separate interesting analytical cross-sections.

[30] Each entrepreneur submitting a project financing application, which foreseen the achievement of a defined set of indicators concerning the project. The form of detailed indicators was to cover a list of monitoring indicators declared by the beneficiary for a given project and finally confirmed in the project financing agreement. The form was individually assigned to each of the projects implemented by the beneficiary. When the entrepreneur implemented more than one project – if a period of around a year elapsed from its settlement – the entrepreneur had to fill in separate forms of detailed indicators for each of the implemented projects. Moreover, if the entrepreneur has relatively recently implemented some other projects under any of the measures, in the case when a period of 16.5-19.5 months elapsed from their settlement, the entrepreneur was to obtain information including a request to fill in the form for the next project.

[31] The form of general indicators was to be used to collect statistical information on the effects of all projects (implemented under measures 2.1 and 2.3 of SOP ICE) within the area of enterprises activity. The beneficiaries were informed that on the basis of the SOP ICE Programme Complement, 2004-2006, PARP is obliged to collect information on overall effects of the awarded grants in the form of indicators. This was not possible to do because of the measurement of unique project indicators, which were impossible to be aggregated at the level of the measures. A uniform set of indicators was established for the purpose, which concerns the activity of the supported enterprises (beneficiaries) that was to make it possible to point to the overall effects of the grants (from the resources of the European Union and the Polish State resources) awarded under SOP ICE.

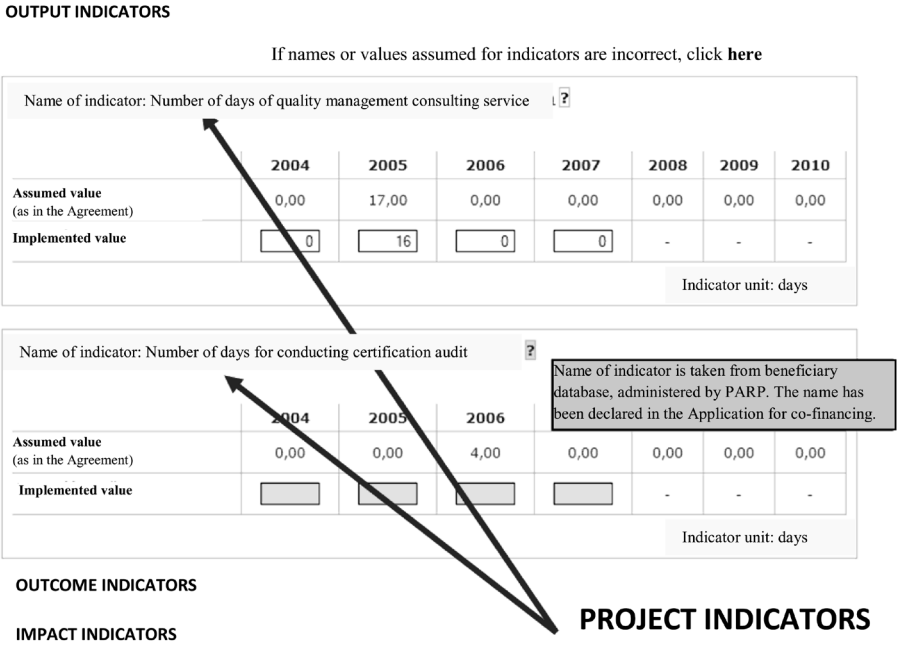


Fig. 2. Detailed form

Source: Evaluation of the results of Measures 2.1 and 2.3 of the SOP ICE after 18 months, PARP

a reference aid (hidden under the question mark next to the indicator – see Figure 3). It contained, *inter alia*, definitions of the indicators, the manner of measuring their value and other information useful to properly fill in the form.

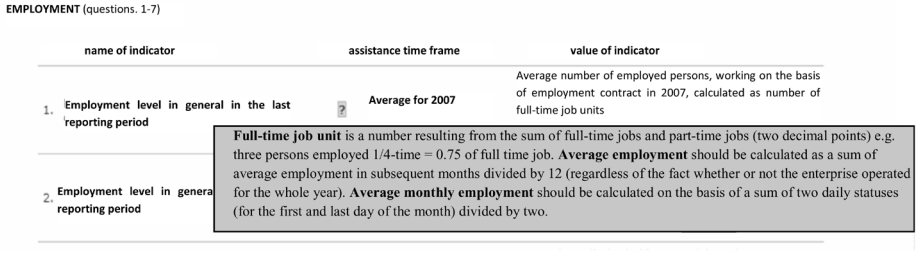


Fig. 3. General form – reference aid

Source: Evaluation of the results of Measures 2.1 and 2.3 of the SOP ICE after 18 months, PARP

The survey was to evaluate the effects of measure and it included questions concerning the impact of the grants on the activity of the supported companies (in the opinion of their owners or representatives of the management), as well as the particular questions. These were traditional survey questions: open (respondent was to write down a brief answer in the balk spaces) and closed (respondent was to indicate a relevant answer or several relevant answers in the designated box or to identify answers on the scale or in a table). Moreover, the projected on-line survey was to be dynamic at time of its completion – its

shape was dependant on the answer given by the respondent. At the beginning, only the first question was visible, and when it was completed subsequent ones appeared. This solution ensured many benefits – e.g. it prevented from overlooking some questions by the respondents, the survey was transparent, and the progress bar informed the respondent on the advancement of the process of survey completion.

Moreover, the tool had to have a number of additional elements that ensured functionality and quality. These covered among others, tabs with the information on the research, full manual, the so-called FAQ tab that were periodically supplemented by the project team, contact to the researcher (employees of the PARP evaluation unit). The standard features covered functions associated with the operation of the tool in the network: login panel, navigation and save keys, form of reporting errors joined with the e-mail address of the employee of the evaluation unit, print preview windows and many others. The research website could be entered, on the basis of the transferred ID and a password, by different persons in a company (e.g. the financial data was filled in by an accountant, detailed indicators by the project coordinator, survey by the owner of the company). It was to additionally facilitate the work with the tool to the entrepreneurs, at the same time, increasing the quality of the collected data.

In order to implement the objectives of the research it was necessary to develop a very complex and advanced tool as compared to the standard CAWI tools used by the majority of companies researching the market and opinions. For example, there appeared nested forms, encrypted connections based on the SSL certificate issued for the domain of the Agency, the format of saving information in detailed forms integrated with the databases system used by PARP.

Although the survey was relatively difficult and it required from the beneficiaries a lot of effort, the number of mail or phone questions was exceptional as compared with other research of this type implemented on a similar group. But, for example, it turned out that the databases of the implemented projects and beneficiaries contain a lot of errors and, what is more, they became very outdated since the moment of applying for the support. On the one hand, the data introduced under an annex to the agreement was not updated in the databases (among others information on the expected indicators), on the other, even apart from it within 18 months after settling the projects a great number of companies changed their name, official seat, contact data (e-mail, phone) or even transformed the partnerships (e.g. mergers or takeovers often caused that the NIP (tax identification number) of the beneficiary changed – basic change, a number constituting the identifier of the respondent under the research). All the above resulted in additional work under the project (it required to update the database and improve the respondent's profile on the website of the research). Moreover, the above circumstances caused that on many occasions the covering letters sent via mail reached a wrong addressee or the reminding letters were unsuccessful (reminding mailing). Furthermore, individual communication channels were used in contacts with the respondents (scans of covering letters sent to the new e-mail addresses or direct phone contact). However, the lessons learned during the first cycle of the CAWI research made it possible to work out a model solutions for individual problems, control the course of the subsequent cycles of the research and develop the mechanism for automatics of the process, repeatability of defined activities, which ensured high quality and stable progress of the implementation stage.

It is difficult to point to all the factors that contributed to the success of the project. They concerned both the projecting area, as well as the implementation and management areas. The recipients of research appreciated, above all, the reliability of the conducted research and interesting results thereof, based on hard data provided by the indicators. The project, most certainly, constitutes an added-value

of the activity of the Agency and it made a significant contribution to the development of the on-going evaluation in Poland, as well as the application of the CAWI technique in evaluation research. During the implementation of subsequent cycle of the "Evaluation of the results", the Agency implemented other evaluation projects that used the CAWI technique. A number of experiences gained during quarterly researches of the Measure 2.1 and 2.3 of the SOP ICE were used when introducing improvements to the new projects[32].

Moreover, it was planned that the experience gained during the project implementation will be further developed in the on-going evaluations dedicated to new measures implemented by the Agency (among others within the frameworks of the following Operational Programmes: Innovative Economy, Human Capital or Development of Eastern Poland for 2007-2013). In agreement with the Steering Group for the Evaluation of the IE OP it was also planned that in 2009 the on-going "Evaluation of the results" will be completed with research of an appropriately selected control group in order to carry out the analysis of the net effects (estimate the scale of changes in the supported enterprises, caused only by the intervention after separating the impact of external factors, such as e.g. changes in the economic trends, effects of other grants, etc.).

Of course, it is probably also possible to separate a number of shortcomings or weaker points under the project, which were impossible to overcame (e.g. better coordination in the implementation stage between the two measures, relieving the burden on respondents, depending on the type of project, increasing the automatics of the implementation stage, and thus relieving the research team, perhaps also better indicators and measurement tools or optimization of the indicators measurement horizon). These are major challenges under the new wave of the on-going evaluation, which will be implemented by PARP in the future, to a large extent, on the basis of the CAWI technique.

