

Tomasz Kupiec

Uniwersytet Warszawski, Centrum Europejskich Studiów Regionalnych i Lokalnych (EUROREG), ul. Krakowskie Przedmieście 30, 00-927 Warszawa;
e-mail: tomasz.kupiec@uw.edu.pl; ORCID 0000-0002-6469-7746

DO WE LEARN HOW TO LEARN? EVOLUTION OF FUNCTIONS OF EVALUATION SYSTEMS — THE CASE OF POLISH REGIONAL ADMINISTRATION

Abstract: The article examines the evolution of functions of evaluation systems. It is based on the comparison of evaluation studies conducted in eight evaluation systems in the Polish regional administration in two periods: 2007–13 and 2014–20. The findings are to some extent contradictory to the expectations formed on the basis of the existing literature of the subject. Although the analysed systems were established in response to external pressures, they support accountability as well as learning. These systems do not focus on procedural issues only, and generation of strategic knowledge increases over time. Numerous regulations imposed on the analysed systems suggest, however, that the use of evaluation in the analysed systems may be symbolic in nature.

Keywords: evaluation functions, evaluation use, evaluation systems, learning, accountability, regional operational programmes

CZY UCZYMY SIĘ UCZYĆ? EWOLUCJA FUNKCJI SYSTEMÓW EWALUACJI — PRZYKŁAD POLSKICH SAMORZĄDÓW WOJEWÓDZTWA

Streszczenie: Autor analizuje, jak ewoluowały funkcje systemów ewaluacji. Bazuje na porównaniu badań ewaluacyjnych realizowanych w ośmiu systemach ewaluacji w polskim samorządzie województwa w dwóch okresach: 2007–2013 i 2014–2020. Uzyskane wyniki podważają w pewnym stopniu oczekiwania wynikające z dotychczasowej literatury przedmiotu. Choć badane systemy powstały w odpowiedzi na zewnętrzną presję, to wspierają zarówno rozliczanie, jak i uczenie. Systemy nie koncentrują się tylko na zagadnieniach proceduralnych, a produkcja wiedzy strategicznej wzrasta w czasie. Liczne przepisy regulujące sposób funkcjonowania sugeruje jednak, że wykorzystanie ewaluacji w badanych systemach może mieć charakter symboliczny.

Słowa kluczowe: funkcje ewaluacji, wykorzystanie ewaluacji, systemy ewaluacji, uczenie, rozliczanie, regionalne programy operacyjne

Programme evaluation is a systematic inquiry that assesses the quality of the implementation process, results and impact of a given intervention. When properly applied and used, it helps to understand, and improve, the performance of intervention (Cronbach, 1980). In this sense, it contributes to social and economic development.

Systematic practice of evaluation was introduced in Poland by the European Union regulations on Cohesion Policy in 2004. From the very beginning, regional administration has been an important actor of this practice. There are 16 active operating regional evaluation units – each supporting the implementation of one of 16 regional operational programmes (ROPs). Since 2007, these units have completed more than 500 evaluation studies, each of them at least 20.

As a large group of active entities, ROP evaluation units are important for the practice of evaluation in Poland; they are also a sufficient sample to study the evolution of this practice. On the other hand, the ROP is an essential instrument of development policy of regional governments. Therefore, the quality of ROP evaluation and the ability to learn from it impacts the well-being of regional communities. For this reason, I chose ROP evaluation as the most appropriate context to analyse the functions of evaluation systems. The primary research question is: *Do ROP evaluation systems support learning or accountability, and how does the proportion of these two functions change over time?*

The article is divided into five sections. The first describes evaluation functions and introduces two dichotomies which guide the analysis. Section two is a short review of the literature on evaluation systems, and is concluded with the statement of hypotheses. Section three describes the sample and the analytical procedure, and section four presents the findings. The findings are then discussed and concluded in section five.

Functions of evaluation

Use is often referred to as the key concept in the field of evaluation (Ledermann 2012; King and Alkin 2018) and the most studied theme in the literature on evaluation (Christie 2007). Closely related to it, and also popular, is the issue of evaluation functions. Focus on *use* well reflects the utilitarian nature of evaluation, but it is probably the discussion about *functions* that has more impact on how we define and understand evaluation practice. Some also argue that *use* is too narrow a concept when we analyse evaluation from the systemic perspective; instead, we should pay more attention to the functions of evaluation systems (Hanberger 2011; Leeuw and Furubo 2008; Liverani and Lundgren 2007).

Of many suggested functions of evaluation, several broader categories can be identified. Their short description is provided below, together with specific names that can be found in the literature of the subject (Widmer and Neuenschwander 2004; Boswell 2008; Hanberger 2011; Chelimsky and Shadish 1997; Batterbury 2006; Mark et al. 1999; Hanberger 2006; Patton 1996):

- providing assessment of performance for external audience / principal – *accountability, oversight, compliance*;
- inducing change in behaviour, adjustment of programme / strategy – *learning, improvement (of performance, planning), knowledge creation, building capacity*;

- justifying decisions based on other considerations than evaluation, building appearances of a learning organisation – *legitimising, substantiating, justifying, sanctioning*;
- involving stakeholders in the evaluation process, ceding power to those normally excluded from the decision-making process - *engaging, empowering, developing a sense of ownership*.

