



Mechanisms Shaping an Evaluation System—A Case Study of Poland 1999–2010

Karol Olejniczak

To cite this article: Karol Olejniczak (2013) Mechanisms Shaping an Evaluation System—A Case Study of Poland 1999–2010, *Europe-Asia Studies*, 65:8, 1642-1666, DOI: [10.1080/09668136.2013.832996](https://doi.org/10.1080/09668136.2013.832996)

To link to this article: <https://doi.org/10.1080/09668136.2013.832996>



Published online: 17 Oct 2013.



Submit your article to this journal [↗](#)



Article views: 119



View related articles [↗](#)



Citing articles: 4 [View citing articles](#) [↗](#)

Mechanisms Shaping an Evaluation System—A Case Study of Poland 1999–2010

KAROL OLEJNICZAK

Abstract

A criticism of evaluation practice in regional policy in Central and Eastern Europe (CEE) countries is that it stresses its function in terms of accountability and formal reporting on progress, rather than its role in policy or institutional learning. This article aims to test this critique through the use of ‘system thinking’ and focuses on the case of regional policy in Poland, where evaluation practice has developed dynamically since 2000. It contributes to the literature by developing an analytical framework that sets out the role of evaluation in regional policy learning and identifies key factors and mechanisms that determine the development of an effective evaluation system for regional policy. The analysis draws on all Polish evaluation studies conducted between 1999 and 2010, interviews with civil servants and experts involved in the implementation of cohesion policy programmes, as well as secondary data from earlier studies. The essay identifies key mechanisms and factors that determine the main function of evaluation as a learning tool that produces and utilises knowledge in the decision-making process. These main elements are: growth in funding of the policy; stability of institutions; motivators; and system architecture.

EVALUATION APPEARED IN THE REGIONAL POLICY FIELD IN THE 1930S with the aim of assessing and informing the value and merit of public interventions (Chelimsky 1985, p. 7; Forumier 2005). What differentiated evaluation studies from traditional socio-economic studies was their pragmatic, practice-oriented approach; reports had clear components of judgement and recommendations for improvement.

Evaluation became popular in the 1970s as one of the tools for rationalising, in fact cutting, public expenditures in the era of crisis. Widespread use across the world came with an introduction of performance-based management (Derlien 1990). A rise of evidence-based policies gave a further push in the development of evaluation (Pawson 2006), although the policy scope and institutional embeddedness of the evaluation practices vary across countries (Furubo *et al.* 2002; Dahler-Larsen 2005).

In the EU, the main impulse for spreading evaluation practices, especially in the new member states of Central and Eastern Europe (CEE), has been cohesion policy (Ferry 2009).

The author gratefully acknowledges valuable comments from the journal’s reviewers as well as the support given at the stage of data collection by the staff of the National Evaluation Unit in the Polish Ministry of Regional Development and the Evaluation Section of the Polish Agency for Enterprise Development.

As Bachtler and Wren indicate (2006, p. 143), the combination of the increasing amount of expenditure with the contested decision on its usage resulted in greater pressure for accountability and evidence-delivery. As a result, evaluation became a regulatory requirement as well as the practice of EU funds implementation.

Over the years, evaluation practice has developed a number of approaches and methods that allow it to go beyond accountability and legitimacy and fulfil different functions in public policies. These include improving quality of performance, improving planning,¹ developing ownership,² empowering the marginalised or excluded stakeholders of policies,³ and learning and adapting theories, assumptions and change mechanisms that underpin public interventions.⁴

Modern public management literature and international practice sees evaluation as one of the central elements of both institutional learning (Preskill & Torres 1999; Barrados & Mayne 2003) and policy learning (Dahler-Larsen 2005; Davies *et al.* 2009; Leeuw *et al.* 1994; Van der Knaap 2000). As Michael Quinn Patton points out, evaluation helps to develop public initiatives through continuous adaptation, especially under conditions of complexity (Donaldson *et al.* 2010, p. 25).

In assessing the development of evaluation for learning in the CEE context, specific conditions must be taken into account. On the one hand, the capacity of evaluation systems in CEE has developed significantly in recent years, with substantial credit being given to cohesion policy evaluation requirements. On the other hand, cohesion policy evaluation is perceived to have been tied to the programming cycle rather than the wider organisational and institutional system. Moreover, in the past it has been perceived as operating according to a relatively narrow definition, emphasising accountability and legitimacy, improved planning, and quality and performance, while neglecting learning and capacity building (Batterbury 2006, p. 186). This critique is important in the CEE context. Here, processes of capacity-building for evaluation are starting from a traditionally weak base and evaluation has been perceived as another bureaucratic burden, mechanically executed and often associated with monitoring or control. In the current programming period, these countries are receiving significantly increased levels of cohesion policy funding and implementing a larger number of operational programmes with more ambitious aims and more sophisticated implementation mechanisms. In this context, it is clear that the focus on evaluation for scrutiny, accountability and control remains fundamental. However, this essay argues that there is evidence of more dynamic processes in the development of regional policy evaluation systems in CEE that incorporates scope for increased institutional and policy learning.

In order to explore these issues, the essay applies 'system thinking' to set out the role of evaluation in modern regional policy and to identify the mechanisms responsible for the development and utilisation of evaluation activities in real policy settings. Poland has been chosen as a case study because it provides a prominent example of a CEE country that, in a

¹The main methods used here are logic models and macro-economic modelling.

²This function has been supported mainly by the participatory evaluation approach (Cousins & Whitmore 1998; King 1998).

³This purpose is addressed by empowerment evaluation (Fetterman & Wandersman 2007).

⁴This purpose of evaluation has been a focus of the theory-driven approach (Astbury & Leeuw 2010; Donaldson 2007; Olejniczak 2011) and developmental evaluation (Patton 2010).

short period of time and from an extremely limited base of experience, managed to develop a quite dynamic evaluation system that covers both national and EU-funded programmes (Ferry 2009; Olejniczak *et al.* 2011). The analysis covers 11 years, from 1999 to 2010, of Polish experiences in the field of evaluation studies. It is based on secondary, quantitative data as well as the results of interviews with Polish civil servants and experts on regional policy, conducted within a framework of two earlier studies.⁵

In the next section, the system thinking approach is briefly introduced and its utility for the analysis of evaluation of public policy is explained. Based on this, the essay's analytical framework is also presented. The following section applies the analytical approach to evaluation in regional policy and asks what mechanisms are responsible for the development of evaluation as a learning tool in regional policy. We apply system thinking to address the question of the role of evaluation in the wider system of regional policy. The case of Poland is then introduced, setting out trends and events in Polish evaluation practices between 1999 and 2010. The analytical framework is applied in order to assess these trends, identifying the main structural factors and mechanisms responsible for the dynamism of the Polish evaluation system. Finally, three sets of conclusions are presented: general remarks on the usefulness of system thinking in analysis of regional policy evaluation; the main conclusions drawn from the case study of the Polish evaluation system; and wider lessons for CEE countries on developing their evaluation system.

The system thinking approach

The system thinking approach is an attempt to analyse dynamic, complex and multilayered entities in a holistic way (Anderson & Johnson 1997; Serman 2000). The approach is based in the understanding that systems are nested in an environment (a certain context), which can itself be part of a bigger system (Ostrom 2005). As a result, system thinking can operate on different, interconnected levels of analysis, focusing on systems within systems, or 'zooming out' to consider the larger entity that embraces a sub-system.⁶ Observed effects are understood as a combination of different factors, some of which are delayed over time or coming from a different location. The behaviour of systems is analysed in a dynamic way taking into account a number of factors. Key amongst these factors are 'trends', which are changes in stocks and flows that can be observed over time, forming patterns of behaviour that repeat in time. Also considered are 'events and factors' that influence trends in stocks and flows, and 'mechanisms' that underlie system behaviour, which are responsible for

⁵Parts of the data come from two *ex post* evaluation studies executed for the European Commission–DG Regio (Kozak *et al.* 2009) and for the Polish Ministry of Regional Development–NEU (EGO *s.c.* 2010b). The first study included 34 interviews with two groups of interviewees: heads of agencies and their departments involved in the implementation of programmes co-funded by structural funds; and Polish experts in public administration and regional development. The second study included 88 interviews with three groups. The biggest group of interviewees included key civil servants responsible for the coordination and implementation of operational programmes. Two other groups were Polish experts in public administration and representatives of key NGOs involved in cohesion policy in Poland. Below we will refer to these sources as 'Interviews 2009' and 'Interviews 2010'.

⁶For example, regional policy is a system of its own—with a clear function of 'supporting regional development'. However, it is also a sub-system within a bigger entity of public policies, which has a function of supporting overall national growth and wellbeing.

TABLE 1
BASIC DEFINITIONS USED IN SYSTEM THINKING

| <i>Term</i> | <i>Definition</i> |
|----------------|--|
| System | A group of interacting, interrelated elements that form a complex coherently organised and unified whole. It produces a set of behaviours, called 'function' or 'purpose' ⁴ |
| Stocks | An accumulation of things, materials, information, resources that build up in the system over time |
| Flows | Interactions and actions that mobilise stocks and changes their level over time |
| Feedback loops | Mechanism for the transmission and return of information that, over time, changes the balance in the system |

Note: ⁴ Anderson and Johnson provide a straightforward explanation of differences between system and a simple collection of elements: 'If you can take or add components to a set of elements without affecting its functioning, then it is just a collection, not a system' (1997, p. 3).

Source: Anderson and Johnson (1997) and Meadows (2008).

repeated trends and patterns of behaviour and for causal connections between events.⁷ Definitions of basic concepts used in system thinking are summarised in Table 1.