Conclusions

The broadly conceived development of the Internet, popularisation of the access to the www network and the increased social, economic and cultural activity, both of private persons and different types of organisations, institutions or their networks, which takes place now "right before our eyes" cannot remain without impact on the research industry (including evaluation research). As the number of Internet users continually and dynamically increases, the population of potential respondents of research implemented through the network also increases. This research touches upon more and more extensive range of socially significant problems, phenomena and events. Together with the development of the Internet new directions of development are identified and new applications of the CAWI technique. Evaluation research can most certainly be a perfect material for the purposes, at the same time, brining benefits for the improvement of the technique, but first of all, for the quality of the results of evaluation research provided on the basis of this technique. Already today, it is difficult to identify the types of programmes that are subject to evaluation, the types of respondents (beneficiaries or other potential informers engaged in the

[32] For example, under the research on the effects of measure 1.2 of the SOP ICE (which concerned, *inter alia*, recapitalisation of micro-loan funds and loan guarantee funds) in order to make the work of the researches more flexible in the stage of implementation a special administrative panel was established (a solution defined as the Content Management System), which makes it possible, without the need to use the IT language, to generate statistics on implementation, sent reminders via e-mails, data export, update of the FAQ tab, edit the questions in the survey and other information visible for the respondents on the research website, etc.).

programme) or areas of intervention, under which it would be impossible to apply the Internet research techniques. Without a doubt, as was shown in many years of experience gained by, *inter alia*, the Polish Agency for Enterprise Development, CAWI is willingly used to carry out research implemented among entrepreneurs (primarily, among small, medium-sized and large companies and the micro-enterprises based on innovative technologies or operating within the sector of e-services). Various theoretical works, which aim at grounding the new technique and practice of commercial research, point to the fact that CAWI is more and more extensively used. As regards the need to implement the CAWI techniques to a wider (and better) scope in the evaluation research the statement of Taylor and Terhanian uttered in the context of the on-line research may be repeated: "it is a speeding train that cannot be stopped and accelerates even more. Those who do not get on it risk that they will remain far behind[33]".

Jacek Pokorski is a graduate of the Institute of Political Science of the Cardinal Stefan Wyszyński University in Warsaw and the Chair of Project Management of the Warsaw School of Economics. He is related with the Evaluation Section of the Polish Agency for Enterprise Development for more than five years now. He is a specialist in evaluation of the competitiveness support and innovative economy programmes, especially, those targeted at the SME sector and the business environment. Pokorski is the manager and the co-author of many research projects (including evaluation research implemented with the application of the CAWI technique). He has experience in the evaluation on programmes financed from the pre-accession and Structural Funds of the European Union (especially Phare Economic and Social Cohesion for 2000-2003, Improvement of the Competitiveness of Enterprises Programme for 2004-2006, the Innovative Economy Programme for 2007-2013). He is a member of the Steering Group for the Evaluation of the Innovative Economy Operational Programme and the Working Group for evaluation by the OP IE Monitoring Committee. Pokorski is currently engaged in the preparations of a complex effects measurement system for the 13 measures of the OP IE implemented by PARP. He is the author and co-editor of studies and publications devoted to evaluation and the effects of the projects co-financed from the EU resources. For several years, he has been teaching at postgraduate studies and conducts trainings on evaluation for representatives of public institutions; he also organizes conferences devoted to the issue.

Sources

Babbie E., *Badania społeczne w praktyce*, PWN, Warsaw 2003, p. 299.

Batorski D., *Korzystanie z technologii informacyjno-komunikacyjnych*, in: Czapiński J., Panek T., *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, pp. 279–307.

Batorski D., Olcoń M., *Prowadzenie badań przez Internet: Podstawowe zagadnienia metodologiczne*. Studia Socjologiczne, 3(182), 99–132.

ESOMAR, *Global Market Research 2009. Esomar Industry Report*, Amsterdam 2009.

Ewaluacja Narodowego Planu Rozwoju i programów operacyjnych w Polsce. Poradnik, National Evaluation Unit, Department of Structural Policy Coordination, Ministry of Economy and Labour, Warsaw 2005.

[33] H. Taylor, G. Terhanian, *Heady days are here again: Online rolling is rapidly doming of Age*, Public Perspective, no 10, 1999.

Domański H., Lutyńska K., Rostocki A (ed.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, IFiS PAN, Warsaw 1999.

Central Statistical Office, *Wykorzystanie technologii informacyjno-komunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.* , Warsaw 2008.

Haber A., Pokorski J., *Ocena rezultatów SPO WKP. Systemowe podejście do mierzenia długofalowych efektów programu.* A lecture at the 3rd Evaluation Conference entitled *Rola Ewaluacji w kreowaniu polityki spójności* (Warsaw, 25 October 2007) organized by the Ministry of Regional Development and the Polish Agency for Enterprise Development.

Hewson, C., Yule, P., Laureat, D., Kogel, C. (2003). *Internet Research Methods. A practical guide for social and behavioral sciences*, London: Sage Publications

Mattelart A., *Społeczeństwo informacji. Wprowadzenie*, Universitats, Kraków 2004.

Pokorski J., *Specyfika zarządzania w projektach ewaluacyjnych, na przykładzie Oceny Rezultatów wybranych działań Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw* (praca dyplomowa), Warsaw School of Economics, Warsaw 2009.

Pokorski J., *System ewaluacji wykorzystania funduszy strukturalnych w Polsce. Ramy instytucjonalne i prawne* (praca dyplomowa), Cardinal Stefan Wyszyński University in Warsaw, Warsaw 2008

Polskie Towarzystwo Badaczy Rynku i Opinii, *Katalog PTBRiO 2008. Rynek badań. Badacze. Firmy Badawcze. XIII edycja. [PTBRiO 2008 catalogue. Research market. Researchers. Research companies. 8th edition]*

GfK report, *Net Index* of March 2009 (http://www.gfk.pl/podstrona.php?page=/page.php?id=235).

Gemius S.A. report, *Polski Internet 2008/2009 [Polish Internet 2008/2009]* of 2008 – (http://pliki.gemius.pl/Raporty/2009/02_2009_Polski_Internet_2008_2009.pdf).

Gemius S.A. report for Google Poland, *Google Poland Finance Study 2008* of September 2008 (http://www.Internetstats.pl/index.php/2009/01/internauci-o-uslugach-finansowych-google/#more-469).

Millward Brown SMG/KRC report for the pracuj.pl portal, *Oferty pracy i poszukiwanie pracy w Internecie,* December 2008, (http://www.Internetstats.pl/index.php/2009/03/oferty-pracy-i-poszukiwanie-pracy-w-Internecie-pracuj.pl).

Millward Brown SMG/KRC report for ING, *Internetowy Barometr ING*, March 2009 (http://media.ingbank.pl/PressOffice/get/PressKit.PressKit.926.po?field=125125).

Report entitled *Badanie zapotrzebowania na działania wspierające rozwój usług świadczonych elektronicznie (e-usług) przez przedsiębiorstwa mikro i małe*, Pentor RI for PARP, Warsaw 2009.

Rossi P., Freeman H., Lipsey M., *Evaluation - a systematic approach*, 6th edition, Sage Publication Inc., Thousand Oaks 1999.

Sielicki J., *Badania w sieci, poprzez sieć i dla projektów internetowych*, PTBRiO training materials (21–22 April 2009 Warsaw).

Taylor H., Terhanian G., *Heady days are here again: Online rolling is rapidly doming of Age*, Public Perspective, no 10, 1999.

The GUIDE – *Evaluation of socio-economic development*, European Commission 2004, (www.evalsed.info)

Zając J. Materials for a training entitled *Wykorzystanie Internetu w badaniach społecznych*, University of Warsaw 2007.

Zając, J. M., Batorski, D, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*. Psychologia Społeczna. Social Psychology, no 3 (2007), 234–247.

Jan M. Zając, Dominik Batorski

Methods of increasing the responsiveness in online surveys

Introduction

In recent years, research, including evaluation studies, carried out by online surveys (CAWI – Computer-Assisted Web Interview) has been growing more and more popular. This method has enormous advantages, such as the possibility to carry out research fast and on large samples at a low cost or the possibility of reaching very specific populations that are often hardly accessible otherwise. At the same time, however, it is not free from limitations, many of which involve low response rates which reduce the generalisation potential of obtained results.

Some research types are better fit for online studies than the other. As presented in this chapter, the specific nature of evaluation research makes it in numerous situations more suitable to carry out an online survey than using any other technique. At the same time, such research may prove fully valuable – not worse, and often even better than those carried out by traditional methods, insofar as relevant methodological standard requirements[1] are met.

One of the most important problems concerning the implementation of questionnaire-based research via the Internet is the assurance of a sufficiently high response rate which is the percentage of completed questionnaires in relation to sample size[2]. If the response rate is low and the motivation to participate in a survey is substantially connected with the survey subject, then the survey results might be biased or even worthless. This results from the occurrence of non-response error, i.e. the effect of the fact that not all individuals drawn in a sample can or want to participate in a survey. Answers of persons who did not participate in the survey can vary significantly from the answers of persons that did participate. In such a case, the generalisation of results from a population sample becomes illegitimate[3].