Accountability and learning are usually considered the main functions of evaluation (Van Der Meer and Edelenbos 2006). Building on that, the other functions may be considered as supplementary or derivative of the main two. The need to legitimise organisation and its decisions is driven by the fact that it is held accountable for its actions and results. At the same time there are growing expectations that organizations will learn from evaluation and support decisions with evidence. Some organisations respond to that by demonstrating even more of legitimising or symbolic use of evaluation.

While learning is inward oriented, i.e. takes place in the organisation that is evaluated or implements an evaluated intervention, the other three groups of functions are outward oriented. Evaluated organisations are accountable to superior bodies, beneficiaries, society. They try to legitimise and justify their actions usually in the eyes of the same actors that they are accountable to. Empowering and engaging are also related to external stakeholders.

Functions of evaluation may not all be fulfilled at the same time. That refers to single evaluation studies as well as evaluation systems operating at a certain time. Mark and collaborators (1999) indicate that most actual evaluations serve one or two functions at most. Referring to a single evaluation study, Olejniczak (2008) points to a contradiction between accountability and learning resulting from the different roles played by evaluators in each of these cases. In the context of evaluation systems, Raimondo argues that focusing on accountability strengthens a different set of norms and beliefs than the one needed for organisational learning.

Considering the above and the opposing orientations of the two main evaluation functions (accountability – outward, learning – inward) leads us to the first of the two dichotomies guiding the research presented in this article, viz.:

- **Evaluation is oriented either on accountability**, i.e. providing assessment of effectiveness, efficiency, value for money to external audience, e.g. supervising authorities, media, society (Batterbury 2006), **or on learning** – dedicated to actors within the implementation system with the purpose of informing decisions, e.g. modifications of programme or implementation procedures.

The second dichotomy is inspired by the concept of learning loops (Argyris and Schön 1978). If organisational learning is perceived as the improvement of the organisation's behaviour based on knowledge and understanding (Fiol and Lyles 1985; Huber 1991), learning loops indicate the scope of knowledge used and of potential improvement. Single loop learning refers to simple adjustments to procedures and routines based on basic operational knowledge. Double loop learning represents deeper reflection, questioning of strategic choices, objectives and norms based on strategic knowledge. From that perspective, the second dichotomy may be formulated as follows:

- **Evaluation provides either operational knowledge** (insights into the quality, barriers of implementation process) **or strategic knowledge** (findings on the effects of the interventions, mechanisms that explain success or failure of programmes).

Although derived from the concept of organisational learning, the division into operational and strategic knowledge also applies to the accountability function of evaluation. One may easily point to evaluation studies assessing, and informing about, the effects and impacts of intervention, as well as about the quality of procedures and implementation processes.

Combining the two dichotomous characteristics produces a two-dimensional matrix with four possible model orientations of evaluation studies or evaluation systems as a whole. Although quite a similar concept was presented in our earlier publication (Olejniczak et al. 2017), it was not used then as a research framework.

	Accountability	Learning
Strategic	<i>Accountability for results / impact</i>	<i>Learning about what works and why</i>
Operational	<i>Accountability for sound implementation</i>	<i>Improving the implementation process</i>

Figure 1. Four potential orientations of an evaluation system

Source: own elaboration based on Olejniczak et al. (2017)

Evaluation systems for accountability or learning

One of the developments in the recent years in the field of evaluation is the application of the system thinking (Leeuw and Furubo 2008), focusing not on single studies but streams of studies flowing through evaluation systems (Rist and Stame 2006), searching for the explanation of how evaluation is used in the organisational context, and looking at the design of the system (e.g. Hanberger 2011; Højlund 2014). Below you can find a short review of the literature on evaluation systems, which is used as the basis for stating the research hypotheses.

Not surprisingly, the problem of evaluation functions is one of those analysed from the perspective of evaluation systems. Picciottio (2016) focuses on accountability in discussing the role of evaluation in the context of agency dilemmas, asymmetry of information and moral hazard. Evaluation helps the principal to secure information about the agent's activities, and in that way is essential to align the interests of the agent and the principal.

Institutional theorists argue that the first and foremost role of most evaluation systems is legitimisation (Ahonen 2015; Dahler-Larsen 2012). As Leeuw and Furubo (2008) put it, they only provide procedural assurance. On many occasions organisations adapt evaluation practices (voluntary or not) because it is expected by the environment in which they operate (Powell and DiMaggio 1991). It is

worth noting that accountability and legitimacy were also long recognised as the primary functions in the EU Cohesion Policy (Batterbury 2006), which is the context for the evaluation practice in the Polish regional administration.

Organisational learning from evaluation is, therefore, a rare phenomenon. Of the four possible modes of evaluation adoption in Højlund's framework (2014), only one is characterised by a high probability of learning. It is when the organisation has a high internal propensity to evaluate and a low external pressure. As a result, evaluations rarely change the policies (Patton 1997; Pawson and Tilley 1997), and evaluation systems produce information that confirms rather than questions the policies (Leeuw and Furubo 2008), which is sometimes referred to as *designed blindness* (Friedman 2001). If learning occurs, it is usually single loop learning based on largely routinised, operational information (Leeuw and Furubo 2008).

