

Variations of evaluation functions in different administrative settings

Tomasz Kupiec, University of Warsaw, t.kupiec2@uw.edu.pl

Dominika Wojtowicz, Kozminski University, doma@kozminski.edu.pl

Paper prepared for the International Conference on Public Policy (ICPP4), Montreal 26th - 28th June 2019, Panel : POLICY DESIGN, POLICY ANALYSIS, EXPERTISE AND EVALUATION

Introduction – problem of evaluation use

The importance of evaluation in public sector lies in its potential to improve public interventions (programs, projects and legal regulations), which should eventually lead to social betterment (e.g. Christie 2007; Mark, Henry 2004).

In the traditional rational approach evaluation provide information about the validity of the policy theory and the efficiency of the means employed. Policy-makers were expected then to use this information directly to change their policy scheme and improve its quality (Van der Knaap, 1995). This type of evaluation use was later labelled as *instrumental use*.

The reality is however different. A number of studies, starting from 1970s (e.g. Weiss, 1972; Alkin et al., 1979; Weiss and Bucuvalas, 1980) informed about disappointingly limited use of evaluation, defined as above. Since then *use* became the key concept (Ledermann, 2011) or hot topic (Alkin and Coyle, 1988) in the field of evaluation, and it remains a central topic of the debate until now (King, Alkin, 2018).

Rich body of literature offers numerous classifications and models of factors of evaluation use, sometimes multi-layered and comprising more than 50 factors (Huberman, Gather Thurler, 1991; Leviton, Hughes, 1981). Among notable works are classifications based on extensive reviews of empirical and theoretical studies by Cousins and Leithwood (1986), Shulha, Cousins (1997), Johnson et al. (2009). They all group factors of evaluation use in two categories:

- characteristics of evaluation process – e.g. credibility, relevance, timeliness;
- decision or policy settings – e.g. type of information needs, decision characteristics, political climate.

Despite efforts from evaluation theorists and practitioners results are still often regarded as “disappointingly inconclusive” (Ledermann 2012). In large sets of factors there is often no prioritization, so the relative importance of factors is not clear. As some argue, findings of the research on the use of evaluations often lack sufficient scientific credibility. They might be useful as a guidance for professional evaluators, but they do not contribute to theory building (Brandon, Singh, 2009; Johnson et al., 2009).

Serious shortcoming of the previous literature is that while it presents conditions necessary for the occurrence of desirable types of evaluation use - *instrumental use* (evaluation directly informs decisions) or *conceptual use* (evaluation alters awareness and attitudes of decision-makers), it does not explain why evaluation is conducted so often for *symbolic* purpose only (i.e. to show that a programme or organisation is trustworthy because it values accountability and learning).

One of the suggested explanation for that situation is the fact that most of the work on evaluation use so far has taken the perspective of specific single evaluation study, which has left factors in the institutional context at the periphery of the analysis (Raimondo, 2018). Good example of that are works of Balthazar (2006,2009). The author declares to deal with institutional context, but he does it from the perspective of single evaluation study. As a result the factor he focuses on – distance between evaluator and evaluatee – is related much more to the characteristics of the process of conducting particular study than to the institutional context in which the organization conducting (or commissioning) evaluation operates.

System thinking in evaluation

Recently proposed and introduced remedy for the described shortcoming is the application of the system thinking to the field of evaluation (Leeuw, Furubo 2008), focusing not on single study, but streams of studies flowing through evaluation systems (Rist & Stame 2006), searching for the explanation of evaluation use types and functions in the organizational context, and design of the system (e.g. Hanberger 2011).

A few worth noting contribution were made as a part of this approach. Højlund (2014), drawing from earlier literature on organizational institutionalism, developed a framework explaining types of evaluation use observable in particular organizations as a resultant of two variables:

- external pressure on adoption of evaluation – regulation, cultural constraints, uncertainty or normative expectation in environment;
- internal propensity to evaluate – depending on the characteristics of the organization (whether it is political or action).