It does not mean, however, that the application of online surveys is always connected with lower response rate. Literature and reports provide examples of CAWI surveys, including evaluation ones, with response rates as high as to 80-90%, i.e. higher than in standard non-Internet surveys. The method in

---

[1] D. Batorski, M. Olcoń, *Prowadzenie badań przez internet: Podstawowe zagadnienia metodologiczne*, „Studia Socjologiczne", 3(182), 2006, p. 99–132; Zając, J. M., Batorski, D., *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna", No. 3, 2007, p. 234–247.

[2] It is worth noting that an alternative term for „response rate" is frequently used in the literature and it takes into consideration all the commenced surveys, irrespective of the fact whether they have been filled in completely or at least to the extent that allows to include the results thereof. In Internet surveys, due to quite often phenomenon of dropout during the survey, there can be a significant difference between the percentage of respondents commencing and completing a survey.

[3] Obviously under the condition that the surveys are carried out using random samples since only in such a case it is possible to apply the method of statistical inference. In practice many surveys carried out on the Internet are conducted using nonrandom samples.

which particular survey is carried out may considerably increase or reduce the willingness to participate therein.

Herein we included specific hints for the purpose of increasing the response rate in CAWI polls at various stages of the research process. The presented recommendations concern the surveys from questionnaires published on websites, irrespective of the invitation method. The problem of increasing the percentage of respondents commencing the survey is presented below – survey starting rate, along with the problem of reducing the percentage of persons resigning during the survey – dropout rate. In other words, the discussed questions include both convincing the respondents to commence the survey and to maintain their interest and motivate them to complete the survey.

Inclusion of evaluation surveys via CAWI methods in a project and within its structure

It might appear that CAWI method allows for a very fast and simple evaluation survey, however, an appropriate preparation is of essential importance for the quality of achieved results. Therefore the preparation of evaluation surveys should be commenced at a possibly earliest stage of project implementation that is subject to evaluation or even earlier – before it is commenced. It is not only about the decision on applying CAWI, but primarily about taking actions that will contribute to a higher survey response rate, while in the case of low response rate and unrandomised results of non-response error, they will allow to improve the possibilities of generalising the survey results by means of analytical weighting (more information on that in the final section of this chapter).

First, the means should be provided for exerting pressure on the persons and institutions participating in the programme and covered by evaluation in order to encourage them to take part in the evaluation survey. It is particularly important in the case of surveys that are extended in time. The obligation to take part in a survey can be indeed one of the conditions for joining a programme. It is also important to collect up-to-date and precise contact data of potential respondents. It is equally important to gather the information on the participants of evaluated measures that will be suitable e.g. to correct the results. The greatest possibilities are provided by the collection of the most important information about all participants (cf. section *Response rate after survey completion*).

Sample selection and increasing the response rate

Already at the stage of selecting the research sample, solutions can be introduced that will affect the response rate. First of all, it should be considered whether this method is appropriate for a given population at all. Surveys using the CAWI method are not suitable for populations where the use of the Internet is rare or the tendency towards using the Internet among its members is very diverse. The latter situation results from the fact that persons using the Internet more frequently and actively will be more willing to participate in a survey. Underrepresentation of certain groups that are harder to reach in an online survey, like seniors or persons with lower education, can substantially bias the results of a survey. In some cases mixed-mode data collection are applied allowing filling-in the surveys also via telephone or in writing. This increases the chances of participation by respondents that would not fill in a survey in the Internet. It should be remembered, however, that the same surveys carried out through various channels

may not be equivalent. An online survey is a self-administered interview whereas computer assisted telephone interviews (CATI) are carried out as a polled interview. These differences may exert influence on the provided answers not only due to the phenomenon of interviewer's effect, but also because some questions – just like the multiple-choice ones – are perceived differently by the respondent when displayed on the screen and in a different way when they are read by the interviewer.

If the survey organiser has some freedom in the selection of population to be surveyed, the advised solution is to select a less numerous one that is clearly defined and interested in the survey subject or convinced that the survey concerns it directly. There is no point examining a population that is overwhelmed by unwanted messages and advertisement information (spam) where the response rate will be particularly low[4]. At this point it should be noted that in the case of evaluation surveys, we have to do with clearly defined populations much more frequently than in other survey types – it is much more often known in detail which individuals belong to which population and which do not. Additionally, the respondents in evaluation surveys will be convinced more often that the survey concerns them (maybe with the exception of surveys regarding very broad consequences of evaluated measures).

The quality of contact information is of huge importance as well – if that information is of low quality, e.g. there are many out-of-date addresses or addresses receiving a lot of spam, the response rate will be lower. Also in this case in a large part of evaluation surveys we should not have this problem[5]. The condition for that, however, consists in collecting contact data (mainly up-to-date e-mail addresses) of persons and institutions participating in the programmes subject to evaluation. It is practiced frequently. With regard to the subsequent evaluation, more focus should be applied towards the collection of that contact information.

In the case of evaluation surveys, units of analysis are quite often not individuals, but organisations and institutions. In Poland, **c**ontrary to individuals, among which only 51% use the Internet[6], almost all companies and organisations have access to the Internet and an e-mail address. According to CSO (Central Statistical Office; Polish: GUS) research of 2008, 93% of companies employing at least 10 persons have access to the Internet[7]. In smaller companies the situation is even worse, the study under *Social Diagnosis* of March 2009, indicates however that over 80% of private entrepreneurs (mostly private businesss activities) use the Internet. Therefore CAWI surveys carried out among such populations make more sense.

The marginal costs of reaching more respondents regarding online surveys are often non-existent. Therefore, organisers often decide to examine the entire population instead of selecting a sample. Such measures, however, are not advised since a well selected sample provides the opportunity of getting the results that do not differ to a large extent from the results for the entire population. At the same time, in the case of more surveys being carried out for the same population, involvement of fewer people in the surveys means that the population members will feel less tired by constant requests to participate in the surveys.

---

[4] K. Rakocy, J.M. Zając, *Czy wierzysz w słowa te? Wiarygodność odpowiedzi uzyskiwanych w ankietach internetowych.* Poster presented at the National Scientific Conference: Social Aspects of the Internet SAI'06, Warsaw, 8–10.10.2006.

[5] In practice a better situation than in evaluation surveys can occur only in organisational surveys, in particular in organisations and institutions in which all the employees have e-mail account.

[6] D. Batorski, *Korzystanie, z technologii informacyjno-komunikacyjnych,* [in:] J. Czapiński, T. Panek, *Diagnoza społeczna 2009. Warunki i jakość życia Polaków,* Warsaw 2009, p. 279–307.

[7] *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.,* GUS 2008.

Research tools and response rate

The features of a research tool affect primarily the willingness complete a survey by the persons that commenced it (referred to as retention rate, i.e. the percentage of respondents that completed a survey in relation to those that commenced it; leaving survey before completion is defined as dropout, as mentioned above). These features affect the decision on commencing the survey to a lesser extent, especially because the respondent usually cannot see the tool before the decision on participation. Furthermore the tool features are of great importance due to the measurement error, this issue is beyond the interest of this article however.

Control of access to the study

Control of access to the study is an important element of online surveys. It is applied mainly in order to assure that one person does not participate in a survey more than once and to control who in the sample has already participated in a survey and who has not. If an invitation to a survey is sent via e-mail, a simple means of control consists in assigning an individual link to the survey to each respondent; it is defined by some authors as automatic login[8]. The advantage of this method is the simplicity of reaching the survey by respondents since all they have to do is to click on the link. Assigning individual links is more and more one of the functions of online survey software and due to that such a message can be sent automatically. This access control method seems currently to be most common in scientific practice.

Another solution is to require a manual login to the webpage containing the survey by separate access passwords and, if need be, also logins for each respondent. It means more trouble for the respondents due to the need to type in the access passwords, but at the same time, from the beginning it exacts more involvement and gives the feeling of confidentiality. Research shows that indeed the response rate is lower in comparison with automatic login. Data of higher quality is collected however, which is indicated by the dropout rate, longer time for (completing the questionnaire) and longer answers to open questions[9]. Similar solution consists in enabling the automatic login to the page by using an individual URL with manual provision of a simple password, e.g. 4-digit PIN code[10]. Such a control access method resulted in the highest response rate in the quoted survey and at the same time it was of only slightly lower quality than in the case of manual login. However, caution must be exercised while using the passwords and access codes. They may cause difficulties for less experienced Internet users.

Manual, semiautomatic or particularly automatic login to the survey are efficient methods of controlling the access to the survey, which should not substantially reduce the response rate at the same time. Survey access control using cookie files and all the more using the IP number is not advised due to lower precision. Additionally it can cause further problems and in extreme cases - lead to bias in result.

---

[8] S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden,* "Social Science Computer Review", 19, 2001, p. 146–162, D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic login procedure to control web survey access,* "Social Science Computer Review", 21 (2), 2003, p. 223–234.

[9] S.D. Crawford, M.P. Couper, M.J. Lamias, ibidem; D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers,* "Social Science Computer Review", 20, 2002, p. 10–21; D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic login procedure, to control web survey access.* "Social Science Computer Review", 21 (2), 2003, p. 223–234.