Systems thinking has relied heavily on quantitative methods. However, the 'soft system' approach uses a mixture of quantitative and qualitative methods and focuses on an exploration of socio-economic systems (including organisations), treated as epistemological constructs rather than given entities (Checkland 1999). The soft system approach uses a three-step analytical procedure (Anderson & Johnson 1997). In step one, key elements of the system are defined (its stocks and flows) and each element is described through a list of variables (indicators). In step two, changes of each indicator over time are analysed. In step three, based on the trends analysis, key problem questions are raised, indicators are compared and possible causal connections are drawn.⁸ This allows the identification of key factors and mechanisms responsible for system performance. In the following sections, the soft system approach is applied first to define the role of evaluation in regional policy and then to identify underlying mechanisms of its development in Poland.

The role of evaluation in regional policy—a system perspective

When describing regional policy as a system, it is important to define the policy's purpose (Table 1). The overall purpose of contemporary regional policy is to facilitate growth and employment by ensuring that regions are able to maximise their development opportunities (OECD 2001, pp. 23–24). In order to implement this purpose, the regional policy system needs to perform three interconnected sets of operations. Each of them is a separate sub-system, with associated resources (human and physical), institutions, procedures and processes. These are presented in Figure 1.

⁷For example, events include elections, civil servants' rotation and delay in EU expenditure in a particular year; visible trends are the increasing rotation of civil servants every four years and the repeated problem of fund absorption. An example of an underlying, causal mechanism is when new elections trigger changes in civil service personnel, and those who are most experienced in EU management leave or are removed by new politicians. That in turn, after a delay, causes a problem with EU fund management and absorption.

⁸To illustrate these connections in a graphical way causal-loop diagrams can be used (Lane 2008).

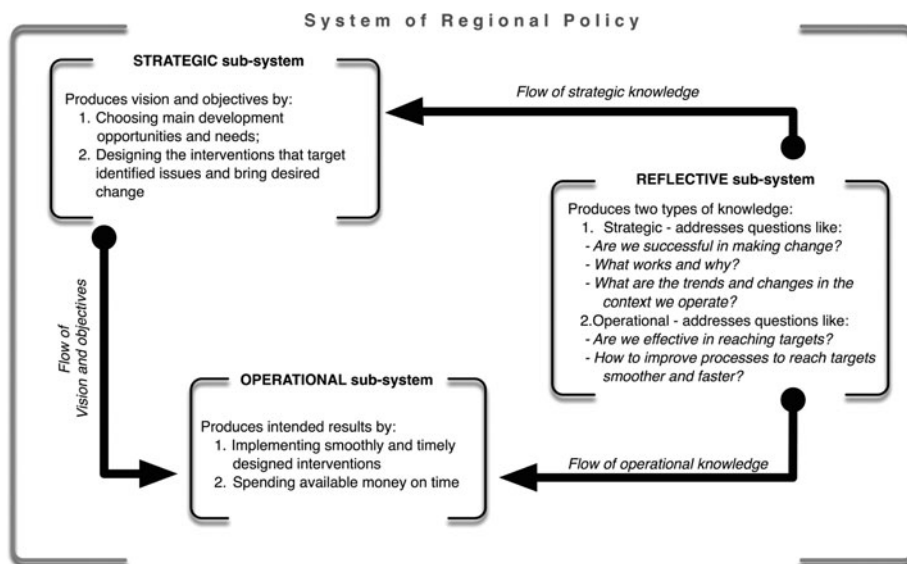


FIGURE 1. REGIONAL POLICY AS A SYSTEM.

As can be seen, the purpose of the strategic system is to deliver a vision of development and translate this vision into objectives.⁹ These decisions, in turn, determine what should be done by the operational system, the aim of which is to deliver what has been planned. Decisions made here refer to how things should be implemented smoothly (that is in a timely manner and according to the legal regulations) within a given framework.¹⁰ The third, reflective system is an engine of learning. Its purpose is to produce knowledge that can be used for the improvement of both strategic and operational activities. Strategic and operational sub-systems send their information needs to it. The reflective system processes information, chooses what should be explored further and then feeds back to the two other sub-systems by providing two substantially different types of knowledge.

The first type of knowledge explains the context in which the operations take place and the real change mechanism responsible for the success or failure of interventions.¹¹ When this type of knowledge is taken on board by the strategic system (key decision makers), it leads to a fundamental change in thinking about the interpretation of needs and opportunities, about policy assumptions, and about the shape of strategic objectives. Double loop learning occurs (Argyris 1977; Preskill & Torres 1999, pp. 44–48).

⁹This in practice means choosing between competing paradigms of regional development and deciding on overall funds allocation. In the field of cohesion policy, it is obtained using three processes: diagnosis and planning (including designation of assisted areas); consultations and negotiations; and coordination.

¹⁰In cohesion policy it covers three fields of activities: communicating and promoting the availability of funding for potential beneficiaries; running the application and selection process of projects; and allocating funding and ensuring financial transfers.

¹¹Strategies and programmes are usually treated as theories that are verified by life. Reflection focuses on proving a causal relation between what was implemented and what actually has changed in reality for the target group, and explaining why and how things worked (Astbury & Leeuw 2010).

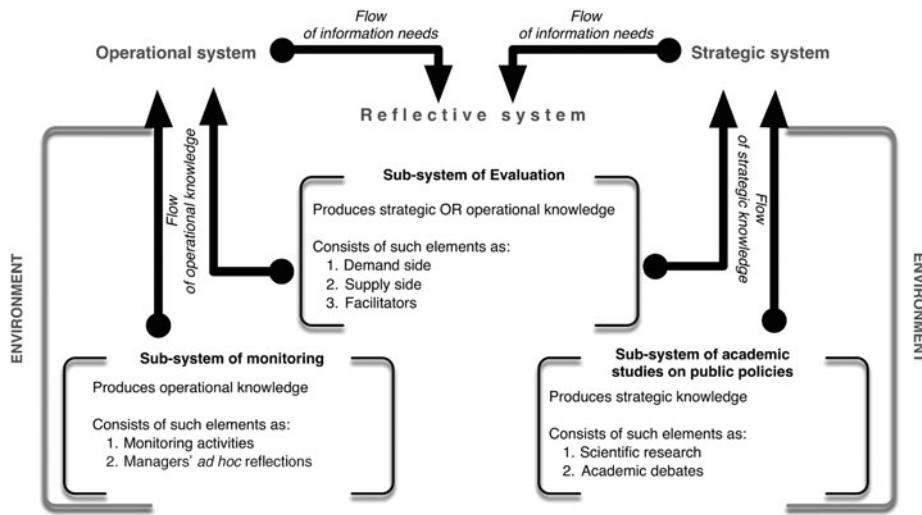


FIGURE 2. REFLECTIVE SYSTEM AND ITS SUB-SYSTEMS.

The second 'type' of knowledge informs about current progress, detects mistakes and brings solutions to short-term implementation problems. When managers of the operational system use it, some adaptation of procedures and everyday implementation practices occurs. Single loop learning takes place and the regional policy system carries on with its present policies and objectives (Argyris 1999; Bennet & Bennet 2004, p. 442).

Focusing more specifically on the details of the reflective systems, it is possible to identify and explore the type of sub-systems that produce knowledge (see Figure 2).

The production of strategic knowledge is a domain of two sub-systems: evaluation research and academic studies. They provide a better understanding of the socio-economic mechanisms responsible for change and development. Operational knowledge can be produced as part of the monitoring activities—discussions in monitoring committees, between policy managers, or *ad hoc* reflection in working groups. It is also produced by evaluation studies oriented to processes. Thus, we can see the function of the evaluation sub-system is caught between everyday managerial information needs and the more strategic, holistic needs of decision makers.

Finally, it is important to define the stocks and flows of the evaluation system itself. As noted, its function is to produce knowledge and feed it to the strategic and operational systems of regional policy. The main product of the evaluation system is the stock of accumulated knowledge, in the form of reports and studies. Its main outcome is knowledge flow in the form of a utilisation of findings from the studies by decision makers and managers. In order to produce knowledge three other stocks are used. First, characteristics of the contracting side of the evaluation process, in this case the public institutions that contract out evaluation studies, are responsible for 'evaluation demand'. Second, 'evaluation supply' is provided through the evaluation expertise in firms, institutes, think tanks and experts that execute evaluation contracts. Third, 'facilitators',

TABLE 2
INDICATORS FOR A REGIONAL POLICY EVALUATION SYSTEM'S STOCK AND FLOWS

| <i>Element</i> | <i>Key issues and indicators</i> |
|----------------------------|--|
| Environment of the system | Timeline of policy cycle Amount of public funds available for regional interventions and its absorption Existence of an institutional strategic planning centre Level of intensity of public debate—measured by proportion between articles that discuss issues of planning, absorption and current implementation of effects of public funds |
| Demand side of the system | Number of institutions that contract out evaluations Level of experience of contracting institutions (experienced = contracted three or more studies) Size of budgets of individual evaluation contracts |
| Supply side of the system | Number of organisations active on evaluation market (active = executing one or more evaluation contracts) Proportions of type of evaluation suppliers (consultancies, research institutes, universities, others) Quality level of the contractors measured by contracting authorities' opinions |
| Facilitators of the system | Presence of evaluation resources (reports' database, libraries, etc.) Number of postgraduate study programmes on evaluations Number of publications on evaluation co-financed by public administration Number of seminars and conferences on evaluation issues |
| Products of the system | Volume of knowledge produced—number of studies and its type Quality level of products in terms of scientific reliability and validity Quality of products in terms of communication (clarity of language, form) |
| Outcomes of the system | Level of evaluation utilisation in strategic and operational systems measured by: self-assessment of civil servants' formal institutional procedures or practices |

which are activities and structures that build an evaluation culture and strengthen its sustainability (for example, training courses, studies, publications, organised databases), also have an important role (Ferry 2009).