Four possible combinations of those two lead to four *adoption modes* of evaluation practice: coercive, mimetic, normative, voluntary. Every of adoption modes leads directly to particular type of evaluation use.

To some extent similar findings were presented earlier by Boswell (2008), who argues that type of evaluation use depends on:

- the role of organization, and source of its legitimacy;

- features of policy area: 1) degree of contestation, and 2) the mode of dealing with conflicts, whether it is democratic or technocratic.

Eckerd & Moulton (2011) studying non-government sector discovered that organizations adopt particular evaluation practice based on the pressure from environment, but the dominant evaluation use type depends on the role of organization in the society. Therefore the two – type of evaluation practice and type of use – are not interrelated.

Although differing in details, these three frameworks share the same general view that it is the environment (pressure from other institutions, features of the policy domain in which organization operates) and characteristics of particular organization (the role of organization, and source of its legitimacy) that determine the practice of evaluation and the function evaluation serves in evaluation system.

It is a very logical and theoretically well-grounded theory. It may work for most contexts. However there are empirical observations of evaluation systems behaving differently (Wojtowicz, Kupiec, 2018). Based on that we believe that the current literature is not applicable to the specific context of multi-layered evaluation systems in complex policy settings.

In the next two parts we will describe the concept of multi-layered evaluation system and give an example of one – evaluation of European Union cohesion policy.

Multi layered evaluation system

As Williams and Imam (2007) suggest thinking in terms of evaluation systems requires defining boundaries – deciding what lies within and outside of system. With number of varying definitions available it is not easy though to set a clear boundary of evaluation system. Hojlund (2014) writes about “several interdependent organizational entities”. Hanberger (2011) mentions systems set up for a group, but also a single organization. Both authors seem to refer specifically to entities commissioning (and at the same time using) studies. Number of institutions as a criterion of evaluation systems appears also in the definition of Leeuw and Furubo (2008). Among them both producers of knowledge and organization striving to use it are expected.

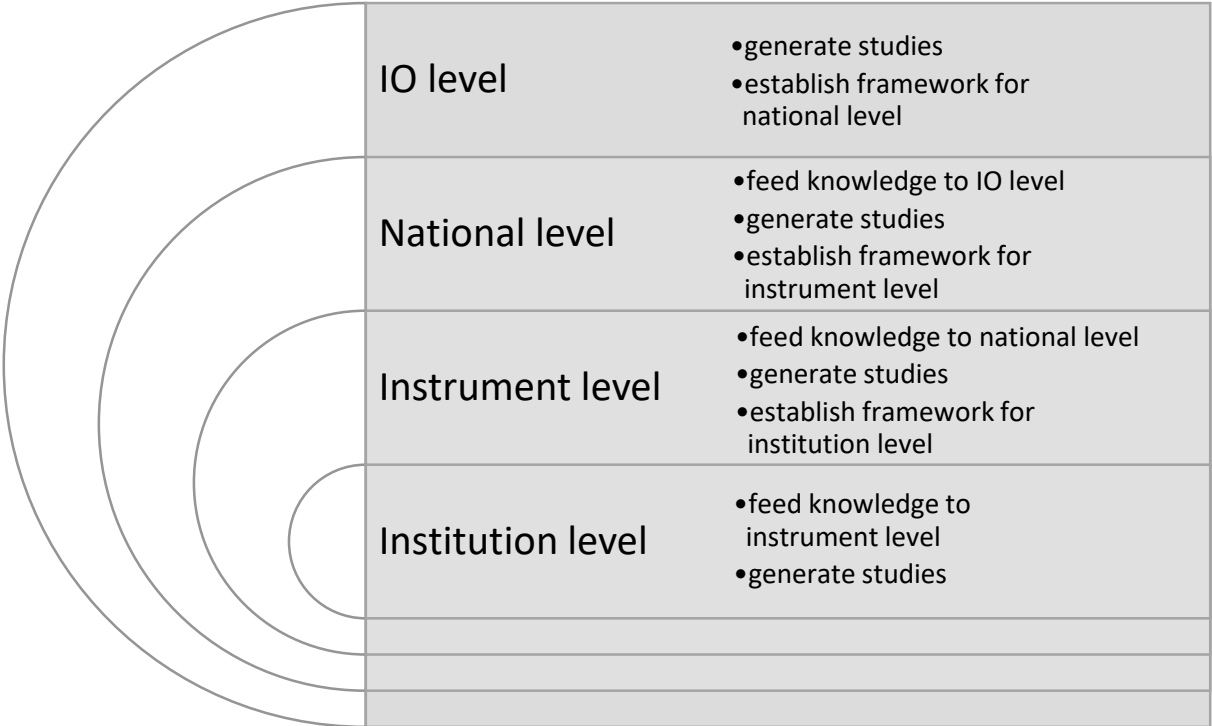
First shortcoming of those definitions is that they ignore unique position and crucial role of evaluation units. Those units are specialized in acquiring evaluative knowledge, usually by commissioning studies to external contractors. At the same time they should not be regarded as ‘users’ since their job is just to feed knowledge to other decision making units in the organization. Evaluation units operate in-between producers and users and deal with knowledge brokering (Olejniczak et al 2016). In our research evaluation units are the building elements of evaluation systems. Other entities / units, especially producers and users of evaluation are, according to system