[10] D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic...*, ibidem.

Survey duration

One of the most important properties of web surveys is their duration. It is of importance for the response rate, primarily for the dropout in a survey. Indirectly it can also affect the inclination to participate in a survey if the invitation contents include the estimated time needed for the provision of all answers[11]. Generally, the surveys carried out online should be short, but deviations from this principle are possible when the surveys are oriented towards a specific, well-motivated population. In the case of long surveys, it is desirable to make it possible to save the survey and return to complete it later on. In such a case reminders should be sent to the persons that have not completed the surveys. Another solution is to divide long surveys into several parts and to send separate invitations for them at a later date[12]. In the latter case it is obviously important to maintain the possibility of respondent identification between the surveys.

Substantially, longer questionnaires may be used in surveys intended for institutions or organisations. However, also in the case when the evaluation survey covers companies or organisations and not individual persons, instead of a single long survey for the entire institution, several shorter ones should be prepared, which additionally can be filled in by various organisation members that have better orientation of particular issues covered by the survey.

The specific nature of CAWI surveys and great sensitivity of respondents to interview duration translate into the fact that short surveys prove more efficient. Therefore it is better to investigate one issue in one survey thus by organising several separate surveys than to investigate many various issues within one survey. The organisation of several similar surveys might be a bit more complicated, but on the other hand the costs of contacts with the respondents or costs of creating several survey tool versions are of very little importance in CAWI surveys.

A factor making the impression of a longer survey is the division thereof into multiple pages. In the case of short and simple surveys, it is advised to place them on a single page since if all questions are not visible, this usually leads to dropouts[13]. Yet, at the same time, in many situation the division of questionnaire is necessary: in the case of very long surveys; when using filtering questions (although some tools allow for filtration within one page); when we do not want the respondents to see the subsequent questions before answering the previous ones; when we want to precisely assess the moment at which a dropout occurs.

In the case of longer surveys and division of subpages, the suggested solution is the progress indicator[14]. It can have various forms ranging from the simplest indicators such as "Page 3 of 7", through percentage and graphical indicators, to dynamic indicators adjusting to the pace of providing answers by the respondent and answers to filtering questions. Progress indicator can be applied as well informing the respondent that the survey will end much earlier than in reality. However, as indicated by results of Kaczmirek[15], it causes in fact a greater response rate during the survey than reliable progress indicators.

---

[11] S.D. Crawford, M.P. Couper, M.J. Lamias, op. cit.

[12] I.S. Choi, *Online research in Korea: Reliability and validity issues.* Speech at the conference: The 1st International Workshop on Internet Survey, Daejeon, Republic of Korea, 9-10.09.2009.

[13] D. Batorski, *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe wspólnoty.* (KBN Report 1-H02E-029-27). Warsaw: the University of Warsaw, Institute for Social Studies, 2005.

[14] S.D. Crawford, M.P. Couper, M.J. Lamias, *op. cit;* L. Kaczmirek, *Human-Survey Interaction: Usability and Nonresponse in Online Surveys.* Cologne: Herbert von Halem Verlag (hard copy).

[15] Ibidem.

The sequence of questions is of special importance, especially in the case of longer surveys. A survey should start with simple questions that are interesting at the same time. Questions about personal data, in which the respondents are least interested, should be located at the end of the survey, with the exception of situations when socio-demographic features are most important in the survey[16] since locating them at the end entails omitting them more frequently by the respondents. Conversly, as indicated by research by Reips[17], asking questions about personal data at the survey beginning increases the motivation of respondents – by answering the questions about gender, age, etc., the respondents reduce their feeling of anonymity and while feeling less anonymous, they feel more obliged to participate in a survey and drop out less frequently prior to survey completion. Seemingly, the results of various surveys might seem contradictory to each other. However, one should expect that the results of long and boring questions about personal data, which may discourage from participation, are different if they are located at the beginning of a questionnaire and other if in 3-4 short questions about the most general data.

Answer format

The answer format is of certain importance for the participation of respondents in a survey. Some studies[18] prove that respondents prefer radio buttons to drop-down menus. Such format is simply more comfortable, especially in the case of short lists (several elements). In the case of populations of active Internet users that deal with these formats when filling in various forms on a daily basis, the format can be of lesser importance. Open questions can be difficult for respondents, narrative questions in particular. Answers to them require more effort than marking one of the provided answers in a closed question. In such a situation the respondents might resign from the survey at all or not answer these questions (non-response item).

It is relatively easy in online surveys to prevent the respondent from proceeding to the next question without providing an answer to the previous one. Such solutions assure that closed surveys have all answers completed. Nonetheless, at the same time it often causes an increase in dropouts, especially in the case of sensitive questions, for instance about income. In such situations the respondent must not be forced to provide the answer or answer "I do not know / refuse to answer" should be added.

Technical aspects of questionnaires

Technical solutions can prevent some respondents from participating in a survey or impede it. While designing the research tool, one should consider the equipment and software used by the surveyed sample[19]. Generally, the simplest solutions prove to be most efficient. Even such popular technologies as JavaScript, Java or Flash are not used by all Internet users. More serious problems might occur in the case of playing video materials. It seems that Flash is much more popular than any other video player (e.g. Windows Media Player or RealPlayer). It should be checked as well whether the research tool is displayed properly in various web browsers (at least Internet Explorer and Firefox) and with various screen

[16] D.A. Dillman, R.D. Tortora, D. Bowker, *Principles for Constructing Web Surveys.* (SESRC Technical Report 98-50). Pullman, Washington, 1998.

[17] U.-D. Reips, *Standards for Internet-based experimenting,* Experimental Psychology", 49 (4), 2002, p. 243–256.

[18] E.g. D.A. Dillman, D.K. Bowker, *The Web questionnaire challenge to survey methodologists,* [in:] U.-D. Reips and M. Bosnjak (ed.), *Dimensions of Internet science,* Lengerich: Pabst Science Publishing, 2001, p. 159–178.

[19] L. Kaczmirek, ibidem.

resolutions. Popularity statistics of such solutions can be found for instance on the following website www.ranking.pl.

Technical problems may be solved neither by asking before the survey about the used equipment and software since a part of respondents might not know that, nor by providing a download location of the necessary programmes since a part of respondents might not be able to install it or have the necessary authorisations. It is possible, however, that a website with the survey verifies some technical parameters such as what web browser is used and whether Flash or Java is installed. Therefore it is possible to display various website versions, depending on certain technical parameters of the computer or of web browser used by the respondent. On the other hand, minimal requirements are listed in some surveys, such as the installation of an additional programme, by adding a link where the respondent can download the given programme.

The speed at which the page is uploaded is of certain importance as well[20]. When survey websites are slow, the respondent might be discouraged and resign from the participation. Admittedly with technological development and popularisation of broadband Internet, these features are less important, but elements should be avoided that are too "heavy" and it should be verified whether a large number of simultaneous website visits slows the server down.

Survey management tool

What should be mentioned at this point is the tool for survey management that can substantially facilitate the researcher's work. The tool should allow for simplified sending of invitations and reminders to the respondents, preferably with the possibility of automatic personalisation of the mail contents, including individual links allowing for automatic login to the survey. Monitoring of the response rate during the survey is also beneficial, tracking of dropout at particular questions since it happens that questions exacting an answer may generate a large increase in the resignations from survey. It should be also possible to verify which respondents have not yet participated in the survey and which have done so, but have not finished filling in the survey, and to send repeated invitations and reminders to those persons. The possibility of generating simple reports during the survey is also convenient.

Duration of survey and sending invitations

During survey planning one should take into consideration the rhythms of activity of the surveyed population. Sending an invitation for a survey at an inappropriate moment reduces the chances that the respondents react to them or receive them at all. If possible, the survey should be carried out at the moment when there are the highest chances that the invitation has been received by respondents at the most favourable moment. The surveys carried out during vacation season, winter or summer school holidays, etc. are less efficient since the potential respondents use the Internet less frequently or even do not use it at all[21] whereas in the organisational context – in the period of particularly intensive work.

[20] D.A. Dillman, R.D. Tortora, D. Bowker, ibidem; U.-D. Reips, ibidem.

[21] S.D. Crawford, M.P. Couper, M.J. Lamias, ibidem; S.J. Sills, Ch. Song, *Innovations in survey research: An application of Web-based surveys*. "Social Science Computer Review", 20, 2002, p. 22–30.

What is more, surveys of very diverse populations confirm that a gross of respondents participates in online surveys shortly after the reception of invitation[22]. Similarly, short-lasting increases in the number of survey completions can be anticipated shortly after sending the reminders[23]. It should be taken into consideration since in the case of applying faulty technologies in the survey; a too large number of simultaneous survey website visits may cause technical problems, including the impediment of survey implementation.

It should be taken into consideration that many Internet users are overburdened with the inflow of information. If they do not react shortly after the reception of invitations, it is probable that they will not remember it and will not react at all. Goeritz and Steiger[24] suggest that in surveys registered for the participants of scientific panel during the first 3 days of surveying, 70-90% of all answers are delivered. Hamilton[25] indicates in his meta-analysis that over a half of respondents participated in surveys during the past 24 hours. He points out as well that the invitations sent between 6 and 9 am on working days were most efficient. In another survey[26] the most efficient invitations were the ones sent in the morning (7:30-8:30) and at the end of working day (15:30-16:30). These results appear to be important when the respondents participate in the survey at work. However, if we expect them to do it during leisure time, invitations sent in the afternoon may be more efficient, so that the respondents read them when they seat at the computer after coming back home from work or school.