According to system thinking, stocks and flows of the system should be translated into relatively straightforward, measurable indicators. A proposed list is presented in Table 2. Since environment is of a crucial nature in system analysis, indicators are added to describe the main characteristics of the environment in which the regional policy system operates.

In summary, we can see that the evaluation system consists of stocks of demand, supply and facilitators that are organised around the function of knowledge production and delivery. Although it is complex, it is only one part that contributes to a wider aim of regional policy decision making.

Analysis of trends in the evaluation system in Poland

The previous section described regional policy as a system of stocks and flows between three main sub-systems: strategic, operational and reflective. Within this, evaluation activities are understood as a key component of systemic learning. In this section, this approach is used to analyse trends in Polish evaluation practices. This creates the basis for

identifying the main mechanisms and factors responsible for the development of evaluation as a learning tool.

The dataset for this analysis starts with 1999 (the beginning of Polish negotiations for accession to the EU) and ends in 2010 (the last year available). It draws on the combined databases of the Polish Ministry of Regional Development, the Office of the Committee for European Integration and the Polish Agency of Enterprise Development. The unit of analysis is evaluations of programmes, sub-programmes and horizontal studies.¹² All of these are concerned with interventions co-financed by the EU (either in the form of pre-accession aid or structural funds). During the analysed period there were no evaluations devoted to purely national interventions, a peculiarity that will be touched on in the final section. The dataset includes all existing records (after an initial data clearing) despite the fact that in the largest database (by the Ministry of Regional Development) only 20% of the studies (of 541 in total) are labelled as regional and territorial development. The argument is that since all of the interventions co-financed by the EU have a spatial impact (for example the European Social Fund (ESF) funded trainings, the cohesion fund's transport projects, interventions targeted at SMEs and innovation), isolating only those labelled by the Ministry would be highly artificial. This approach is in line with the latest moves towards an integrated approach to different strands of structural funds that merge at different levels of EU territory.¹³

Based on the data, as well as findings of earlier studies, the main changes in the environment in which the evaluation system operated can be discussed, including trends in its demand side, supply side, facilitators, products and effects (see [Table 2](#)).

The environment of the evaluation system

Reviewing the timeline of the policy cycle, three stages in the planning and implementation of European funds in Poland can be distinguished: first, the period before Polish accession to the EU which was characterised by EU external aid and, after 2002, formal pre-accession aid for Poland; second, the first programming period (2004–2006)—starting with the accession of Poland to the EU and the ending of all its funding in 2009 (due to the extended financing rule called the $n + 3$ rule); and third, the second programming period (2007–2013), ending, due to the $n + 2$ rule, in 2015. There are clear overlaps between these periods.¹⁴

Negotiations on the new EU budget—a financial cornerstone of the second period—started in 2005 when the first period was still underway. Moreover, programming for each period (that is, the design of programmes and strands of interventions) started usually one year before its formal beginning. In addition, EU funds are spent according to the $n + 2$ rule

¹²There is no database of evaluations of single projects in Poland. Although they are numerous they have little policy relevance. They are usually performed as an ESF obligatory requirement and part of projects' implementation, with little attempt to trace effects and explain projects' theories.

¹³Furthermore, the perception of the Polish ministry also evolves in this direction. Initially, the bulk of EU funding was labelled in Poland as regional policy; later it turned into cohesion policy and presently it is described as development policy. This change in wording mirrors changes in the philosophy regarding EU-funds as part of integrated development-oriented national activities.

¹⁴At the beginning of each programming period annual financial allocations are made to each cohesion policy programme. Programmes are then required to spend funds by the end of the second, or in some cases third, year following the year in which they are allocated. These targets are called the $n + 2$ and $n + 3$ rules.

which means that funds available for a particular programming period can still be contracted out and spent within the next two years after the formal ending of that budgetary period.¹⁵ As a result of these overlaps, there are transition periods when decision makers and managers have to work on two simultaneous tracks—closing the old programming period and starting a new one. It should finally be noted that the first programming period after the Polish accession to the EU was especially intensive since it was shorter than the EU budgetary standard (for old member states it was 2000–2006 + two years).

In terms of funding levels available to Poland during those periods, we can see a very sharp rise from €3.6 billion available in 1999–2003 during the pre-accession period, to €17.1 billion available for the first post-accession period and €85.6 billion available for the second period (2007–2015) (EGO s.c. 2010b, p. 15). Over the decade, Poland became one of the main beneficiaries of EU funds.

Despite substantial development needs, the use of these funds became a serious challenge. An immediate challenge was to design interventions—creating a consistent vision of functional, multi-sector development, describing desired outcomes and translating them into measurable objectives, programmes and projects. This was a completely new way of thinking both for Polish politicians and for the traditional Polish bureaucracy, which was process-oriented and divided into sector silos. Planning for the purpose of EU funds prompted a completely new approach, which for a long time became the only field covered by strategic planning. Over time, strategic planning for EU funds became the key axis of strategic thinking about regional development in Poland.¹⁶ Civil servants involved in implementing EU funds became pioneers, applying results-oriented public management in the face of traditionally bureaucratic procedures (Kozak 2007).

A further challenge was simply to spend money on time and in line with complex EU regulations. This struggle is illustrated by a review of the absorption levels during the first programming period. Member states are obliged to spend EU funds within the required time limit or they are returned to the EU. In 2005, Poland managed to spend only 7% of its available funds. After the rapid application of a special improvement plan, spending in 2006 climbed to 25%. In February 2007, the level jumped to 50% and later reached 75% in 2008. Poland managed to use all of its funds only thanks to an extended period for spending (until 2009) granted by the European Commission (EGO s.c. 2010b, pp. 135–36). Thus, throughout the whole period, the risk of losing resources was high.

Polish accession to the EU and the level of resources available for public investment guaranteed substantial public attention. In theory, the topic of public debates should shape the directions of public policies (in this case, regional policy) and create information needs that could be filled by evaluation studies. This issue is explored by systematic review of articles related to the first programming period published in the four main Polish newspapers.¹⁷ The results are presented in Figure 3.

¹⁵For example, in the financial package for the 2007–2013 period, funds can be contracted until 2013 and spent until 2015. In the 2004–2006 budgetary period, the Commission granted new member states a three year extension of funding (until 2009), instead of two years.

¹⁶It is worth pointing out that it was only in 2009 that the first attempt was made to start a policy discussion on development needs that was not driven by the EU funds perspective (Boni 2009).

¹⁷The review was done on the basis of data from media monitoring contracted by the Ministry of Regional Development. It covered all articles from the 2002 to 2009 period related to EU funds for 2004–2009 and published in four newspapers: *Rzeczpospolita*, *Gazeta Wyborcza*, *Polityka* and *Wprost*. Analysis of the articles' contents was carried out by Evaluation for Government Organizations (EGO s.c. 2010b).

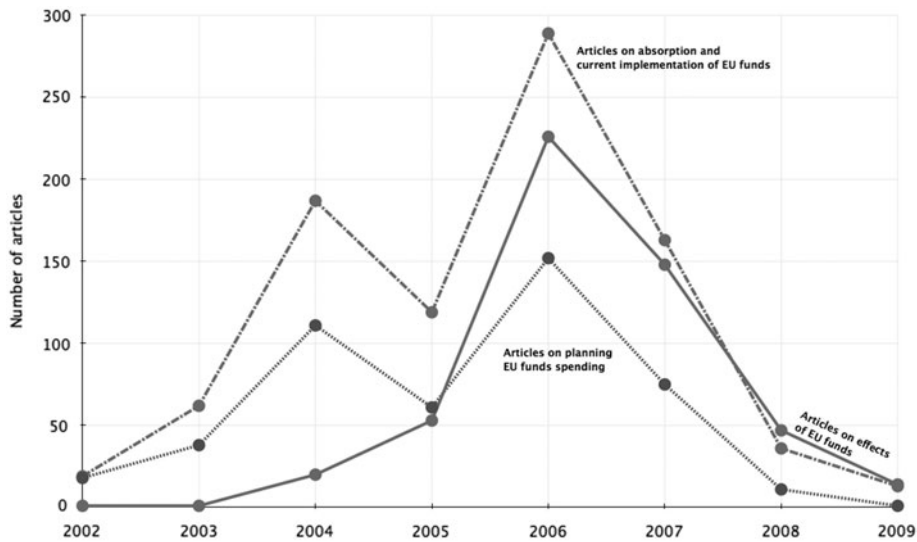


FIGURE 3. ARTICLES ON STRUCTURAL FUNDS PUBLISHED IN FOUR LEADING POLISH NEWSPAPERS 2002–2009.

This analysis raises a number of insights. Articles about planning of regional interventions started to appear in 2004, that is, after the strategic choices were already made (by that time plans were being implemented). Publications focused mainly on the absorption issue, with a dramatic peak in 2006. This was to a large extent a delayed alarm, since improvement measures had already been introduced in 2005. Exploration of real effects was limited mainly to projects, and this disappeared by 2008 when, in fact, the first real effects began to be visible. Thus, we can conclude that the Polish press was rather reactive to the developments in regional policy and focused exclusively on the issue of money absorption, not the rationality of the investments. This means that there was no public debate about real strategic choices (such as the directions and desired effects of EU aid in Polish development) which evaluation studies could feed into.