theory, considered as components of internal environment over which system has some control and closely interacts with it.

Complex policy settings (like the shared management approach in European Union policies – see next part) determines another feature of evaluation system – it may be multi-layered and nested. The elementary evaluation system operates at the level of single institution managing policy instrument (program, or part of it, i.e. set of priorities). If program management is shared between several institutions their evaluation units combined together constitute second level of evaluation system – program level. Next level - national - comprises of units dealing with all instruments in particular state under certain policy. The last and highest level might be international organization (IO) level, e.g. European Union.

Evaluation studies are conducted at each level of the system. Higher level of evaluation system additionally creates regulatory and organizational environment for the subordinate one. Lower levels feed knowledge to the superiors.

Figure 1 Multi-layered evaluation system



Source: own elaboration

European Union cohesion policy – a case of complex policy settings with multi-layered evaluation system

The subject of our interest in analyzing multi-layered evaluation system is the European Union's cohesion policy evaluation system. The reason for this is three fold. First, many of the EU countries (mostly the ones that joined the Community in the 80. of past century and afterwards) had scant experience and very limited practices in conducting public policies and programs evaluations. There is growing evidence of the role of Cohesion policy evaluation encouraging spillovers into the evaluation practice within the domestic policies of EU Member States (Bachtler et al, 2010). It is worth stressing that cohesion policy comparing to other EU policies has always been comprehensively scrutinized and evaluations has played a significant role in its planning and implantation. This mostly due to the size of the budget allocated, but also to the fact that it covers activities under many other EU policies (e.g. relating to innovations, SMEs, industrial policy). It has experienced the emergence of an evaluation culture that is generally more developed and sophisticated than that of other EU or domestically funded regional development policies, which is reflected in the large number of evaluation studies produced by MS and by European Commission itself (Fratesi, Wislade 2017).

Second, cohesion policy is one of the EU policy, which is being implemented on the principles of so called *shared management*, the essence of which is the division of duties resulting from the implementation of arrangements adopted as part of the political negotiations process between the EU and the Member States. The idea of shared management is to achieve common goals, leaving the Member States and their administrations with a large dose of autonomy in adopting the institutional solutions (Scharpf, Fritz 1994). The European Commission for each seven-year perspective develops guidelines obligating countries to undertake evaluation activities. These guidelines are defined by the Regulation, and therefore are identical for all beneficiary countries, but they do not specify the structure of the system to be provided for the required assessment of the implementation of the funds and their effects. The European Commission shall provide guidance on how to carry out evaluations. The role of the European Commission is seen as a supporter and promoter of good practices, rather than an institution enforcing mandatory solutions. As it is stated in the common regulatory framework, evaluation systems in Member States should guarantee the accurate assessment of the effectiveness, efficiency and impact of assistance from cohesion policy funds in order to improve the quality of design and implementation of programs, and to determine their impact in relation to the targets set out by the EU. Since 2014, an evaluation plan should be drawn up at the national level in order to ensure more coherent, complex and structured actions concerning evaluations undertaken in the Member States. The European Commission provides support in preparing evaluation plans to institutions involved in implementation of cohesion policy programs, giving its comments on the scope and methodology proposed in the plan. It should be stressed that the regulations on evaluation plans are not very tight, leaving space for managing authorities to decide on the timing, scope and methodology of evaluations