Martinaitis, Christenko and Kraučūnienė (2018) noted that the previous literature on evaluation systems describe them statically, in the sense that it does not explain if, and how, evaluation systems once established might evolve from one type of use (or function) to another in the process of maturing. These allegations, however, are not entirely correct. Raimondo (2018) observed that when an evaluation system is institutionalised for accountability purposes, it establishes and preserves a set of certain norms and beliefs, which makes it very challenging to reorient evaluation to a learning function. Additionally, although evaluation systems are often a response to organisational loose coupling, they may in fact contribute to further decoupling of organisations. This results in a symbolic use of evaluation, i.e. maintaining an appearance of learning rather than actual learning.

Equally pessimistic perspective was suggested by Leeuw and Furubo (2008). According to them, evaluation systems, just as any other public sector unit, are focused on budget maximisation. This leads to the "shopping-for-clients" behaviour, displacing independent thinking and critical analysis with findings and conclusions tailored for justificatory and substantiating purposes.

From the perspective of the research question posed in the introduction and based on the literature discussed above, we may formulate several hypotheses regarding the behaviour of evaluation systems. None of them has been thoroughly verified so far. This study is intended to address this gap.

- H1: Evaluation systems established due to an external pressure do not support organisational learning.
- H2: Evaluation systems initially established for the accountability function stick to it also at the later stages of operation; it is not easy to reshape such systems to support learning.
- H3: If evaluation systems support learning, it is only single-loop learning based on routinised, operational information, with no real impact on policy assumptions.

Research design

To test my hypotheses, I analysed the data on evaluation studies conducted by evaluation units operating in the Polish regional administration dealing with the implementation of the EU Cohesion Policy operational programmes. These units are the cores of what can be considered evaluation systems according to the criteria proposed by Leeuw and Furubo (2008).

1. Organisational responsibility – evaluation units operate as a distinct team in larger organisational structures, usually named departments, which are a part of the regional government office. Other units of these departments are responsible for programme formulation and implementation¹, and are the potential users of the evaluation findings. On the basis of EU and domestic regulations, the voivodeship executive boards serve as the Managing Authorities (MA) for regional operational programmes (ROP); their formal responsibilities include programme evaluation.
2. Permanence – evaluation units in the regional administration have operated incessantly since 2007, i.e. they commission evaluation studies every year. To date, they completed over 500 evaluation studies, most of them in the Śląskie and Kujawsko-Pomorskie Voivodeships (44 each), and the least in the Dolnośląskie Voivodeship (20). Evaluation activities are subordinated to the programming cycle. The first evaluation plans were adopted for the years 2007–13, and they evaluated the ROP for the same period. Now, the second round of evaluation plans is being implemented for the ROPs implemented in 2014–20. As postulated by Leeuw and Furubo, the operations of the evaluation system are permanent and planned.
3. Distinctive epistemological perspective and focus on the intended use – as already mentioned, the evaluation process of ROPs is based on formally approved plans. The plans were supposed to be the outcomes of a collaborative effort engaging all the stakeholders. We might, therefore, expect that they present the shared views on what would be evaluated and why. The plans indicate what kind of knowledge would be gained at a certain time and who should be the user of that knowledge.

Out of 16 ROP evaluation systems, eight were randomly selected for the analysis (Dolnośląskie, Kujawsko-Pomorskie, Lubelskie, Łódzkie, Małopolskie, Mazowieckie, Pomorskie, Śląskie). The previous studies (Kupiec 2014b; Wojtowicz and Kupiec 2018) show that the ROP evaluation systems and their activities are similar enough to allow for the generalisation of the results from such a sample to the total population of 16. Selected systems account for 159 of a total of 301 ROP evaluation studies conducted in the period of 2007–13, and 197 out of 337 planned for the period 2014–20. Both samples allow a 5% margin of error, with a 95% confidence interval.

The aim of this research was to compare the evaluation studies from two programming periods, and to draw conclusions about the evolution of the evaluation

¹ Sometimes these functions are separated among two distinct departments.

systems based on the observed differences. To do that, each study was analysed and classified as:

- providing either operational or strategic knowledge,
- oriented either on accountability or learning.

Additionally, it was checked whether the study is compulsory (required by regulations or guidelines) or not.

The classification of completed studies was based on a systematic review of the reports. The analysed elements include: title, goals and the executive summary. If those were not clear enough, the conclusions and recommendations were also analysed. In the case of studies planned for the future, fiches of the studies were analysed (including the scope, type of study, goals, justification, key questions/problems). Although each of the characteristics is supposed to be dichotomous, in reality many of the analysed studies provide both strategic and operational knowledge or they seem to be oriented to accountability and learning at the same time. In such cases, the studies were assigned to both groups.