Demand side

The demand side can be assessed by the number of contracting institutions (and their level of activity per year), their experiences, as well as funds allocated to evaluation studies. This information is presented in [Figures 4 and 5](#).

The core of the demand side consisted of evaluation units. These are small teams of civil servants located in different public institutions (central ministries or marshal offices of regional governments) responsible for contracting, supervising and collecting evaluation studies. Most of the evaluation units usually combined their evaluation responsibilities with monitoring or promotion and less frequently with strategic programming.

[Figure 4](#) shows that the demand side grew from one post in 2000 to 20 posts (seven units) in 2004 at the beginning of Polish EU membership, and up to 57 units and 150 posts in 2010.

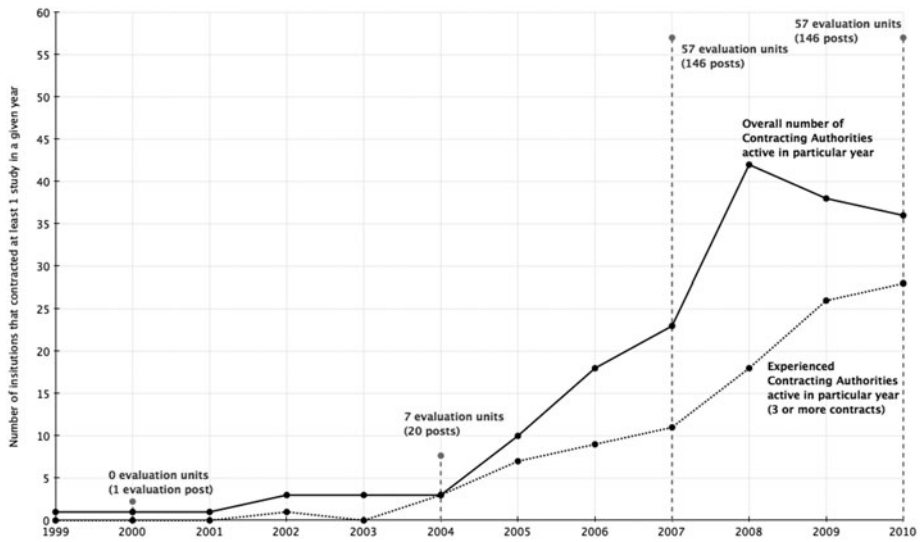


FIGURE 4. CONTRACTING INSTITUTIONS FOR EVALUATION STUDIES.

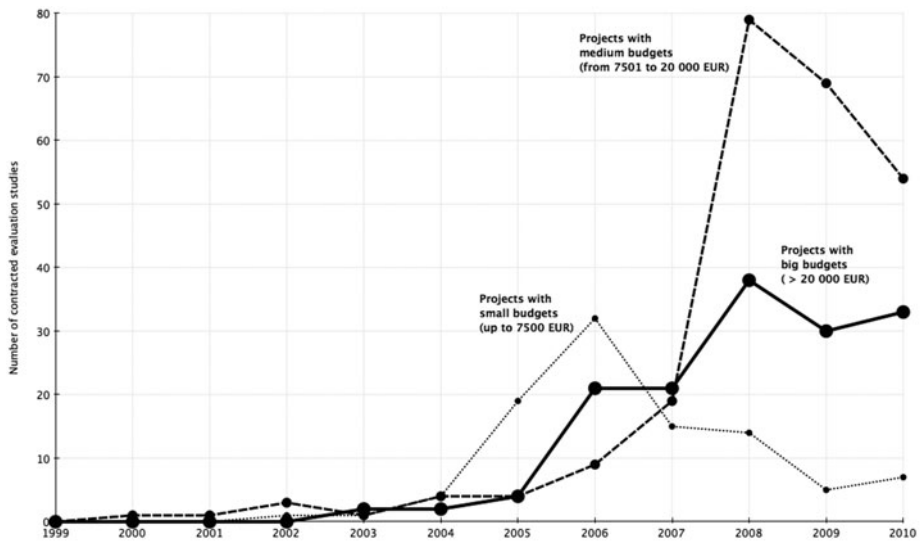


FIGURE 5. EVALUATION CONTRACTS BY THE SIZE OF THEIR BUDGETS.

This dynamic growth was clearly part of the overall development of management and implementation structures for EU Structural Funds in Poland.¹⁸

¹⁸Putting this in context, for the period 2004–2009 the system consisted of 65 institutions and 4,500 civil servants. The system was almost entirely developed within a period of 36 months. In 2007 it extended further because of the new programming period and the introduction of a decentralised implementation system for regional operational programmes.

Developing the experience of evaluation units and their staff was a process of ‘learning by doing’ (Bienias 2009). Thus, it is appropriate to see what the trend changes were in the number of units that had a record of three or more supervised studies. When comparing this with the trend in overall unit activities, we can see a number of peculiarities. There is a clear drop in the years 2002 and 2003. This indicates a transition between the pre-accession period and the first structural funds programming period. During this transition, old units were either dissolved, or taken over by another, or assigned new responsibilities. The new units created at this point (such as the National Evaluation Unit (NEU) in the Ministry of Regional Development) had no experience and no memory of past evaluation expertise collected by the old institutions. Thus, first experiences and initial, very limited know-how were lost during this transition time.

Also of note is the widening gap in the 2006–2008 period between the number of active units and their experience.¹⁹ This indicates a transition between the first and second programming periods. New units were created for new programmes at the central and especially the regional levels.²⁰ However, this second transition was substantially different from the previous transition. Although some units disappeared because their programmes were closing, most experienced actors continued to operate, and they became central to the expanding network of evaluation units.²¹ Moreover, staff from dissolved units joined them. In particular, the NEU became the central node of the system. It was situated in a strategic department of the Ministry of Regional Development and was responsible for the coordination of all activities related to EU Structural Funds. Thus, the memory loss of the system was much smaller and a relatively smooth transition allowed for the accumulation of experience.

The process of narrowing the gap that started from 2009 is another peculiarity.²² This shows that the demand system stabilised and matured. No new units were created, existing entities continued to operate and they contracted out fewer, more targeted studies.

A simple but straightforward indicator for the experience of evaluation units can be the quality of the terms of reference (ToR) of studies they contract out. This issue has not been explored in a regular manner across the years. However, analysis of interviews, experts’ observations and single studies reveal a number of salient points (Olejniczak *et al.* 2008; EGO *s.c.* 2010a; ZBN PTS 2010). Over time, the ability to form a limited number of precise, focused evaluation questions increased significantly.²³ Contracting authorities also learned to better manage the framing requirements of particular studies that refer to evaluation questions, scope, methods and resources (time and money).²⁴ Although ToR are constructed in a more rational and logical way, they often remain overambitious (in terms of questions concerning scope,

¹⁹In 2005, 70% of active units were experienced. In the following three years these were: 50%, 48% and 43%, respectively.

²⁰Units were created in all 16 regional marshal offices. They focused on evaluation of the regional operational programmes in their regions.

²¹These are notably: the NEU (Krajowa Jednostka Oceny), the European Social Fund Unit (both located in the Ministry of Regional Development) and the Evaluation Unit in the Polish Agency for Enterprise Development.

²²In 2010 only 22% of operating units had no experience in constructing or supervising evaluations.

²³The most extreme examples from the pre-accession period consisted of over 100 evaluation questions raised in a single evaluation study. The standard for the first period were ToR with up to 20 research questions. Currently ToR prepared by the NEU usually include between five and seven key questions.

²⁴For an extensive, empirical discussion on the importance of balance between these four elements see Ferry and Olejniczak (2008, pp. 63–66).

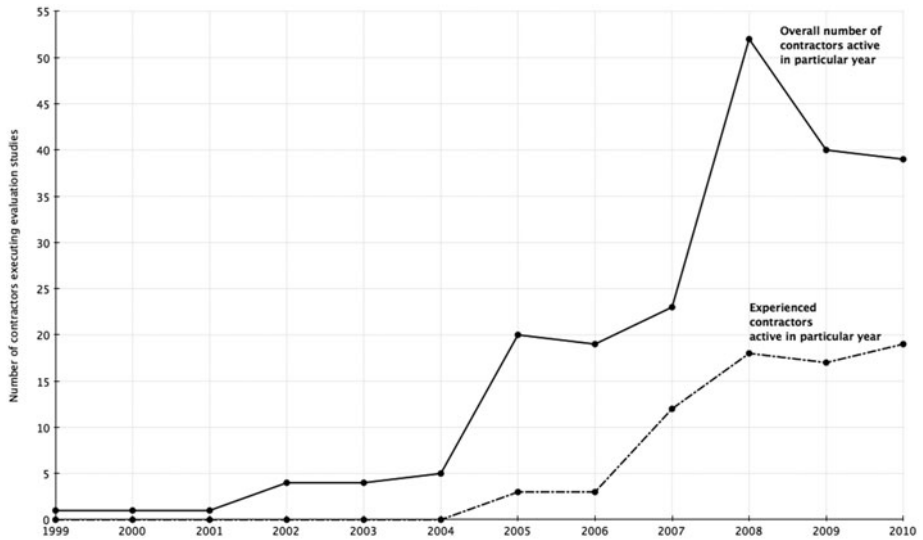


FIGURE 6. CONTRACTORS CONDUCTING EVALUATION STUDIES IN POLAND.

time and money assigned), especially in units with less experience. Price continues to be a key selection criterion since on average it is responsible for 60% of the final scoring of bids for evaluation studies (PTE 2011). Often, civil servants do not judge the quality of the offer for an evaluation study at all and select an offer exclusively on the basis of low price. Only the most experienced units tend to use price as 40% of the final selection grade.