conducted. Hence, Member States may decide to be more ambitious and to go beyond what is required by the Regulation.

In principle, the shared management cohesion policy implementation should ensure a greater sense of responsibility for the results to be achieved and ability of the Member States' administrations to effectively and efficiently manage policy. Such concept can be reduced to the statement "different paths, same goal". The common provisions included in the EU Regulations point out the functions that the evaluation should perform, however - according to the shared management concept - there are no details how to organize the evaluation systems with a silent assumption that it will make the systems tailored to national administrations, so that they can effectively implement the same, shared functions.

And finally, cohesion policy is a suitable case of a multi-layered evaluation system as its implementation involves as many as three administrative levels, i.e. EU, national and regional. Member State and its regions are responsible for carrying out evaluations and making use of their results during the implementation period. The Commission executes a supervisory role by checking that the arrangements for the management and control systems adopted by each Member State are in line with the guidelines.

Research goal

The aim of our research is to refine the theory of evaluation use in evaluation systems by identifying factors determining the organization and orientation of complex evaluation systems.

Two hypotheses will be verified:

- Contrary to current views in the literature, organizations under the same regulatory pressure and characterized by the same (or lack of) internal propensity to evaluate, may develop different evaluation practice – systems organized in different ways and fulfilling different functions
- The method of organizing the evaluation system affects its orientation – knowledge provided, functions performed, target users

Research design and methods

Our main subject of interest – unit of analysis - is the evaluation system at national level. However to capture implications of choices made at national level we studied also evaluation units at organization level.

The study covers Cohesion Policy evaluation systems in 8 out of 13 countries of central and south-eastern European countries that jointed the EU in 2004 or later, i.e.: Bulgaria, Croatia, the Czech Republic, Hungary, Poland, Romania, Slovakia and Slovenia. We concentrated on these particular countries for several reasons. On one hand, these countries have common features, which make

them comparable as case studies. This regards the fact that they represent group of the greatest beneficiaries of CP funds in per capita terms due to the level of their economic development measured by GDP, which is one of the element of so called Berlin algorithm basing on which the EU assistance in granted. The countries are also characterized by comparable administrative culture, as all of them were part of ex-Soviet Block and have been continuing the socio-political transformation since the fall of Berlin Wall in 1990. And finally, all these countries share quite similar paths of evaluation practice development, with evaluation being enforced by the regulations after joining the EU, and remaining limited almost exclusively to CP programs.

The analysis is based on the following data sources:

- (1) a survey of heads of all Cohesion Policy evaluation units in eight countries conducted from June 5 to August 01 2017. Invitation to participate in the survey was sent to a total of 80 units, and we received 74 complete responses, what states the response rate on 93% level. The details of the country cross-section are presented below.

Country	Surveys sent	Surveys received	Response rate
Bulgaria	7	5	71%
Czech Republic	12	11	92%
Croatia	5	4	80%
Hungary	1	1	100%
Poland	36	36	100%
Romania	3	3	100%
Slovakia	15	13	87%
Slovenia	1	1	100%
Total	80	74	93%

- (2) 27 interviews with representatives of: central coordinating bodies for evaluation, leading evaluation units, local evaluation experts or members of evaluation societies in each country (not less than 3 interviews per country);
- (3) Interviews with European Commission representatives;
- (4) Desk research of existing English-language sources.

All studies were conducted from May to October 2017.