Three basic measures were applied to test the hypotheses: the number of studies dedicated to accountability in relation to the number of studies dedicated to learning in the 2007–13 perspective (H1); the share of studies dedicated to learning in the 2014–20 perspective in relation to the same share in the 2007–13 perspective (H2), and the ratio of studies providing operational and strategic knowledge to the total number of studies supporting learning.

Findings

Exactly two-thirds of the 159 analysed studies from the 2007–13 perspective can be described as potentially supporting learning. Almost the same share (65%) focused on accountability. Therefore, although the evaluation practice of ROP was clearly coerced in 2007 by EU regulations, the systems, even in the initial period of 2007–13, were not entirely devoted to accountability. Instead, they served both learning and accountability in balanced proportions.

What draws the researchers' attention is the fact that a large majority of learning concerned only operational issues (53% of studies) – identifying problems in the implementation process and suggesting simple improvements of procedures. It represents single loop learning (Argyris and Schön 1978). At the same time, strategic, double loop learning, i.e. substantial revision of premises underlying the interventions and changing the instrument settings, was rarely supported by evaluation studies in the 2007–13 period (16% of studies). Therefore, when focusing only on strategic knowledge, we can observe a substantial advantage of accountability studies over learning studies (40% to 16%).

When we refer to the entire sample of studies, both strategic and operational, hypothesis 1 stating that evaluation systems established due to external pressure do not support learning appears false. However, if we limited our scope of interest to the – more important – strategic evaluation, the hypothesis would hold true.

The share of studies oriented on learning has not changed between the periods 2007–13 and 2014–20. In fact, we can observe a shift in that direction as the share

of such studies has increased from 65% to 84%. Again, the picture is different if we focus on strategic studies only. Although the share of *strategic accountability* studies outbalances *strategic learning* ones in both analysed periods, the difference is visibly smaller in 2014–20.

Therefore, as in the case of hypothesis 1, conclusions concerning hypothesis 2 differ depending on the scope of the evaluation studies we take into account. When all the studies are included, H2 holds true. The ROP evaluation systems established due to an external pressure from the EU did not focus more on learning in the period 2014–20. *Accountability* studies were frequent in the first programming period, and are even more frequent now. On the other hand, one might argue that H2 becomes irrelevant from the perspective of the findings. We cannot expect reorientation on *learning* because it was already an important function in the 2007–13 period and it stayed equally important in 2014–20, with two-thirds of the studies supporting it.

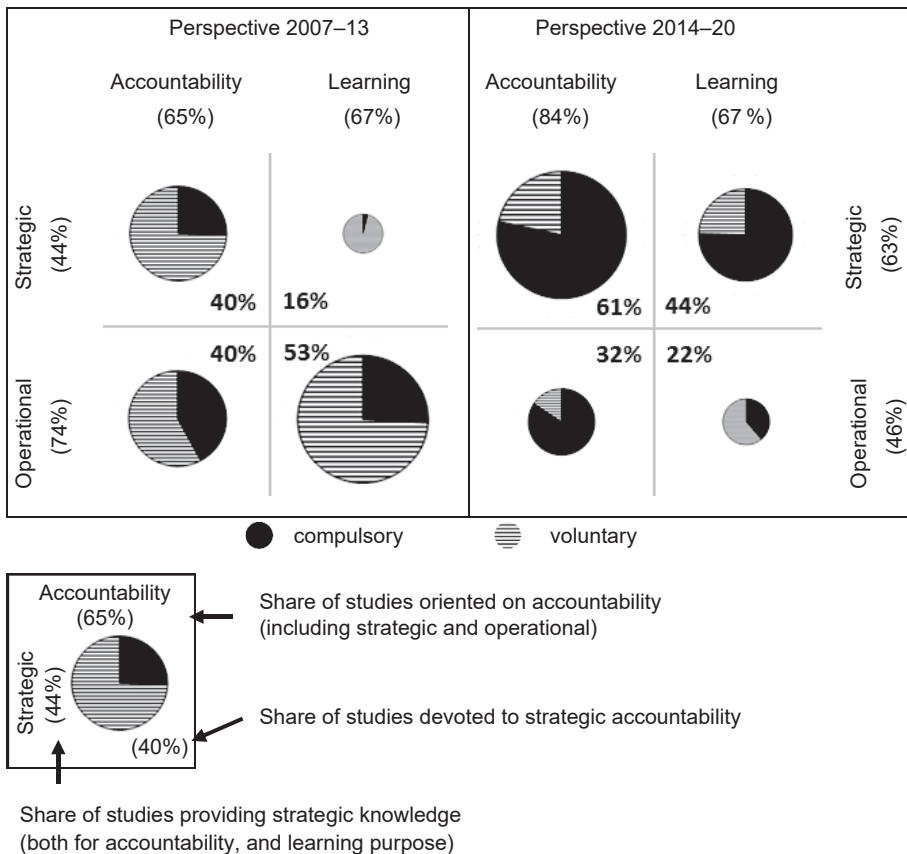


Figure 2. Orientation of ROP evaluation systems

Source: own elaboration

If we focus on strategic evaluation only, H2 is definitely relevant, but it appears false. Accountability was a dominant function of strategic studies conducted by ROP evaluation systems in 2007–13. However this dominance was significantly limited in 2014–20 period, when learning is not supported by only one-third of strategic evaluation studies.