Field time is also important. A short period is usually enough (up to 2 weeks); although, at the same time a survey should last at least one week due to various weekly activity rhythms. The reason for prolonging this period may be unlimited accessibility of the surveyed population or a part thereof during the survey period or will to apply other channels to reach the respondents under mixed schemes, especially the mail. As far as a notification on the completion date is provided or not in the invitations and the fact whether the time limit should be short or relatively longer, the results of previous surveys are not conclusive[27]. It seems, however, that the information about the data collection deadline should be found in the invitation for organisational purposes.

Inclination to fast response to the invitations is sometimes used in case of very fast surveys when the results are needed within 2-3 days. It is assumed in such a case that a substantial reduction of time for data collection will cause a slight decrease in response rate since a gross of answers would be collected shortly after the start. Of course, such a solution reduces the representativeness and prevents the application of certain techniques of increasing the response rate.

---

[22] P. Mazurek, J.M. Zając, K. Rakocy, *Online digital surveillance: What they know about you depends on what you know.* Speech at the conference: General Online Research 07, Leipzig, 26-28.03.2007; J.M. Zając, *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej,* [in:] M. Sokołowski (ed.), *Oblicza internetu: Internet w przestrzeni komunikacyjnej XXI wieku,* Wydawnictwo PWSZ w Elblągu, 2006, p. 167–178.

[23] S.D. Crawford, M.P. Couper, M.J. Lamias, ibidem.

[24] A.S. Goeritz, S. Stieger,. *The impact of the field time on response, retention, and response completeness in list-based Web surveys,* "International Journal of Human Computer Studies", 67, 2009, p. 342–348.

[25] M.B. Hamilton, *Online Survey Response Rates and Times: Background and Guidance for Industry.* SuperSurvey White Paper, 2003.

[26] Granello and Wheaton, 2004, quotation of: V.M. Sue and L.A. Ritter, *Conducting online surveys*. Sage, 2007.

[27] A.S. Goeritz, S. Stieger, *The impact of the field time on response, retention, and response completeness in list-based Web surveys,* "International Journal of Human Computer Studies", 67, 2009, p, 342–348; S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates,* "Public Opinion Quarterly" 67, 2003a, p. 579–588.

Invitation to participate in a survey

The importance of invitation to a survey for response rate has been known to methodologists for a long time. An appropriate method of contacting the respondents attracts their attention and motivates to participate. Invitation is particularly important in online survey since in contrast to mail polls, for instance, the respondents usually do not watch the survey tool before the decision on the participation[28]. It is important especially in a situation when the following parts of the questionnaire are located on various websites and in the case of controlling the access to the survey (e.g. by means of a password). What is more, invitations to a survey in the Internet should be short and concise. Optimisation of the invitation seems to be a relatively simple method of increasing the motivation of respondents in online surveys that additionally does not require particular financial outlays by the researcher, such as gratuity.

Contacts before the survey

Pre-notification contacts with respondents prepare for the participation and facilitate the subsequent communication. They perform mainly informational function, however, they can contribute to an increase in the response rate. At this stage the quality of mailing list or another applied contact method can be verified. If the survey concerns an organisation, support can be provided to it in the preparation for the survey, for instance by means of assigning relevant persons or collecting necessary information. Social influence technique called "foot-in-the-door" technique can be used at this stage, namely a small request preceding the actual survey in this case[29]. According to the impact mechanism described by Cialdini[30], the respondents committed to the fulfilment of a small request are more willing to fulfil larger ones as well. Caution should be exercised though to prevent such measures from affecting the results of the actual survey.

Method of delivering the invitations

The selection of a proper communication form to send an invitation depends to a large extent on the available information on the examined sample. Currently, the most common one consists in sending invitations via electronic mail that has been the most popular communication form in the web for many years. According to the research Social Diagnosis 2009, 90% of Polish users older than 16 years use it, and thus more or less 45% of adults. Similarly, an invitation can be sent by means of a web communicator[31].

If the survey organiser does not know such contact data or wants to conduct the survey using a sample of volunteers, invitations to the survey can be published on the websites visited by the local population. Invitations can be additionally distinguished by graphical means; however, they should not resemble advertisements too much. Invitation by means of websites is applied frequently when the survey organiser does not have the data allowing for contact with the respondents via e-mail or beyond the Internet. Unfortunately this solution does not provide an opportunity of controlling the respondents and precise determination of the response rate.

Another solution is to display the invitations in pop-up windows displayed during website visits. It

[28] S.D. Crawford, M.P. Couper, M.J. Lamias, ibidem.

[29] N. Guéguen, *Foot-in-the-door technique and computer-mediated communication,* "Computers in Human Behavior", 18, 2002, p. 11–15.

[30] R. Cialdini, *Wywieranie wpływu na ludzi.* Gdańsk: GWP, 1996.

[31] D. Batorski, ibidem.

allows to randomly selecting respondents among the website users on the basis of cookie files when there is no other base list. This method can be applied especially when there is no possibility of displaying the invitations on profile website. It has significant limitations at the same time. Emerging pop-up windows resemble disliked advertisements and some users block such forms in their web browsers and therefore the chances of attracting attention to the invitations decrease. What is more, closing such a window causes that a return to the survey at a more convenient moment becomes impossible. Therefore, if possible, online surveys invitations thereto included in emerging pop-up windows should be avoided.

Means used beyond the Internet – mail, telephone or direct meetings – can be utilised to invite people to participate in a survey and to remind them about it. Especially telephone reminders can be efficient since assistance can be provided to the respondents at the same time, e.g. in the case of technical problems. Contact with the respondents should be planned beforehand. A great advantage of evaluation research is constituted by the fact that many programmes subject to evaluation allow for collection of contact data of participants throughout their duration.

Invitation sender

In the case of surveys carried out beyond the Internet, the information on organisers and sometimes on the sponsors is used in order to authenticate the survey and to convince the respondents to participate therein[32]. The sender that is more reliable and enjoys an authority should be more convincing for respondents. In the case of online surveys, such techniques are not always efficient. There are several reasons for that: application of similar methods of authentication and attracting attention by the senders by means of spam, therefore many Internet users do trust such manipulations[33]; specific nature of behaviour and the feeling of anonymity in the Internet[34] causing that Internet users can be less willing to subordinate to authorities; and finally the fact that Internet users usually do not read carefully the received messages and only browse through them.

Contrary to appearances, the sender of higher social status and authority does not always improve the response rate. Admittedly a university professor in the experiment conducted by Guéguen and Jacob[35] reached a higher response rate than a student, but in the research of Porter and Whitcomb[36] and Zając[37] the manipulation of various sender traits did not exert any influence on the inclination of respondents to participation in the survey. It seems though that a reference to the power of the organiser and sponsor of the survey over the respondents is efficient[38]. Therefore, the conclusion is drawn that invitations issued

[32] R.M. Groves, F.L. Fowler Jr., M.P. Couper, J.M. Lepkowski, E. Singer, R. Tourangeau, *Survey methodology* Wiley, 2004.

[33] S.R. Porter, M.E. Whitcomb, *E-mail subject lines and their effect on web survey viewing and response,* "Social Science Computer Review", 23 (3), 2005, p. 380–387.

[34] A.N. Joinson, *Internet behaviour and the design of virtual methods,* [in:] C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet*, Oxford: Berg, 2005.

[35] N. Guéguen, C. Jacob, *Solicitation by e-mail and solicitor's status: A field study of social influence, on the Web,* "CyberPsychology & Behavior", 5, 2002, p. 377–383.

[36] S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates,* "Public Opinion Quarterly" 67, (2003a), p. 579–588.

[37] J. M. Zając, *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej,* [in:] M. Sokołowski (Ed.), *Oblicza internetu: Internet w przestrzeni komunikacyjnej XXI wieku,* Wydawnictwo PWSZ w Elblągu, 2006, p. 167–178.

[38] A.N. Joinson, U.-D. Reips, *Personalized salutation, power of sender and online survey response rates to web-based surveys,* "Computers in Human Behavior" 23, 2007, p. 1372–1383; A.N. Joinson, A. Woodley, U.-D. Reips, *Personalization, authentication and self-disclosure in self-administered Internet surveys,* "Computers in Human Behavior" 23, 2007, p. 275–285.

under the auspices of, for instance, company heads or directors of programmes covering numerous enterprises are more efficient. A possibility of an earlier commitment made by the participants of such programmes to participate in evaluation surveys later on could be considered as well.