The pattern of the number of evaluation studies and the level of resources assigned to studies is presented in Figure 5. In general, the number of contracts and spending on evaluation grew from single studies per year at the beginning of the millennium to over 100 studies per year in the last three years. Annual spending grew from €100,000 in 2004 to almost €2.5 million in 2010.

Looking at trends in the number of studies and their budget size, the following points should be noted. Small and medium-sized budget studies clearly dominated. Small budgets obviously determined the scope of studies. Moreover, the sizes of these budgets were extremely small when compared to the sizes of studies in other countries. As a result, they were not attractive for international research and consultancies involved in evaluation work. Only during the last three years has a wave of *ex post* horizontal studies from the 2004–2006 period increased the proportion of large budgets to around 30% of the overall number of contracted studies. Thus, there continues to be a large group of studies that are relatively small in financial terms, and as a result, narrow in their scope.

Supply side

The changing situation on the supply side is presented in Figure 6. Concerning the overall number of active contractors and their experience, some clear trends are apparent. From 2001 to 2006, there was virtually no accumulation of experience. This was because different companies, with no intention of specialising in evaluation, had executed contracts. The gap

between the number of active companies and their experience started to narrow in 2007. However, the dynamic extension of the market into the regional level drew regional and local firms into the field of evaluation. This process widened the gap again. During 2008–2010, it narrowed again, although still, quite surprisingly, every second contractor had no previous experience in executing evaluation studies. They tended to use their broader experience in social research or surveys.

From a review of other assessments of the supply side in the Polish evaluation system, the following findings emerge (EGO s.c. 2010a; PTE 2011). Over 70% of companies involved in studies are Polish owned. Large consultancies are almost absent in the Polish evaluation market and small and medium-size companies dominate. Sometimes these are formal branches of international firms, however they are fully staffed by Polish researchers. Research institutes and universities appear occasionally as contractors, although in recent years they have become more active, especially in *ex post* evaluations. University experts tend to work more as experts in studies coordinated by firms. Research firms (specialised in survey studies) have become more active in the evaluation domain in the last year. Interestingly, they usually form consortiums with other entities. Finally, companies tend to specialise in evaluation slowly, although it is still very early and their contract portfolios are still rather driven by the number of contracts available.

In summary, the supply side is still at an early stage of development, highly fragmented, with lots of inexperienced and contract-driven rather than topic-driven players. Not surprisingly, the quality level of the contractors is still quite low in the opinion of representatives of the contracting authorities.²⁵

Facilitators

The next element of an evaluation system is facilitators. This relates to all the activities that support and promote evaluation culture and practices (publications, conferences and regular training programmes). The main trends in this element of the system in Poland are presented in Figure 7.

Figure 7 reveals a clear difference in volume between the pre-accession and post-2004 periods. The real beginning was clearly the year Poland entered the EU. However, what is not visible is the fact that some activities in the first programming period were co-financed with pre-accession money. The next jump in quantity takes place in 2008, when regional authorities enter the field of evaluation. What is interesting is the growing competition between regions in terms of publications and conferences. A clear regularity of events is also apparent. Most conferences are annual; publications are usually made as part of a series, and training is in the form of one year, post-diploma courses.

In the field of training, we can see an interesting, long-term market response to the needs of the Polish administration. Instead of short training (popular in the pre-accession period), universities offer one-year post-diploma programmes for practitioners. All of these programmes, with one exception, are run on the market basis, and they educate both civil servants and evaluators.²⁶

²⁵Interviews 2010.

²⁶The only exception is the Academy of Evaluation organised by EUROREG-University of Warsaw and the Ministry of Regional Development. This one year study programme is co-financed by the EU European Regional Development Fund and it is designed exclusively for civil servants.

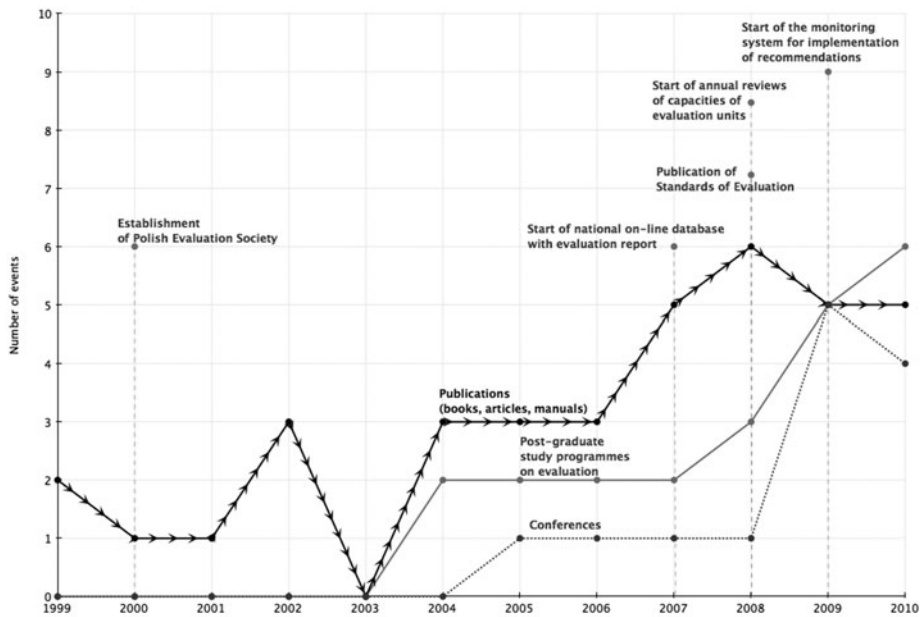


FIGURE 7. EVENTS FACILITATING AN EVALUATION CULTURE IN POLAND.

Crucially, there are clear, stable drivers behind the development of facilitators—these are the NEUs, the unit in the Polish Agency of Enterprise Development and, recently, the revived Polish Evaluation Society. These drivers cooperate in running annual conferences and seminars, sponsor and edit the main publication series (often on advanced evaluation methodologies), and introduce systemic improvements. Thus, there is a stable, growing system of evaluation promotion and development. The system has been based firmly on the dynamism and ideas of the Polish administration, supported by some academic institutions.

Products

Figure 8 shows a growing number of evaluation studies. This indicates the volume of knowledge produced. So far, there have been over 540 reports published. In terms of the evaluation type, 14% of them were *ex ante*, 76% on-going and 10% were *ex post* studies.

Concerning the type of knowledge produced by these evaluations, operational knowledge dominates (62% of all studies) although this proportion is not as large as may have been expected. The one point of time when the strategic focus clearly outweighed operational issues was in 2006—the time of *ex ante* studies as part of preparations for the second programming period. Beyond sheer volume, it is also important to assess the quality of these evaluations. Surprisingly, so far, there has been little discussion and virtually no systemic assessment of the scientific reliability and validity of Polish evaluation studies.

The picture that emerges from interviews with civil servants and an unstructured review of reports produced highlights the fact that recent studies of central programmes have clearer structures and more conclusive results. They go beyond reporting of the programme process or just describing problems and move toward explaining causal mechanisms and

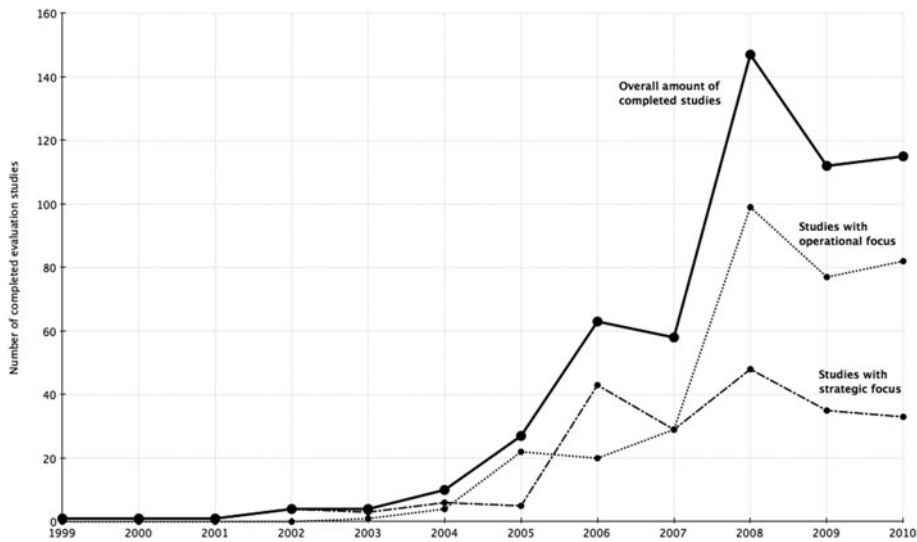


FIGURE 8. COMPLETED EVALUATION STUDIES IN POLAND.

their effects. This can be attributed both to more demanding civil servants, but also to the nature of the recent wave of more horizontal, *ex post* studies. The clear progress being made in methods is also apparent. Simple desk research and surveys with beneficiaries are accompanied by in-depth scientific literature reviews, system analysis and use of longitudinal secondary data. Despite these advances, there have been no standards developed in terms of defining ‘good research quality’ in evaluation study. For example, the trend in methodological innovations has been counterbalanced by the growing practice of replication. Contractors often copy solutions that have been accepted already by other contracting authorities. They do it without adaptation for the different context and new projects (this is especially visible at the regional level). In theory, the application of the same methods should allow for comparisons; however, in practice, these solutions are poorly executed and have low validity.