The most striking observation from the comparison of the programming periods is the shift from operational to strategic evaluation. In the 2007–13 perspective, three-fourths of the studies focused on the implementation process, with 44% focusing on the effects. The proportion has switched in the current period: the majority (63%) of the studies analyse the effects, while only 46% provide operational knowledge. The shift concerned both accountability and learning studies, but it was with greater intensity in the subset of studies devoted to learning. This observation is in opposition to hypothesis 3. Contrary to the view of Leeuw and Furubo (2008), ROP evaluation systems in the period 2014–20 do not focus entirely on the implementation process and procedures, but provide more knowledge on the effects of the interventions and mechanisms explaining them.

After verification of the hypotheses, the categories of *accountability* and *learning* will be further decomposed to provide more insights into the nature of the outcomes of ROP evaluation systems. The section is concluded with the discussion about the impact of legal requirements on the structure of evaluation studies.

Accountability

There are at least few actors that the ROP Managing Authorities are accountable to. The first is the European Commission (EC) – the EU executive body responsible for the implementation of the Cohesion Policy at EU level and – in simple terms – for providing funds for ROP implementation. At national level, the key addressee of ROP evaluation is the Ministry of Investment and Economic Development, coordinating implementation of the Cohesion Policy at national level. Evaluation studies conducted to satisfy those institutions represent the *external accountability* function.

The voivodeship executive boards are also required, or at least expected, to inform actors at regional level about the implementation and effects of ROPs. Those include programme beneficiaries (among them, local governments), the regional community in general (and, more specifically, voters), the media. The studies addressed to those actors were classified as fulfilling *internal accountability*.

In the period 2007–13, the proportion of external and internal accountability studies was roughly even, with a slight advantage of internal accountability. Typical examples of *strategic external accountability* were studies estimating the target values of impact indicators with the HERMIN model, measuring the impact of ROP on the goals of the Lisbon Strategy and the Europe 2020 Strategy, or judging the compliance of ROP with EU horizontal principles (sustainable development, equal opportunities, non-discrimination and gender equality). Most common *operational external accountability* studies were focused on information

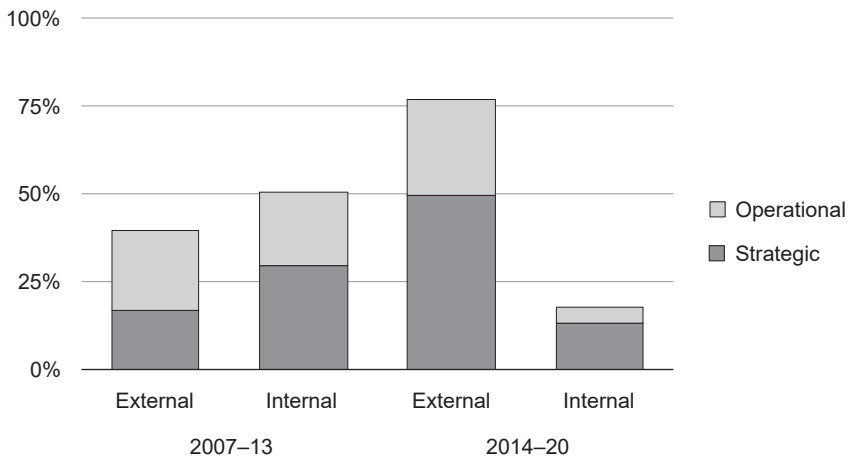


Figure 3. External and internal accountability*

* share in the total number of evaluation studies in a given programming period

Source: own elaboration

and promotion activities of the MA ROP. Typical examples of *internal accountability* were ex-post evaluations. Since they dealt with the entire scope of the ROP intervention, they could not go into details of the mechanism of change. They usually presented only general information on financial and material progress, and indicator values. Therefore, they supported neither learning nor future programming, but only informed about what had been achieved.

The key change in the current perspective (2014–20) is the dynamic increase of *strategic external accountability* studies. It was caused primarily by the new EU regulations requiring each priority of operational programme (such as ROP) to be evaluated at least once during the programming period. As a result, many ROP evaluation units planned a set of evaluations of effects – one per priority – declaring that it is a response to the EU obligation. At the same time, most ROP evaluation units have reduced the number of operational studies, e.g. assessing the readiness for the implementation of major projects. As a result of those two changes, there are four times more *external accountability* studies than *internal accountability* studies in the 2014–20 programming period.

This shift is not in line with the observations concerning international organisations – where evaluation was also originally introduced to ensure accountability. According to Raimondo (2018), accountability in international organisations has historically been oriented upward and externally to oversight bodies, but over time these relationships have become more complicated, and the importance of giving account downwardly to clients, and internally to themselves, has increased.

Learning

The distinction between operational and strategic learning is in practice not very clear. For the purpose of this research, only evaluation studies dealing exclusively with implementation process issues (e.g. assessing and improving selection criteria, application and contracting process, list of indicators, capacity of applicants) were considered as supporting *operational learning*. All other studies including any considerations on ROP results or goal achievements were classified as potentially supporting *strategic learning*.