Results

a. Variations in evaluation system design

Although developed under the same UE rules evaluation systems in V4+4 countries varies in at least two dimensions, i.e.: (1) level of (de)centralization and structure, and as a consequence of that – number of evaluation units, (2) existence and activity of central coordination body.

Structure

The most common solution in terms of (de)centralization is the one found in Czechia, Slovakia, Bulgaria and Croatia. Number of evaluation units operate in each of those systems. They are located in ministries who act as managing authorities (MAs) for certain operational programs (OPs). In those systems evaluation unit is usually responsible for evaluating single OPs that is implemented by the ministry it is part of.

Poland is the example of even deeper decentralization. Evaluation units operate not only at the level of ministries – MAs, but also implementing bodies – usually government agencies – responsible for implementation of selected priorities under certain OP.

At the other extreme lies Hungary – centralized evaluation system with only one evaluation unit. It operates in the Prime Minister's Office and has exclusive competences regarding evaluation of 7 OPs. MAs for those programs are located in different ministries. Similar – single unit evaluation system – exists in Slovenia. It is however to large extent determined by the scale and structure of CP implementation system. Only one OP is implemented in Slovenia, by the Government Office for Development and European Cohesion Policy, which also host the evaluation unit.

Romania constitutes another interesting example of centralized structure. There is a single unit responsible for evaluation and operating within Ministry of Regional Development, Public Administration and European Funds. However this unit is divided into three sub-units – two of them responsible for evaluation of single OP, and the third one deals with five OPs. It needs to be said that all seven OPs are implemented by units located in the same ministry.

As one may expect, the number of evaluation units reflects the level of (de)centralization. Single units in Hungary, Slovenia, and one subdivided into three in Romania is contrasted with 5 in Croatia, 7 in Bulgaria, 12 in Czechia, 15 in Slovakia, and 34 in Poland¹.

Central coordination body (CCB)

In the majority of decentralized systems the network of evaluation units is complemented with and linked by CCB. The scope of activities and the extent to which CCB attempts to regulate evaluation practices differs.

Active CCBs operate in Poland, Czechia and Slovakia. They establish working groups and organize their cyclic meetings, organize evaluation conferences, provide training, organize postgraduate

¹ 18 in Poland if evaluation units at the level of regional government are excluded.

courses (Poland), conduct knowledge sharing activities, assist in public procurement procedure (e.g. elaboration of TORs), support the process of designing evaluation plans.

What distinguishes Polish CCB from the two others is the desire to formally regulate some practices of evaluation unit. While in Czechia and Slovakia CCBs only counseled in the process of drafting evaluation plans, the Polish one formally approved those documents.

Czech and Slovakian CCBs provide some general guidelines on how to assess the quality of evaluation studies. Polish CCB developed standard check list, and require all evaluation units to complete it after every study and upload to the central database. Recommendation follow-up is another obligation of Polish evaluation units fulfilled through central database. Similar solution has been implemented in Czechia, but it was still not operational while we conducted our study². In Slovakia there is only a paper guidelines on this subject, and no actual effort to impose them.

All three – Polish, Czech and Slovak CCB also act as evaluation unit responsible for providing studies on horizontal subjects, as well as evaluating CP effectiveness at the level of Partnership Agreement.