Efficiency of an invitation can be improved by additional information about the sender, e.g. a photo or manipulations of the sender's identity, but the possibility of applying such techniques seems limited. A greater response rate has been confirmed in surveys after adding a photo to the invitation in e-mail[39] and in a pop-up window[40]. It seems, however, that the attractiveness of the person on the photo, their gender and the gender of the respondent, might be of great importance, as well as the interactions between these factors, but this issue needs to be further examined. In another interesting survey carried out by Agnieszka Janiszewska from Gemius[41], it was noticed that new faces in the pop-up invitation attract attention more than a face used for this purpose multiple times earlier and thus well-known to the respondents. Finally Althoff, Greif, Griel and Batinic[42] suggested that invitations sent by a female sender are more efficient (the so-called Anita effect) with the proviso that it is related to a situation where male respondents are predominant in the sample and the survey is carried out using a group of persons that use electronic e-mail intensively and need to be convinced to commit time to participate in a survey. The usefulness of the above hints is small, mainly due to the fact that they are strongly dependent on the survey context. Furthermore, sending untrue information on the survey organiser and performer would be in most cases unethical.

The inclination to participate in a survey is clearly higher in a situation when the person conducting the survey or sender of the invitation is a member of the community, especially when this community has a group identity. This fact might result from a greater inclination to act for the community members. Even if the researcher does not belong to it, it can be helpful to employ a person that is an authority for such a community.

Invitation contents

An appropriate formulation of message informing on the survey can be helpful in improving the response rate. The functioning mechanisms can be explained by referring to the processes of social influence. Caution should be exercised in order not to exaggerate with such techniques, especially when the survey concerns confidential and sensitive topics since the survey situation in such a case will be of social nature too much[43].

In most situations the method of increasing the response rate with no detriment of the quality of data can be personalisation of the invitation. The most common method of addressing a respondent is by using their name and surname (of course the researcher has to know these data before sending the

[39] N. Guéguen, C. Jacob, *Social presence reinforcement and computer-mediated communication: The effect of the solicitor's photography on compliance to a survey request made by e-mail,* "CyberPsychology & Behavior", 5, 2002a, p. 139–142.

[40] J. Przewłocka, P. Janczewski, *Jak e-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online.* Speech at the 8th PTBRiO Congress, Warsaw 18–19.10.2007.

[41] A. Janiszewska, *Działania prowadzone, przez Gemius na rzecz podnoszenia response rate w badaniach online.*. Research paper. Warszawa: Gemius SA, 2009.

[42] S. Althoff, V. Greif, B. Griel, B. Batinic (2006). *Determinants of response rates of online surveys – the Anita effect: Results of a joint project.* Speech at the conference: General Online Research 06, Bielefeld.

[43] A.N. Joinson, *Internet behaviour and the design of virtual methods,* [in:] C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet.* Oxford: Berg, 2005.

invitations!). Most surveys confirm the efficiency of this method[44] with the exception of surveys by Porter and Whitcomb[45]. Joinson and Reips[46], on the basis of results of a series of experiments, suggested that the absence of influence of personalised greeting on the response rate in certain cases results from the simultaneous reduction of the impression of respondent's anonymity.

A simple method of personalising an invitation consists also in sending an individual message to each respondent instead of e-mails sent simultaneously to the entire group (even as undisclosed copies), which makes an impression of a more individual request[47]. Other methods of referring to the exceptionality of the survey and inaccessibility rule[48] can be also tested, but their efficiency might be limited. In the survey by Porter and Whitcomb[49] such as "you are one of the very few people participating in this survey" did not exercise any influence on the response rate.

From the point of view of social influence techniques, it is not surprising that formulating the invitation as a request is helpful[50]. It is important to include information at the beginning about the organisers and contact possibilities with the researchers. Irrespective of the response rate, it is required by the standards of conducting surveys anyway. The respondents have also the right to the information about the accessibility of survey results and the method of using them. What is more, such accessibility and social usefulness of a survey, as discussed below, can be perceived as a form of gratuity and thus they can have a positive influence on the motivation to participate in a survey.

It is very important to define the time needed to fill in the survey in the invitation. A shorter survey requires less effort from the respondent and therefore indicating shorter time increases the inclination to commence the survey. The respondents should not be misled since they are more prone to resign while filling out the survey if they realise during the participation that the survey takes longer than predicted[51].

It is hard to express an opinion on the meaning of invitation subject. In the survey conducted by Porter and Whitcomb[52], it had no influence on the efficiency of invitations. Certainly one should properly consider the title of the e-mail with invitation that at the same time does not discourage the respondents and does not resemble a typical spam title.

---

[44] D. Heerwegh, *Effects of personal salutations in e-mail invitations to participate in a Web survey* "Public Opinion Quarterly", 69 (4), 2005, p. 588–598; D. Heerwegh, T. Vanhove, K. Matthijs, G. Loosveldt, *The effect of personalization on response rates and data quality in web surveys,* "International Journal of Social Research Methodology: Theory and Practice", 8, 2005, p. 85–99.

[45] S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates,* "Public Opinion Quarterly" 67, 2003a, p. 579–588.

[46] A.N. Joinson, U.-D. Reips, Personalized salutation, power of sender and online survey response rates to web-based surveys, "Computers in Human Behavior", 23, 2007, p. 1372–1383; A.N. Joinson, A. Woodley, U.-D. Reips, *Personalization, authentication and self-disclosure in self-administered Internet surveys*. Computers in Human Behavior", 23, 2007, p. 275–285.

[47] D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers,* "Social Science Computer Review" 20, 2002, p. 10–21.

[48] Cf.R. Cialdini, *Wywieranie wpływu na ludzi.* Gdańsk: GWP, 1996.

[49] S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates,* "Public Opinion Quarterly", 67, 2003a, p. 579–588

[50] Cf.A.R. Trouteaud, *How do you ask counts: A test of Internet-related components of response rates to a Web-based survey* "Social Science Computer Review", 24, 2004, p. 385–393.

[51] S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden,* "Social Science Computer Review", 19, 2004, p. 146–62.

[52] S.R. Porter, M.E. Whitcomb, *E-mail subject lines and their effect on web survey viewing and response,* "Social Science Computer Review" 23 (3), 2005, p. 380–387.

Appearance of the invitation

With the development of internet technologies and capacity of internet connections, possibilities of designing specific appearance of invitations emerged, for example by means of graphical elements and an unusual template in general. However, such solutions may cause negative effects leading to a decrease in the response rate. The respondents will be less willing to answers such invitations. It can result from a greater volume and thus longer loading time of such messages, from a feeling of flooding with incoming incentives and mainly from the fact that such methods are associated primarily with undesired mail. Such conclusions are supported by the survey conducted by Whitcomb and Porter[53] who proved a negative influence of coloured background and complex graphics of an e-mail invitation on the percentage of responses. It should be kept in mind as well that graphical elements can be displayed improperly, especially because the users utilise various mail clients and a very large number of people checks mail only by means of websites.

In the case of invitations displayed as pop-up windows, the results of Polish researchers suggest that adding uncomplicated graphical elements to the text can increase its efficiency, but simple and aesthetical looks still remain the most important features[54]. Therefore we recommend moderation and graphical simplicity of the invitations and suggest avoiding long texts[55].

Gratuity

A quite obvious method of influencing the motivation to participate in a survey is remuneration for the respondents. A gratuity can be in financial, material or symbolic form (thanks, feedback or a possibility of comparing one's own results with the others). It can be delivered both after the completion of a survey and before the commencement thereof, with the latter being usually more efficient in the case of payment.

In evaluation surveys, it is particularly important to use symbolic graphics, which results from the standards of conducting research anyway – the respondents should be thanked and provided with the explanation what is their participation needed for. In the case of surveying an organisation or institution, the access to the results and possibility of comparison with the others are particularly important for the respondents. Even if only collective results are provided, they can represent interesting and needed information anyway.

Financial and material gratuity usually has a positive influence on the response rate and limits the dropout[56]. However, the effects are sometimes small and do not always compensate the costs and additional effort concerning for instance the transfer of prizes. There is not much known about the results of gratuities in time when several surveys are being conducted using the same sample[57]. A greater

[53] M.E. Whitcomb, S.R. Porter, *Email contacts: A test of complex graphical designs in survey research*, "Social Science Computer Review" 22 (3), 2004, 370–376.

[54] J. Przewłocka, P. Janczewski, *Jak e-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online*. Speech at the 8th PTBRiO Congress, Warsaw 18–19.10.2007.

[55] J.M. Zając, D. Batorski, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie, realizacji próby*, „Psychologia Społeczna" No. 3, 2007, p. 234–247.

[56] A.S. Goeritz, *Incentives in Web studies: Methodological Issues and a review*, "International Journal of Internet Science", 1(1), 2006, p. 58–70.

[57] A.S. Goeritz, D.G. Goldstein, *Individual payments as a longer-term incentive in online panels*, "Behavior Research Methods", 40 (4), 2008, p. 1144–1149.

result may be provided by investing additional resources for instance in personalisation of invitations or creation of enhanced research tools. What is more, the prizes of greater value are usually only a bit more efficient.

The effects of material gratuity depend strongly on the context and specific nature of a given population. If the respondents participate in a survey with interest and willingness to help, provision of material incentives does not exert significant influence on the inclination to participate. What is more, giving prizes is generally a bad practice since it destroys inner motivation. Gratuity often attracts a specific sort of respondents, which paradoxically might decrease the possibility of generalising the survey outcomes. Additionally, in the case of remuneration for the participation, attempts of multiple participation[58] and dishonest filling in the surveys just to complete them and get the prize are more frequent[59].