The latter constitute a broad and heterogeneous category. It can be divided into three subgroups distinguished by the scope of learning:

- *effects* – studies assessing ROP results in a specific, relatively narrow field, e.g. impact on SME innovation, education, transport. Since those studies focus on some type of intervention, they are quite detailed and show what was effective and what was not. This can serve as the basis for potential modifications of intervention in the future – learning.
- *barriers and facilitators* – studies which move a step further and discuss barriers and facilitators of efficiency, success factors².
- *theories of change* – studies combining barriers and facilitators with intervention assumptions into complete theories of change, revealing actual causal relationships between the actions and the observed effect. As part of the evaluation, the logic model of intervention is identified, verified and modified. This type of evaluation studies provides the highest chances that the double loop learning will actually occur in the programme cycle.

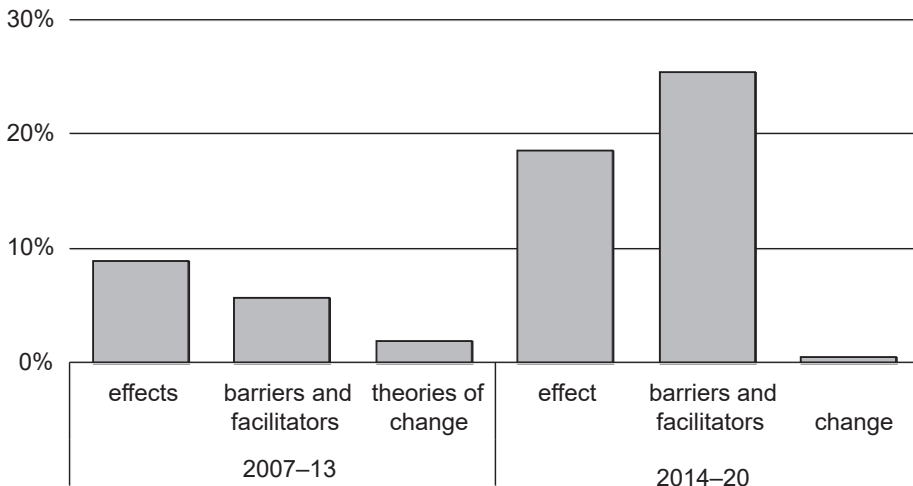


Figure 4. Strategic learning – share in the total number of evaluation studies in particular programming period

Source: own elaboration

² It is worth mentioning that in some evaluations of results, the barriers and facilitators identified by the evaluators relate only to the implementation process, which means that only operational learning is possible.

Evaluation studies developing theory of change were very rare in the 2007–13 period, and this situation has not changed in the current perspective. Therefore, the contribution of ROP evaluation to the formulation of new policy instruments is not as great as we could hope for. As already presented in Fig. 2, the share of strategic learning studies increased substantially in the period 2014–20. Additionally, we can see in Fig. 4 that this was not only a quantitative but also a qualitative change, as the proportion of the studies on effects and studies presenting barriers and facilitators has shifted in favour of the latter.

Compulsory and obligatory studies

In analysing ROP evaluation plans for the current perspective, one may draw conclusions that the single most important cause of differences between the 2007–13 and 2014–20 periods are the obligations included in the EU and domestic regulations. The general requirement of evaluating operational programmes under the Cohesion Policy was, and is in force, in both analysed programming periods. However, the detailed provisions on the types and subjects of studies are more numerous and rigorous in the 2014–20 perspective. As a result, the scope of discretion of ROP evaluation units has been limited and the number of studies that have to be regarded as compulsory has increased substantially.

In total, the share of compulsory studies increased from 26% to 72%, and the direction of change is the same in all four quarters of the evaluation orientation matrix (Fig. 2). The greatest increase of obligatory studies can be observed in *strategic learning*. It was fuelled by the obligation to evaluate the results of each priority of operational programme at least once during the programming period, as well as the requirement to assess the values of indicators related to the European Social Fund³. The share of *operational learning* was influenced by the obligation to evaluate the selection criteria but also informal discouragement from the National Evaluation Unit (NEU) to evaluate information and promotion activities. Mid-term evaluations and studies presenting the general progress of results – both serving *strategic accountability* – are also required in the guidelines issued by the NEU.

Discussion and conclusions

This article compares the products – evaluation studies – of ROP evaluation systems in two programming periods, 2007–13 and 2014–20, to ascertain the extent to which the systems support learning and accountability, as well as the potential shift in this respect over time. The observations described in the findings section are intriguing in the sense that they do not fully support the hypotheses built on the basis of the previous (mostly theoretical) literature on evaluation systems. Although ROP evaluation systems were established due to formal

³ Although it is not explicitly stated that a separate study is necessary for each priority, in most ROP evaluation systems it was interpreted that way.

requirements stated in the EU regulations, in the period 2007–13 they supported learning as much as accountability, which is against the expectations. The picture gets more complicated when we introduce the distinction between strategic and operational knowledge. Such division is justified by the previous research showing that evaluation is more useful and appropriate as a source of knowledge on processes rather than effects (Olejniczak et al. 2017). When referring just to strategic knowledge, we see that the systems were oriented on accountability in the period 2017–13, which supports the hypothesis. However, unexpectedly the learning function receives much more attention in the 2014–20 perspective. Against the expectations is also the significant increase in the production of strategic knowledge, at the expense of operational.