The quality of the studies can be measured not only by the quality of their method (true test) but also by the quality of their communication (utility test) (Weiss & Bucuvalas 1980; Torres *et al.* 2005). A recent meta-evaluation of the language used in evaluations produced some unexpected findings (Broda *et al.* 2011).²⁷ The language of Polish evaluation reports is, on average, more complex than the language of scientific texts. It requires from a reader 17 years of education (PhD level). Furthermore, executive summaries of reports are even less understandable than the main body of texts. The authors of this study suggest that the difficulty of evaluation language has risen in recent years, which can be the result of the use of more sophisticated methods. Ambiguity of language can also be caused by the vagueness of the research questions and research structures. The key conclusion is that the language of

²⁷The clarity of language has been measured by the number of indicators of which a ‘Fog index’ is the most interesting one. It measures the length of phrases and their structure—the ratio between nouns and verbs. The shorter the phrases and the more verbs there are, the less foggy the message is and more understandable to the average reader. This technique is based on the ‘Plain English standard’.

evaluation studies obscures the message that is put forward to the public and decision makers.

Outcomes

Information on the outcomes of the evaluation system in Poland is still very limited. Unlike some other elements, there has been no systematic record of the use of evaluation reports. Only in 2009 did the NEU launch a system for monitoring the implementation of evaluation recommendations. However, the system is in its early stages, and its first critical feedback indicates that too much attention was placed on the mechanical, formal use of technical recommendations rather than real utilisation of knowledge.

This situation to a large extent confirms that the focus of the system is on the production of information (as an easily measurable process) and, to a lesser extent, utilisation of knowledge. Some ideas about the use of knowledge are revealed by the results of interviews conducted in 2010.²⁸ Fourteen senior civil servants involved in the planning and implementation of policy assessed on the school scale (from 2, the lowest, to 5, the highest) the utility of reports' findings in decision making for the first programming period. According to them, operational use scored 3.5 while strategic use scored 3.8. In general, interviewees were critical of the extent to which knowledge was utilised. Surprisingly, the reports were used more for strategic decision making despite the fact that most of the produced knowledge was operational. A recent study also reveals growing concern among civil servants about the ability to synthesise information produced, and as a result, the real utility of the growing population of studies (Olejniczak & Ledzion 2013).

Key mechanisms and factors

Following this assessment of the evolution of the evaluation system in Poland, this section aims to explain the patterns in outcomes of the system (concerning knowledge production and utilisation). From the analysis of trends, four key questions arise. First, what is responsible for the dynamic trend in knowledge production? Second, what causes this majority of produced knowledge to be operational in nature? Third, why is the rise in quantity of the studies not accompanied by an equally dynamic rise in the quality of produced knowledge? Fourth, why is the usage of such a growing pool of knowledge relatively limited in the decision-making process of regional policy? Concerning the first question, knowledge production has been boosted by the combination of factors. The mechanism was clearly triggered by the EU, including the growth in EU funding and requirements on the timing of evaluation. This led to increasing complexity in the regional policy system (that is, concerning the number of programmes and implementing institutions), which in turn resulted in more evaluation contracts performed for new programmes and their supervising institutions. This was very clear in 2005 (following the start of the new system in 2004) and in 2008 (after the introduction of a formally new decentralised system with a new role for Polish regions in 2007). Competition between newly created institutions was another driver. Since evaluation units wanted to justify their existence and fulfil their objectives, they contracted studies. As the main indicator of success

²⁸See footnote 5.

was a high level of activity, units tend to contract more studies. The same competition mechanism was also responsible for growth in facilitators such as conferences and publications—especially those provided by regional offices. The policy cycle was another mechanism responsible for shifts in the number of evaluations. The majority of studies are usually launched at the beginning of a policy cycle, when the implementation system is being designed or when arguments are collected for negotiations of the new budgeting cycle. The number is raised again at certain moments of implementation—when problems or challenges occurred, as illustrated by the 2006 and 2008 peaks.

Concerning the second question, the reasons behind the domination of operational over strategic knowledge were as follows. The mechanism was fuelled by the instability of institutions. The lack of strategic planning traditions in the Polish administration, combined with the instability of the Polish ministries, resulted in the absence of an administrative centre for strategic planning that would allow strategic continuity between EU programming periods. Since the institutions were not sure of their future, they were not interested in planning beyond one policy cycle. This resulted in a focus on day-to-day operations and an information need on how to spend money smoothly. This orientation was reinforced by time pressures and the direction of public disputes. The media focused mostly on current spending and the risks entailed in delays, putting pressure on managers and decision makers to focus exclusively on absorption. Between 2004 and 2006, problems occurred quite often and time pressure was high (with the shorter programming period) and the main interest was on operational know-how and quick, temporary solutions. Evaluation was one of the information sources used to find solutions or to justify managers' decisions.

This mechanism started to work the other way around once some key actors emerged in the system. In the second programming period, the Ministry of Regional Development continued to be the main hub of regional policy. The policy itself grew and slowly turned into a broader development policy. This created a need for more strategic, comprehensive views on planning and its effects, which led in turn to a wave of *ex post* evaluations. In this context, it is important to note that the synthesis of *ex post* evaluation results for 2004–2006 has been used on the policy level during high-level discussions on the directions of the cohesion policy after 2017 (Bienias 2011). Finally, it is important to reiterate that institutional stability corresponds with the strength and scope of strategic knowledge. This knowledge requires more horizontal, long-term, larger studies and data, observation and experience accumulation over time. Clearly, this requires a stable institution that can supervise processes over time and consume knowledge at a more general level.

Concerning the third question, the phenomenon of a gap between the number and quality of the studies can be explained by several factors. A key factor is the time delay between performing evaluations and learning from this process. Put simply, accumulation of experience needs time. It also needs institutional stability over time. In Poland, the number of contracts available grew rapidly, and new players appeared quickly (both on contracting and contractors' sides). This was combined with institutional shifts between programming periods that led to discontinuity of experience on both the supply and demand sides (especially during the 2002–2004 transition there was a loss of knowledge). The stability that the system gained over time allowed contracting authorities to gain experience and become more demanding when it came to the quality of products provided by consultancies. This quality pressure started only when clear institutional hubs of the system stabilised in the NEUs and a unit in the Polish Agency for Enterprise Development (PAED). They set

standards and continued regular activities such as conferences, publications and studies that raised the quality threshold.

Another factor in the gap between the number and quality of the studies was the competitive impulse on the demand and supply sides of the evaluation system. On the demand side, the main motivator for evaluation units was the number, rather than the quality, of contracted evaluations. As most of the units were beginners, they tended to base their ideas for contracting studies on other units. Waves of similar study topics were often apparent. Quality was not monitored and never clearly demanded from evaluation units until 2009 (with the first ranking for performance of evaluation units). On the supply side, competition between companies did push them to experiment with methods and approaches. However, this impulse was undermined by the selection criteria produced by evaluation units in their offers (where the cheapest offer wins), the size of budgets and the similarity in evaluation topics. Small budgets and repeating topics discourage innovation and encourage imitation.

Finally under this heading, the strong relationship between the evolution of evaluation methodologies and the quality, accessibility and user-friendliness of communication must be taken into account. The use of more advanced, sophisticated methods often resulted in the use of more technical jargon and difficulties in communication. Furthermore, until recently, there was no consideration given to language quality either on the demand or the supply side—the level of communication was never part of the assessment criteria for authorities contracting evaluations.

Turning to the fourth question, the issue of relatively limited use of the evaluation results can be explained by a clear mismatch between information needs and provided results. This was due, in part, to the characteristics of operational knowledge. Although there was pressure to spend EU money quickly, the ways to do it were not of interest to the general public or the media. Evaluation results, therefore, represented a source only for a relatively narrow group of policy managers. Moreover, detailed operational issues of the myriad of interventions caused information inflation that was difficult to synthesise and use (even for operational purposes) in future programmes. Its technical, specialised character was also responsible for its quick deterioration and loss of relevance. Another barrier to the use of evaluations was their use of the language of policy managers, including technical jargon, deeply rooted in the terminology of particular interventions. Furthermore, studies were often very descriptive and inconclusive which resulted in their limited utility. The timing of the results was a further barrier to their effective utilisation. In the shortened programming period 2004–2006, the operational knowledge from evaluation arrived too late for managers. Moreover, *ex ante* studies in 2006 for the 2007–2013 period addressed programme designs that had already changed during the process of consultations on the future programmes. Thus, in most cases, operational knowledge was not delivered on time from the perspective of decision makers. Problems with the ‘system architecture’ compounded the shortcomings of the evaluations themselves. Evaluation units were quite loosely connected to units in strategic and operational systems. In other words, there was limited contact between evaluation units and the potential users of their knowledge. As a result, questions raised in evaluation studies had limited relevance to the information needs of strategic and operational managers. Moreover, evaluation units had too weak a position in the system to set the topics for system-wide discussions.

Finally, the main indicator of evaluation units’ performance was fulfilment of an evaluation plan and the number of completed studies. The extent of utilisation of studies

(and even dissemination) was not coded in the system of performance indicators. This created unintended consequences, which have been well described in public management literature (Smith 1995). Such consequences include concentration only on activities included in annual reviews, which Smith terms 'tunnel vision'. The result is a focus on reports completed and insufficient attention being given to issues that are more qualitative in nature. 'Suboptimisation' is another consequence. In this case, civil servants pursue their own narrow objectives, focussing on completed studies at the expense of the overall outcome of the reflective function.