In the case of financial and material prizes, it is possible to reward everyone that participated in a survey and to organise lotteries in which a defined number of persons wins the prize of the highest value. In the case of a lottery, a large amount can impress the respondents. Furthermore, the award of a greater prize to one person instead of multiple small prizes makes it easier to send and the costs of delivering the prizes are lower. There is no certainty though whether lotteries are more efficient than prizes awarded to everyone[60] and whether more impressive prizes with lower odds of winning in fact attract more than lesser prizes with a greater probability of winning, as suggested by Deutskens and co-researchers[61]. However, if a lottery is organised, its results should be available for the researcher immediately after the survey completion[62] since the respondents may not trust the lottery.

It should be mentioned as well that currently the use of material prizes does not have to be limited to sending the prizes beyond the Internet. It is possible to use material gratuities in the Internet, for example such as access to paid websites, a possibility of downloading materials, discount coupons or loyalty points on commercial websites[63]. The results of a survey by Birnholtz and co-researchers suggest that cash is more efficient since electronic coupons are of objectively lower value for the respondents than the equivalent value in cash.

---

[58] D. Batorski (2005), ibidem; J.A. Konstan, B.R.S. Rosser, M.W. Ross, J. Stanton, W.M. Edwards, *The story of subject naught: A cautionary but optimistic tale of Internet survey research,* "Journal of Computer Mediated Communication", 10 (2), 2005.

[59] A.S. Goeritz, *Incentives in Web studies: Methodological issues and a review,* "International Journal of Internet Science", 1(1), 2006, p. 58–70.

[60] Cf. D. Batorski, *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe. wspólnoty.* (KBN Report 1-H02E-029-27). Warsaw: the University of Warsaw, Institute for Social Studies, 2005; S.R. Porter, M.E. Whitcomb, *The. impact of lottery incentives on student survey response, rates,* "Research in Higher Education" 44 (4), 2003 b, p. 389–407. T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate, versus delayed notification of prize draw results on response behavior in web surveys: An experiment.* "Social Science Computer Review" 22 (3), 2004, p. 377–384.

[61] E. Deutskens, K. de Ruyter, M. Wetzles, P. Oosterveld, *Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study,* "Marketing Letters" 15 (1), 2004, p. 21–36.

[62] T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate versus delayed notification of prize draw results on response behavior In web surveys: An experiment,* "Social Science Computer Review", 22 (3), 2004, p. 377–384.

[63] J.P. Birnholtz, D.B. Horn, T.A. Finholt, S.J. Bae, *The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents,* "Social Science Computer Review", 22 (3), 2004, p. 355–362; A.S. Goeritz, *incentives in Web studies: Methodological issues and a review,* "International Journal of Internet Science" 1(1), 2006, p. 58–70, T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment,* "Social Science Computer Review", 22 (3), 2004, p. 377–384.

Response rate during survey

Response and dropout rates should be certainly traced during the survey since a lot valuable information can be gathered this way. For instance a high percentage of respondents dropping out during a survey, especially if it concerns one of the questionnaire elements, may be an indication of an improperly formulated question or technical problems. If the response rate is very low, additional means should be undertaken, for instance additional methods of contacting the respondents described below or even changes in the survey procedure and for instance contacts with respondents also beyond the Internet or organising another way of providing answers, for example via telephone.

A good method of increasing the inclination to participate in a survey is to send reminders and mobilise the persons that have not participated immediately[64]. The respondents can be contacted beyond the Internet, for instance via mail, telephone or during meetings. In order to avoid bothering the respondents, the reminders should be sent only to the persons that have not yet filled in the survey. If there is no such information, the solution is to send a reminder to the entire surveyed group that is combined with thanks to the ones that already participated therein and with information on the survey completion date[65]. Reminders are advised in a situation when the respondent might interrupt filling in a survey and not return to it afterwards.

A reminder brings the desired effects if at the moment of its reception the respondent's situation is different than when they received the first invitation. In the case of online surveys, such a situation may change quickly. For instance when the respondent received an e-mail invitation on the day when they got many other, more important communicates at the same time and overlooked it or postponed for later, they should be reminded about it at a more convenient moment. Therefore it seems that reminders in online surveys can be sent in small time intervals, for instance every 2 days[66], especially when the respondents actively use electronic mail. What is more, after a short period of time it is easier for the respondents to associate the invitation with the reminder whereas after a longer time lapse they might think that a new survey is meant. Sometimes it happens though that reminders are sent less frequently, like in the second, third and fourth week of a 6-week survey[67]. There should not be too many reminders. The final outcomes of subsequent communicates are decreasing, besides they can be tiresome and annoying for the respondents[68].

Response rate after survey completion

After the completion of a survey, the response rate should be subject to assessment. It is not only about the percentage of invited persons that completed the survey, but also about the structure of participation. It is particularly important to have other sources of information concerning the surveyed

[64] D.A. Dillman, *Mail and Internet Surveys: The Tailored Design Method,* New York: John Wiley, 2000.

[65] D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers,* "Social Science Computer Review" 20, 2002, p. 10–21.

[66] S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden,* "Social Science Computer Review" 19, 2001. p. 146–162.

[67] J.P. Birnholtz, D.B. Horn, T.A. Finholt, S.J. Bae, *The effects of cash, electronic, and paper gift certificates as respondent Incentives for a web-based survey of technologically sophisticated respondents,* "Social Science Computer Review", 22 (3), 2004, p. 355–362.

[68] D.J. Solomon, *Conducting web-based surveys,* "Practical Assessment, Research & Evaluation", 7 (19), 2001.

population that will allow for a comparison of the structure of conducted sample with the population structure or, to be more precise, with the structure of drawn sample. It is essential whether there are groups whose participation in a sample is clearly different from the prevalence in the population. If so, and the variable isolating the groups are significantly correlated with the highest variables under research, this probably means that there is a significant distortion of survey results.

Even after the survey completion, there are possibilities of correcting the results and increasing the representativeness of results. They consist in the construction of analytical balances. The results are weighted so that individuals with lesser participation are taken into consideration to a greater extent and less weight is attributed to groups with a higher response rate. There are various methods of constructing such balances, but it should be pointed out that they require the possession of information about the entire population, which might entail the necessity of a reference survey with better characteristics carried out beyond the Internet. Collection of data, even basic, such as gender, age, etc. should be planned already during measures covered by evaluation, particularly when it is possible to acquire such information about all participants.

The most common methods of reducing the non-response error include weighting by means of stratification balances. This method requires only the information on the frequency schedules of most important variables in the population and consists in adjusting the structure of completed sample to the structure of the entire population. However, in online surveys[69], the method of propensity score weighting[70] is applied more and more frequently. It is more difficult, in so far as this method requires the possession of data on all population members, but it is impossible in the case of certain evaluation surveys. On the basis of these data it is possible to define the probability of participation in the survey for each of the persons in a population. The opposite of this probability is later used for the construction of analytical balances used to weight the results. The effect of applying propensity score weighting is the reduction of errors of estimating online surveys[71]. Irrespective of the weighting method, the use of variables that are strongly correlated with the variables interesting for the researchers for the purpose of constructing balances gives particularly good effects.

Summary

Unwillingness to participate in surveys and low response rate are some of the greatest problems concerning online surveys. At the same time many methods of addressing this problem can be indicated. As presented above, measures aimed at increasing the tendency of respondents to commence and complete a survey can be undertaken at every survey stage, ranging from the selection of surveyed population and sample through the choice of survey duration, communication with the respondents, provision of gratuities, appropriate reaction during the survey, to secondary methods of assuring

[69] S. Lee and R. Vallian, *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment* "Sociological Methods and Research" 37(3), 2009, p. 319–343; M. Schonlau, A. van Soest, A. Kapteyn and M. Couper, *Selection Bias in Web Surveys and the Use of Propensity Scores,* "Sociological Methods and Research" 37(3), 2009, p. 291–318.

[70] P. R. Rosenbaum and D.B. Rubin, *Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score.* JASA 79(387): 1984, 516–524; R.J.A. Little and D.B. Rubin (1987), *Statistical Analysis with Missing Data.* New York: John Wiley and Sons, 1987.

[71] S. Lee and R. Vallian, *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment,* "Sociological Methods and Research" 37(3), 2009, p. 319–343.

representativeness, At the same time it should be kept in mind that adequacy and efficiency of particular measures depends on the context of a given survey.

While affecting the inclination to participate in a survey, one should consider whether these operations increase non-random deviations. Such a threat occurs when the efficiency of operations is not uniform due to the features of respondents. This problem is particularly significant when we want to generalise the survey outcomes of the population sample. An increase in the response rate in such a situation does not have to mean an increase in the representativeness, on the contrary – a decrease. Furthermore, some of the methods of affecting the participation in a survey, for example gratuity, can lower the quality of data at the same time. It should be kept in mind as well that some methods of increasing the inclination to commence a survey may cause a growth of dropouts at the same time. The influence exerted on the respondents should not be exaggerated, especially if the goal does not consist in generalisation on the basis the population sample, but rather a search for dependencies between variables.

As presented above, due to many reasons, evaluation surveys are often suitable for implementation by means of CAWI method. We think, however, that evaluation surveys will use the technique of online surveys more frequently in the future. Therefore the ever increasing knowledge on the organisation of surveys and behaviour of respondents should be used and the surveyed persons should be motivated.