The discrepancy between the current state of affairs and the empirical observations presented here may prove that hypotheses and theories on the behaviour of evaluation systems are context dependent and it is risky to generalise them. What distinguishes ROP evaluation systems is the fact that they are part of a larger multi-layered evaluation system operating in complex policy settings. In such multi-layered settings, the distribution of the influence potential between the stakeholders may be a key factor determining the orientation of the systems and the dominant types of evaluation use (Eckhard and Jankauskas 2019). These influences may take the form of informal encouraging and discouraging from conducting certain evaluations, or formal regulations as is the case with ROP evaluation systems.

The high number of obligatory studies is not without influence on ROP evaluation practice. The sets of studies planned and conducted in the analysed systems are very similar to each other. It is what Powell and DiMaggio (1991) termed coercive isomorphism. Since ROP evaluation units are obliged to conduct the same studies, they borrow the detailed concepts from each other⁴. In effect, coercive isomorphism triggers additional mimetic isomorphism, making the products of ROP evaluation system even more similar and reducing collective learning in favour of duplicating similar conclusions from similar studies.

External pressure exerted by the EU regulations and domestic guidelines raises questions about the actual purpose of ROP evaluation systems. Even though from the perspective of the products we can observe a growing focus on strategic learning, there is a risk that to some extent it represents only a symbolic use of evaluation as predicted by Højlund (2014). From the perspective offered by Dunlop and Radaelli (2013), what takes place in ROP evaluation systems is the so-called “learning in the shadow of hierarchy”. With an increase in regulations, learning shifts from the flexible delegation type to the hetero-directed one, which is suitable for coping with instructions or following the procedures, not challenging and changing policy assumptions. A practical manifestation of this may be evaluation studies that deal with strategic issues but then offer only procedural recommendations – a problem already observed in the ROP evaluation practice (Kupiec 2015).

⁴ This problem in the context of ROP evaluation was observed in earlier studies (Kupiec 2014a).

Another practical problem is that despite a growing number of studies classified as supporting strategic learning, representatives of ROP evaluation units declare that, in general, evaluation informs the programming of new interventions to a lesser extent than before, due to bad timing⁵. Although evaluation systems are established to remedy organisational loose coupling (Raimondo, 2018), stricter external rules imposed on RPO evaluation practice and not supported by internal incentives may lead to further decoupling, and not support actual learning.

The first practical implication and recommendation of this study for the evaluation practice is derived directly from the paragraph above. I believe there is a need to revise the regulations (EU and domestic) and reduce the number of compulsory evaluation studies. This number has increased dramatically between 2007–13 and 2014–20. It has resulted in a seemingly positive rise of strategic learning studies. However, in reality it may have led to a reduction in learning from evaluation and an increase of symbolic use. Another recommendation also concerns strategic learning. Despite the growth in the number of studies classified as generally supporting strategic learning, there is a dramatic deficit of studies providing complete theories, explaining mechanisms of change, revealing causal relationships between the actions and the observed effects. Both the authorities responsible for RPO evaluation systems and the national authorities coordinating the entire practice of Cohesion Policy evaluation should join efforts to provide more such studies, as only they provide the proper basis for actual double loop learning – questioning beliefs and norms, redesigning policy choices and objectives.

As for future research on the subject, it seems necessary to explore how the regulatory and organisational context in which evaluation systems at the level of organisations operate determine the evolution of their functions. There are at least several potential factors to examine. Apart from regulatory obligations concerning evaluation, these are regulatory settings of the evaluand (ROP being part of the Cohesion Policy implementation system in our case), structure of the evaluation system at national level or activity of the leading actors of evaluation practice / coordinating bodies in the national context.

References

- Ahonen, P., 2015, “Aspects of the institutionalization of evaluation in Finland: Basic, agency, process and change”, *Evaluation*, 21(3), 308–324.
- Argyris, C., and Schön, D. A., 1978, *Organizational Learning*, Reading, MA: Addison-Wesley Pub. Co.
- Batterbury, S. C. E., 2006, “Principles and purposes of European Union Cohesion policy evaluation”, *Regional Studies*, 40(2), 179–188.

⁵ This problem was reported by several Polish regional evaluation units interviewed as part of the evaluation study “The Role of Evaluation in Cohesion Policy”, commissioned by the European Parliament (REGI Committee) (report to be published).