Conclusions

This review of the evolution of the evaluation system in Poland produces three sets of conclusions: on the usefulness of system thinking in analysis of regional policy; on the key findings coming from the case study of the Polish evaluation system; and, building on these, wider lessons for CEE countries on developing evaluation systems.

The analysis portrayed evaluation practices as part of a wider system of regional policy decision making. This wider system, in order to be effective, needs to perform three interconnected sets of operations. First, it must choose development alternatives and define strategic objectives (strategic function). Second, the system has to implement them smoothly, in a timely way, and in line with regulations (operational function). Third, it has to learn from success and failures on what interventions work, for whom and under what conditions (reflective function).

Evaluation practice falls into the reflective function. The role of evaluation studies is to produce knowledge based on evidence from performed interventions and to feed this knowledge into key strategic discussion or everyday managerial practices. The application of system thinking provides us with flexibility and clarity in moving between different aspects and levels of regional policy implementation. On the one hand we can go beyond the myriad of managerial activities and see the bigger picture—three main sets of processes and the main function of policy. On the other hand, we can go into details of the processes of a particular sub-system (as was done in this article for the reflective part) without losing sight of the overall purpose of the policy.

This approach has been especially helpful in overcoming the dominant perspective on the effectiveness of cohesion policy. Both public opinion and administration of CEE countries tend to focus on the level of money spent as a key indicator of successful policy. Instead, this approach perceives spending as just one of the output measures that should contribute to the overall outcome of the regional policy system—facilitating growth, employment and development opportunities.

System thinking also makes visible the dynamic interrelations between system elements. As a result, it is possible to assess the intensity and quality of feedback mechanisms. This essay shows how the limited flow of information needs from strategic and operational systems to reflective systems leads to the limited utility of evaluation activities.

To sum up, there are promising perspectives in applying system thinking to analyses of regional policy implementation. Future studies could focus more on the relations and sometimes conflicts between functions. As Hummelbrunner (2011, pp. 262–63) notes, stakeholders involved in different aspects of policy implementation often have very diverse pictures of interventions and associated processes. Another possible topic to explore is

feedbacks in the system and their relationship with an environment—what impulses shape strategic visions, what types of information feed into the strategic sub-system, how and in what form strategic objectives trickle down into the operational system. Those types of studies would link the regional policy field to wider debate on policy making (Stone 2011), evidence-based policies and a use of knowledge in public policies (Nutley *et al.* 2007).

In terms of findings from the Polish case and wider lessons for CEE countries, a basic conclusion is that there are basic challenges to the development of evaluation for learning common to all evaluation systems, including Poland. The analysis shows that the natural tendency of the evaluation system is to focus on more technical, process-oriented knowledge. To put it simply, the question of how to absorb resources quickly and smoothly dominates. Moreover, the common combination of mechanisms and factors tends to draw actors' activities towards producing knowledge and increasing its quantity, rather than using knowledge and increasing its quality. These challenges are exacerbated in the CEE context where traditional weaknesses in public administration, lack of evaluation skills and experience in the public and private sector and the pressure to develop evaluation capacity while spending unprecedented levels of EU funds has created significant barriers. Thus, for Poland and all CEE countries who are trying to develop evaluation systems, key challenges are to devote more attention to holistic reflection on determinants of successful socio-economic change and switch the focus from the production of knowledge to its use. This means the use of evaluation as a tool not only for managerial, day-to-day improvements, but also as a source for public discussion on value, merit and directions of regional policy.

Nevertheless, evaluation capacity, including capacity for learning, has developed dynamically 'from scratch' over the past decade. The analysis identified four groups of mechanisms and factors that have facilitated the development of the system and utilisation of evaluation as a learning tool. The first mechanism is simply growth in EU funding. The volume of available resources, the number of interventions, the number of institutions involved in their implementation and the level of public interest created a critical mass for development of an evaluation system. In the Polish case, the evaluation system was clearly a spin-off of the overall EU regional policy system. Due to its size and importance it became visible and started to trickle down into other domestic policies. Thus, cohesion policy in CEE is a strong impulse for introducing modern public management standards and practices of evaluation into public administration.

The second mechanism is stability of institutions. This allows the accumulation of experience on the demand side (that is institutions that contract out studies), it increases the methodological quality of studies and, over time, the institutional capacity for strategic thinking. As the Polish case shows, the initial lack of institutional stability caused a loss of know-how in the first years of the evaluation system. On the other hand, the stable institutional structure after 2004 clearly facilitated the development of regular practices of execution and promotion of evaluation activities. Therefore, the clear lesson for other countries is to allow incremental organisational learning by assuring continuity in institutional structures and responsibilities. One has to remember that what makes institutions are people. In the Polish case, the stability and motivation of staff employed in key evaluation units was a clear factor of system development.

The third mechanism is incentives. The Polish case revealed the importance of competition, level of prices, selection criteria and success indicators. The competition mechanism is simply what motivates institutions on the demand side (contracting authorities)

and firms and research institutes (contractors) on the supply side. In the Polish evaluation system, the competition mechanism was responsible for increasing the quantity of contracted evaluation studies, the development of methods and the establishment of regular practices in the field of evaluation facilitators such as conferences and publications. It is worth pointing out that initially the relatively low prices offered by Polish contracting authorities (in comparison to contracts in older EU member states) discouraged international consultancies and allowed local expertise to grow.

From a negative perspective, other incentives also had a strong impact on the quality of delivered products and caused low utilisation of results, most notably the selection criteria used by contracting authorities in the development of evaluation projects and the success indicator used by evaluation units. In relation to selection criteria, in Poland, those evaluation units that favoured design and methodology criteria pushed the development in methods and evaluation approaches. However, the overall tendency to use price as a key criterion hampered innovation in methods and fuelled imitation practices. In terms of the success indicators, in Poland, this was usually the number of completed evaluations and fulfilment of annual evaluation plans. This resulted in focus on the quantity but not on the quality and use of reports. This experience offers important lessons for other CEE countries. It is apparent that the selection criteria of evaluation contracts are a crucial form of micro-regulation that shapes the quality of whole market trends. Assessment of the quality of the contract (clarity of research question, consistency of research design and novelty of proposed methods) has to be the fundamental element in the public procurement for evaluation studies. The Polish experience reflects and reinforces the basic premise of performance management: 'what gets measured, gets done' (Osborne & Gaebler 1992, p. 146). Choosing volume of studies as a key indicator of system performance focuses attention on the basic production of reports. This indicator has to be balanced by other measures that inform about knowledge use and dissemination.

The final mechanism that shapes the system of evaluation is its architecture. This has two aspects. One concerns the network of evaluation units. In the Polish case, their connection with each other as well as with their international counterparts, such as the European Commission evaluation units, was an important vehicle for knowledge transfer, exchange of know-how and problem solving. A clear lesson for other countries is the importance of those types of communities of practice for the development of system capacities.

The other aspect of architecture was the loose connection and relatively isolated position of evaluation units in relation to strategic and operational functions. This resulted in a limited use of produced evaluations. Evaluation units have limited understanding of information needs in strategic and operational sub-systems. Furthermore, their isolated position gave them little strength to initiate meaningful policy discussions or conduct an effective dissemination of the results. The clear lesson for CEE coming from this case is to think about evaluation units not as mere supervisors of knowledge production but as knowledge brokers. These are intermediates between the world of science and consulting, the world of political appointees and public managers from strategic or operational functions. They have to make evaluation studies meaningful for users. On the one hand, knowledge brokers monitor the information needs of public managers and politicians and then translate them into evaluation orders. On the other hand, they push evaluators and social scientists to communicate findings in understandable and practical language. They are also responsible for feeding the results into decision making and the management process.

Thus, the civil servants who are knowledge brokers have to keep close relationships with their counterparts in the strategic and operational sub-systems.

The need for making evaluation meaningful leads us back to the core of simple evaluation craftsmanship: asking questions that matter for potential users and answering them in a clear, reliable manner. As the Polish case shows, this is a continuous challenge that has to be tackled at the level of a whole country system as well as in single evaluation studies.

University of Warsaw, Centre for European Regional and Local Studies (EUROREG).