Jan M. Zając, PhD – is an Assistant Professor of Psychology at the Faculty of Psychology at the University of Warsaw. He holds M.A. in Psychology (University of Warsaw, 2004) and in Economy (Warsaw School of Economics, 2003). He participated also in international scholarship programmes and studied at Bocconi University in Milan and at Katholieke Universitaet in Leuven in Belgium.His scientific interests regard mainly psychological and social aspects of Information and Communication Technologies (ICT), online research methods and social networks analysis.

Dr Zając authored and co-authored several publications concerning mainly psycho-social aspects of the Internet and methodology of online studies, as well as participated in many academic and commercial projects regarding quantitative research onlineHe has been a member of the International Review Board of the cyclic conferences "General Online Research" (since 2008).

Dominik Batorski, PhD – Assistant Professor of Sociology at the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw. His scientific work encompasses primarily the issues of social networks, as well as dynamic of social processes occurring within networks such as diffusion of innovation or social impact. He also works on social network analysis methods and methodology of carrying out psychological and social research via the Internet. He is a member of the Council for Social Monitoring preparing the cycle of studies called Social Diagnosis. In this research he is responsible for the module devoted to conditions, and consequences of using information-communication technologies. He is a member of several scientific associations; inter alia the International Network for Social Networks Analysis and the Polish Sociological Association.

Sources

Althoff S., Greif V., Griel B., Batinic B. (2006), *Determinants of response rates of online surveys – the Anita effect: Results of a joint project***.** Speech at the conference General Online Research 06, Bielefeld.

Batorski D. (2005), *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe wspólnoty.* (KBN Report 1-H02E-029-27). Warsaw: the University of Warsaw, Institute for Social Studies.

Batorski D. (2009), *Korzystanie z technologii informacyjno komunikacyjnych,* [in:] J.Czapiński and T. Panek, *Diagnoza społeczna 2009. Warunki i jakość życia Polaków,* Warsaw.

Batorski D., Olcoń M. (2006), *Prowadzenie badań przez internet: Podstawowe zagadnienia metodologiczne.* „Studia Socjologiczne" 3(182).

Birnholtz J.P., Horn D.B., Finholt T.A., Bae S.J. (2004), *The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents.* "Social Science Computer Review", 22 (3).

Choi I.S. (2009), *Online research in Korea: Reliability and validity issues*. Speech at the conference The 1st International Workshop on Internet Survey, Deajeon, Korea, 9-10.09.2009.

Cialdini R. (1996), *Wywieranie wpływu na ludzi.* Gdańsk: GWP.

Crawford S.D., Couper M.P., Lamias M.J. (2001), *Web Surveys: Perceptions of Burden*. "Social Science Computer Review", 19.

Deutskens E., de Ruyter K., Wetzles M., Oosterveld P. (2004), *Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study.* "Marketing Letters" 15 (1).

Dillman D.A. (2000), *Mail and Internet surveys: The tailored design method.* New York: John Wiley.

Dillman D.A, Bowker D.K. (2001), *The Web questionnaire challenge to survey methodologists,* [in:] U.-D. Reips and M. Bosnjak (Ed.), *Dimensions of Internet science*. Lengerich: Pabst Science Publishing.

Dillman D.A., Tortora R.D., Bowker D. (1998), *Principles for Constructing Web Surveys*. (SESRC Technical Report 98–50). Pullman, Washington. Downloaded on 2 June 2007 from http://www.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf.

Goeritz A.S. (2006), *Incentives in Web studies: Methodological issues and a review*. "International Journal of Internet Science" 1(1).

Goeritz A.S., Goldstein D. G. (2008), *Individual payments as a longer-term incentive in online panels*. "Behavior Research Methods" 40 (4).

Goeritz A.S., Stieger S. (2009), *The impact of the field time on response, retention, and response completeness in list-based Web surveys.* "International Journal of Human Computer Studies", 67.

Groves R.M., Fowler F.L. Jr., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. (2004), *Survey methodology.* Wiley.

Gueguen N. (2002), *Foot-in-the-door technique and computer-mediated communication.* "Computers in Human Behavior", 18.

Gueguen N., Jacob C. (2002a), *Social presence reinforcement and computer-mediated communication: The effect of the solicitor's photography on compliance to a survey request made by e-mail.* "CyberPsychology & Behavior, 5".

Gueguen N., Jacob C. (2002b), *Solicitation by e-mail and solicitor's status: A field study of social influence on the Web.* "CyberPsychology & Behavior" 5.

GUS (2008), *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*

Hamilton M.B. (2003), *Online Survey Response Rates and Times: Background and Guidance for Industry*. SuperSurvey White Paper. Pobrano 20 sierpnia 2009 from http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.htm.

Heerwegh D. (2005), *Effects of personal salutations in e-mail invitations to participate in a Web survey*. "Public Opinion Quarterly", 69 (4).

Heerwegh D., Loosveldt G. (2002), *Web surveys: The effect of controlling survey access using PIN numbers*. "Social Science Computer Review" 20.

Heerwegh D., Loosveldt G. (2003), *An evaluation of the semiautomatic login procedure to control web survey access*. "Social Science Computer Review", 21 (2).

Heerwegh D., Vanhove T., Matthijs K., Loosveldt G. (2005), *The effect of personalization on response rates and data quality in web surveys*. "International Journal of Social Research Methodology: Theory and Practice, 8".

Janiszewska A. (2009), *Działania prowadzone przez Gemius na rzecz podnoszenia response rate w badaniach online*. Research paper. Warsaw: Gemius SA.

Joinson A.N. (2005), *Internet behaviour and the design of virtual methods, [in:] C. Hine (Ed.), Virtual Methods: issues in social research on the Internet*. Oxford: Berg.

Joinson A.N., Reips U.-D. (2007), *Personalized salutation, power of sender and online survey response rates to web-based surveys*. "Computers in Human Behavior" 23.

Joinson A.N., Woodley A., Reips U.-D. (2007), *Personalization, authentication and self-disclosure in self-administered Internet surveys*. "Computers in Human Behavior" 23.

Kaczmirek L. (2009, hard copy), *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*. Cologne: Herbert von Halem Verlag.

Konstan J.A., Rosser B.R.S., Ross M.W., Stanton J., Edwards W.M. (2005), *The story of subject naught: A cautionary but optimistic tale of Internet survey research*. "Journal of Computer Mediated Communication" 10 (2). Downloaded on 28 August 2009 from http://jcmc.indiana.edu/vol10/issue2/kosntan.html.

Lee S. and Vallian R. (2009), *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment*. "Sociological Methods and Research" 37(3).

Little R.J.A. and Rubin D.B. (1987), *Statistical Analysis with Missing Data*. New York: John Wiley and Sons

Mazurek P., Zając J.M., Rakocy K., (2007), *Online digital surveillance: What they know about you depends on what you know*. Speech at the conference General Online Research 07, Leipzig, 26–28.03.2007.

Porter S.R., Whitcomb M.E. (2003a), *The impact of contact type on Web survey response rates*. "Public Opinion Quarterly", 67.

Porter S.R., Whitcomb M.E. (2003b), *The impact of lottery incentives on student survey response rates*. "Research in Higher Education" 44 (4).

Porter S.R., Whitcomb M.E. (2005), *E-mail subject lines and their effect on web survey viewing and response*. "Social Science Computer Review", 23 (3).

Przewłocka J., Janczewski P. (2007), *Jak e-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online*. Speech at the 8th Congress of PTBRiO, Warsaw 18–19.10.2007.

Rakocy K., Zając, J.M., (2006), *Czy wierzysz w słowa te? Wiarygodność odpowiedzi uzyskiwanych w ankietach internetowych*. Poster presented at the National Scientific Conference: Social Aspects of the Internet SAI'06, Warsaw, 8–10.10.2006.

Reips U.-D. (2002), Standards for Internet-based experimenting. "Experimental Psychology", 49 (4).

Rosenbaum P.R. and Rubin D.B. (1984), *Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score*. JASA 79(387).

Schonlau M., van Soest A., Kapteyn A. and Couper M. (2009), *Selection Bias in Web Surveys and the Use of Propensity Scores.* "Sociological Methods and Research" 37(3). Sills S.J., Song Ch. (2002), *Innovations in survey research: An application of Web-based surveys.* "Social Science Computer Review", 20.

Solomon, D.J. (2001), *Conducting web-based surveys.* Practical Assessment, Research & Evaluation, 7 (19).

Sue, V. M. and Ritter, L.A. (2007). *Conducting online surveys.* Sage.

Trouteaud, A.R. (2004), *How do you ask counts: A test of Internet-related components of response rates to a Web-based survey* "Social Science Computer Review", 24.

Tuten T.L., Galesic M., Bosnjak M., *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment***.** "Social Science Computer Review" 22 (3).

Whitcomb M.E., Porter S.R. (2004), *Email contacts: A test of complex graphical designs in survey research* "Social Science Computer Review", 22 (3).

Zając J.M. (2006), *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej,* [in:] M. Sokołowski (ed.), *Oblicza Internetu: Internet w przestrzeni komunikacyjnej XXI wieku.* PWSZ Publishing House in Elbląg.

Zając J.M., Batorski D. (2007), *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby***.** „Psychologia Społeczna", No. 3.