- Boswell, C., 2008, "The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy", *Journal of European Public Policy*, 15, 471–488.
- Chelmsky, E., and Shadish W. R., 1997, *Evaluation for the 21st Century: A Handbook*, Thousand Oaks, CA: SAGE.
- Christie, C. A., 2007, "Reported influence of evaluation data on decision makers' actions: An empirical examination", *American Journal of Evaluation*, 28(1), 8–25.
- Cronbach, L. J., 1980, *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*, San Francisco: Jossey-Bass Publication.
- Dahler-Larsen, P., 2012, *The Evaluation Society*, Stanford, CA: Stanford University Press.
- Dunlop, C. A., and Radaelli, C. M., 2013, "Systematising policy learning: From monolith to dimensions", *Political Studies*, 61, 599–619.
- Eckhard, S., and Jankauskas, V., 2019, "The politics of evaluation in international organizations: A comparative study of stakeholder influence potential", *Evaluation*, 25(1), 62–79.
- Fiol, M., and Lyles, M., 1985, "Organizational learning", *Academy of Management Review*, 10(4), 803–813.
- Friedman, V. J., 2001, "Designed blindness: An action science perspective on program theory evaluation", *American Journal of Evaluation*, 22(2), 161–181.
- Hall, P., 1993, "Policy paradigms, social learning, and the state: The case of economic policymaking in Britain", *Comparative Politics*, 25(3), 275–296.
- Hanberger, A., 2006, "Evaluation of and for democracy", *Evaluation*, 12(1), 17–37.
- Hanberger, A., 2011, "The real functions of evaluation and response systems", *Evaluation*, 17(4), 327–349.
- Højlund, S., 2014, "Evaluation use in the organizational context – changing focus to improve theory", *Evaluation*, 20(1), 26–43.
- Huber, G., 1991, "Organizational learning: The contributing processes and the literatures", *Organization Science*, 2(1), 88–115.
- King, J. A., and Alkin, M. C., 2018, "The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research", *American Journal of Evaluation*, 40(3).
- Kupiec, T., 2014a, "Użyteczność ewaluacji jako narzędzia zarządzania regionalnymi programami operacyjnymi", *Studia Regionalne i Lokalne*, 2, 52–67.
- Kupiec, T., 2014b, "Evaluation practice of regional operational programmes in Poland", *Management and Business Administration. Central Europe*, 23(3), 135–151.
- Kupiec, T., 2015, "Program evaluation use and its mechanisms: The case of Cohesion Policy in Polish regional administration", *Zarządzanie Publiczne*, 3, 67–83.
- Ledermann, S., 2012, "Exploring the necessary conditions for evaluation use in program change", *American Journal of Evaluation*, 33(2), 159–178.
- Leeuw, F. L., and Furubo, J. E., 2008, "Evaluation systems: What are they and why study them?", *Evaluation*, 14(2), 157–169.
- Liverani, A., and Lundgren, H. E., 2007, "Evaluation systems in development aid agencies: An analysis of DAC peer reviews 1996–2004", *Evaluation*, 13(2), 241–256.
- Mark, M. M., Henry, G. T., and Julnes, G., 1999, "Toward an integrative framework for evaluation practice", *American Journal of Evaluation*, 20(2), 177–198.
- Martinaitis, Ž., Christenko, A., and Kraučiušienė, L., 2019, "Evaluation systems: How do they frame, generate and use evidence?", *Evaluation*, 25(1), 46–61.

- Olejniczak, K., 2008, *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średnio-okresowych INTERREG III*, Warszawa: Wydawnictwo Naukowe Scholar.
- Olejniczak, K., Kupiec, T., and Newcomer, K., 2017, „Learning from evaluation – the knowledge users’ perspective”, *Evaluation Theory and Practice*, 5(2): 49–74.
- Patton, M. Q., 1996, „A world larger than formative and summative”, *American Journal of Evaluation*, 17(2), 131–144.
- Patton, M. Q., 1997, *Utilization-Focused Evaluation: The New Century Text*, Thousand Oaks, CA: SAGE.
- Pawson, R., and Tilley, N., 1997, *Realistic Evaluation*, London: SAGE.
- Picciotto, R., 2016, “Evaluation and bureaucracy: The tricky rectangle”, *Evaluation*, 22(4): 424–434. *Political Studies* 61: 599–619.
- Powell, W. W., and DiMaggio, P. J. (eds.), 1991, *The New Institutionalism in Organizational Analysis*, Chicago, MA: University of Chicago Press.
- Raimondo, E., 2018, “The power and dysfunctions of evaluation systems in international organizations”, *Evaluation*, 24(1): 26–41.
- Rist, R., and Stame, N. (eds.), 2006, *From Studies to Streams: Managing Evaluative Systems*, London: Transaction Publishers.
- Van der Knaap, P., 1995, “Policy evaluation and learning: Feedback, enlightenment or argumentation?”, *Evaluation*, 1(2), 189–216.
- Van Der Meer, F.-B., and Edelenbos, J., 2006, “Evaluation in multi-actor policy processes: Accountability, learning and co-operation”, *Evaluation*, 12(2), 201–218.
- Widmer, T., and Neuenschwander, P., 2004, “Embedding evaluation in the Swiss federal administration”, *Evaluation*, 10(4), 388–409.
- Wojtowicz, D., and Kupiec, T., 2018, “Reluctant to learn? The use of evaluation to improve EU cohesion policy implementation in Polish and Spanish regions”, *Evidence & Policy*, 14(1), 103–120.