References

- Anderson, V. & Johnson, L. (1997) *Systems Thinking Basics: From Concepts to Causal Loops* (Waltham, MA, Pegasus Communications).
- Argyris, C. (1977) 'Double-Loop Learning in Organizations', *Harvard Business Review*, 55, 5.
- Argyris, C. (1999) *On Organizational Learning*, 2nd edn (Oxford, Wiley-Blackwell).
- Astbury, B. & Leeuw, F. L. (2010) 'Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation', *American Journal of Evaluation*, 31, 3.
- Bachtler, J. (2001) 'Quod erat demonstrandum? Ewaluacja polityki regionalnej', *Studia Regionalne i Lokalne*, 4, 7.
- Bachtler, J. & Wren, C. (2006) 'Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges', *Regional Studies*, 40, 2.
- Barrados, M. & Mayne, J. (2003) 'Can Public Sector Organisations Learn?', *OECD Journal on Budgeting*, 3, 3.
- Batterbury, S. (2006) 'Principles and Purposes of European Union Cohesion Policy Evaluation', *Regional Studies*, 40, 2.
- Bennet, A. & Bennet, D. (2004) 'The Partnership between Organizational Learning and Knowledge Management', in Holsapple, C. (ed.) *Handbook on Knowledge Management 1: Knowledge Matters* (Heidelberg, Springer).
- Bienias, S. (2009) 'Evaluation of Structural Instruments in Poland', paper for the conference *Evaluation of EU Structural Funds: Reinforcing Quality and Utilization*, Ministry of Economics, Vilnius, Lithuania, 26–27 March 2009.
- Bienias, S. (2011) 'Evidence-based Cohesion Policy and its Role in Achieving Europe 2020 Objectives. Conference Background Report', paper to the *Opening Conference of the Polish Presidency in the European Union*, Gdańsk, 7–8 July.
- Boni, M. (ed.) (2009) *Polska 2030—Wyzwania rozwojowe* (Warsaw, Kancelaria Prezesa Rady Ministrów).
- Broda, B., Maziarz, M., Piekot, T., Poprawa, M., Radziszewski, A. & Zarzeczny, G. (2011) *Język raportów ewaluacyjnych* (Warsaw, Ministerstwo Rozwoju Regionalnego).
- Checkland, P. (1999) *Systems Thinking, Systems Practice: Includes a 30-Year Retrospective* (New York, Wiley).
- Chelimsky, E. (1985) *Program Evaluation: Patterns and Directions* (Washington, DC, American Society for Public Administration).
- Cousins, B. J. & Whitmore, E. (1998) 'Framing Participatory Evaluation', *New Directions for Evaluation*, 80, Winter.
- Dahler-Larsen, P. (2005) 'Evaluation in Public Management', in Ferlie, E., Lynn, L. E. & Pollitt, C. (eds) *The Oxford Handbook of Public Management* (Oxford, Oxford University Press).
- Davies, H., Nutley, S. & Smith, P. (eds) (2009) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol, Policy Press).
- Derlien, H. U. (1990) 'Genesis and Structure of Evaluation Efforts in Comparative Perspective', in Rist, R. C. (ed.) *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations* (London, Transaction Publishers).
- Donaldson, S., Patton, M., Fetterman, D. & Scriven, M. (2010) 'The 2009 Claremont Debates: The Promise and Pitfalls of Utilization-Focused and Empowerment Evaluation', *Journal of Multidisciplinary Evaluation*, 6, 13.
- Donaldson, S. I. (2007) *Program Theory-Driven Evaluation Science: Strategies and Applications* (New York, Lawrence Erlbaum).
- EGO s.c. (2010a) *Badanie potencjału oraz stanu dotychczasowych prac w ramach jednostek ewaluacyjnych funkcjonujących w Instytucjach Zarządzających oraz Instytucjach Pośredniczących I i II stopnia w ramach NPR 2004–2006 i NSRO 2007–2013* (Warsaw, Ministerstwo Rozwoju Regionalnego).

- EGO s.c. (2010b) *Ocena systemu realizacji polityki spójności w Polsce w ramach perspektywy 2004–2006* (Warsaw, Ministerstwo Rozwoju Regionalnego—Krajowa Jednostka Oceny).
- Ferry, M. (2009) 'Cohesion Policy Evaluation Systems in the Visegrad Member States: An Overview', in Bienias, S. & Lewandowska, I. (eds) *Evaluation Systems in the Visegrad Member States* (Warsaw, Ministry of Regional Development).
- Ferry, M. & Olejniczak, K. (2008) *The Use of Evaluation in the Management of EU Programmes in Poland* (Warsaw, Ernst & Young—Program 'Sprawne Państwo').
- Fetterman, D. & Wandersman, A. (2007) 'Empowerment Evaluation: Yesterday, Today, and Tomorrow', *American Journal of Evaluation*, 28, 2.
- Forunier, D. M. (2005) 'Evaluation Defined', in Mathison, S. (ed.) *Encyclopedia of Evaluation* (Thousand Oaks, CA, Sage Publications).
- Furubo, J.-E., Rist, R. C. & Sandahl, R. (2002) *International Atlas of Evaluation* (New Brunswick, NJ, Transaction Publishers).
- Hummelbrunner, R. (2007) 'Systemic Evaluation in the Field of Regional Development', in Williams, B. & Imam, I. (eds) *Systems Concepts in Evaluation. An Expert Anthology* (Ann Arbor, MI, American Evaluation Association).
- Hummelbrunner, R. (2011) 'Applying Systems Thinking to Evaluation of Structural Funds', in Olejniczak, K., Kozak, M. & Bienias, S. (eds) *Evaluating the Effects of Regional Interventions. A Look Beyond Current Structural Funds' Practice* (Warsaw, Ministry of Regional Development).
- King, J. A. (1998) 'Making Sense of Participatory Evaluation Practice', *New Directions for Evaluation*, 80, Winter.
- Kozak, M. (2007) 'Managing European Regional Policy in Poland in the Initial Period after the Accession', *Regional and Local Studies (Studia Regionalne i Lokalne)*, Special Issue, available at: http://www.studreg.uw.edu.pl/archiwum_new.php?rok=2007#, accessed 15 August 2011.
- Kozak, M., Ledzion, B., Olejniczak, K. & Wojtowicz, D. (2009) *Ex Post Evaluation of Cohesion Policy Programmes 2000–06 Co-financed by ERDF. Management and Implementation Systems for Cohesion Policy, National Assessment Report Poland* (Brussels, EPRC & EUROREG).
- Lane, D. (2008) 'The Emergence and Use of Diagramming in System Dynamics: A Critical Account', *Systems Research and Behavioral Science*, 25, 1.
- Leeuw, F. L. & Furubo, J.-E. (2008) 'Evaluation Systems: What Are They and Why Study Them?', *Evaluation*, 14, 2.
- Leeuw, F. L., Rist, R. C. & Sonnichsen, R. C. (1994) *Can Governments Learn?: Comparative Perspectives on Evaluation & Organizational Learning* (London, Transaction Publishers).
- Meadows, D. H. (2008) *Thinking in Systems: A Primer* (Vermont, Chelsea Green Publishing).
- Nutley, S. M., Walter, I. & Davies, H. T. O. (2007) *Using Evidence: How Research Can Inform Public Services* (Bristol, Policy Press).
- OECD (2001) *Territorial Outlook* (Paris, OECD).
- Olejniczak, K. (2011) 'Tracing Theories Behind Practice—A Theory-Driven Approach', in Olejniczak, K., Kozak, M. & Bienias, S. (eds).
- Olejniczak, K., Kozak, M. & Bienias, S. (eds) (2011) *Evaluating the Effects of Regional Interventions. A Look Beyond Current Structural Funds' Practice* (Warsaw, Ministry of Regional Development), available at: <http://www.books.google.com>, accessed 15 August 2012.
- Olejniczak, K. & Ledzion, B. (eds) (2013) *Ocena systemu realizacji polityki spójności w Polsce w ramach perspektywy 2007–2013* (Warszawa, EGO s.c. dla Ministerstwa Rozwoju Regionalnego—Krajowej Jednostki Oceny).
- Olejniczak, K., Ledzion, B. & Pander, W. (2008) *Ocena potencjału oraz stanu prac w ramach jednostek ewaluacyjnych funkcjonujących w Instytucjach Zarządzających oraz Instytucjach Pośredniczących I i II stopnia w ramach NPR 2004–2006 i NSRO 2007–13* (Warszawa, EGO s.c. dla Krajowej Jednostki Oceny).
- Olejniczak, K., Strzęboszewski, P. & Bienias, S. (eds) (2011) *Review of the Evaluation Practices in the European Union. Perspectives of 27 Member States (Draft Report for Polish Presidency)* (Warsaw, Ministry of Regional Development).
- Osborne, D. & Gaebler, T. (1992) *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* (Reading, MA, Addison-Wesley).
- Ostrom, E. (2005) *Understanding Institutional Diversity* (Princeton, NJ & Woodstock, Princeton University Press).
- Patton, M. Q. (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (New York, The Guilford Press).
- Pawson, R. (2006) *Evidence-Based Policy: A Realist Perspective* (London, Sage).
- Pollitt, C. (1998) 'Evaluation in Europe. Boom or Bubble?', *Evaluation*, 4, 2.

- Preskill, H. & Torres, D. R. T. (1999) *Evaluative Inquiry for Learning in Organizations* (Thousand Oaks, CA, Sage Publications).
- PTE (2011) *Polski rynek badań ewaluacyjnych w roku 2010* (Warszawa, Polskie Towarzystwo Ewaluacyjne).
- Ramage, M. & Shipp, K. (2009) *Systems Thinkers* (London, Springer).
- Smith, P. (1995) 'On the Unintended Consequences of Publishing Performance Data in the Public Sector', *International Journal of Public Administration*, 18, 2.
- Sterman, J. (2000) *Business Dynamics: Systems Thinking and Modelling for a Complex World* (Boston, MA, McGraw-Hill/Irwin).
- Stone, D. (2011) *Policy Paradox: The Art of Political Decision Making*, 3rd edn (New York & London, W.W. Norton).
- Torres, R., Preskill, H. & Piontek, M. (2005) *Evaluation Strategies for Communicating and Reporting* (Thousand Oaks, CA, Sage Publications).
- Van der Knaap, P. (2000) 'Performance Management and Policy Evaluation in the Netherlands', *Evaluation*, 6, 3.
- Weiss, C. H. & Bucuvalas, M. J. (1980) 'Truth Tests and Utility Tests: Decision-Makers' Frame of Reference for Social Science Research', *American Sociological Review*, 45, 2.
- ZBN PTS (2010) *Badanie metaewaluacyjne treści dokumentacji przetargowych związanych z wykonaniem badań ewaluacyjnych w latach 2008–2010* (Warsaw, Zakład Badań Naukowych Polskiego Towarzystwa Socjologicznego na zlecenie Ministerstwa Rozwoju Regionalnego).