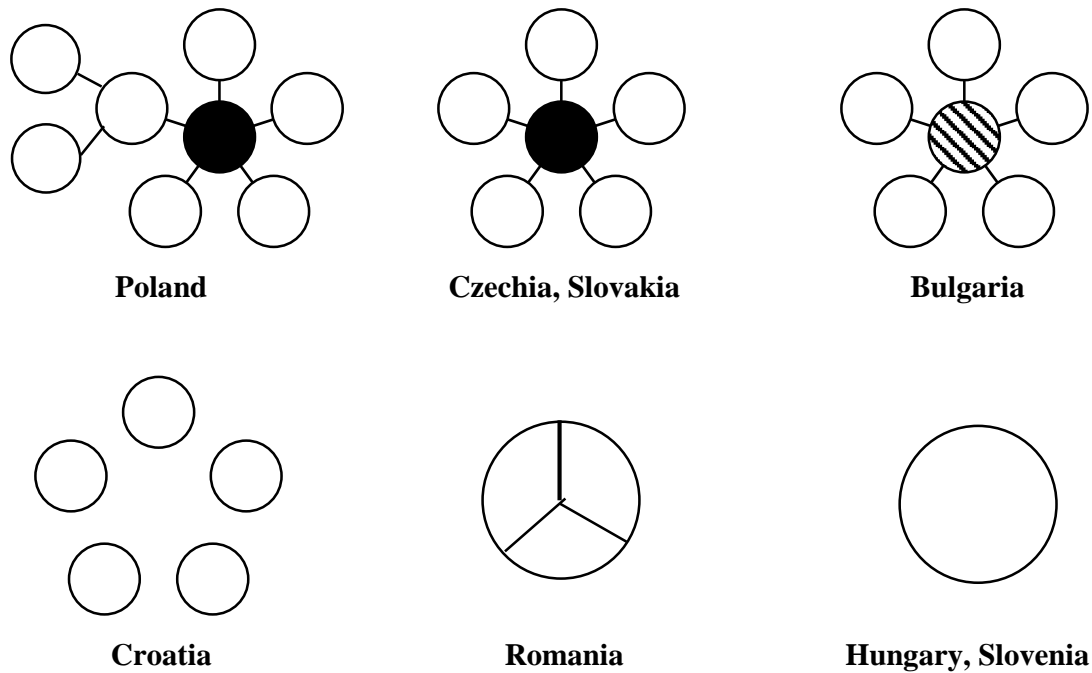
In striking contrast to the three described CCBs is the Bulgarian one. Although present in the system it is not of much use for evaluation units. Its activities are limited to forwarding communications from EC and ensuring conformity to the EC requirements. However evaluation units are left to themselves when it comes to designing evaluation, disseminating results, organizing follow-up processes, etc. No support or training is provided by CCB.

Croatia is the only country with decentralized evaluation system that does not (as of yet) include CCB³, leaving evaluation units on their own.

² The Polish one is in force since 2007-13 perspective.

³ It has to be mentioned that since Croatia joined the EU only in 2013, the evaluation system (as well as the whole CP implementation system) is relatively new and still in the process of formation. The establishment of a CCB was planned in the forthcoming months.

Figure 2 Structure of national evaluation systems



○ - evaluation unit ● - central coordinating & evaluating unit ◐ - central coordinating only unit

Source: own elaboration

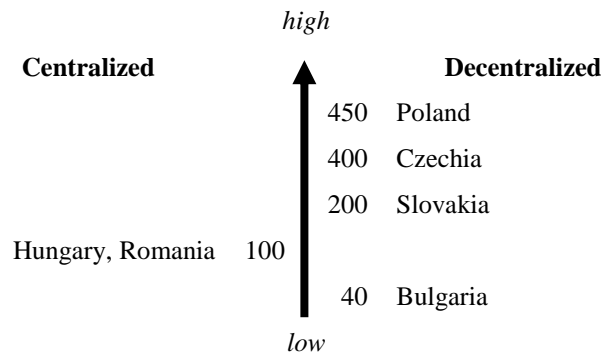
b. Variations in performance

We will discuss the differences in performance of evaluation systems taking the degree of decentralization as the main dividing line, where Hungary, Romania and Slovenia are the centralized systems, and the rest are decentralized. In some cases we will however exclude some countries from the comparison. Explanation for that will be given.

Number of completed evaluation studies is the first obvious difference between analyzed systems. During the programming period 2007-2013 around 100 studies were completed in centralized systems of Hungary and Romania. At the same time 180 studies were completed in Slovakia, over 400 in Czechia, and 450 in Poland. Bulgaria is the exception among decentralized system with only 40 studies⁴. Even including that the dominance of decentralized systems in terms of the number of completed studies is apparent.

⁴ Croatia and Slovenia cannot be compared to the other 6 countries as the scale of cohesion policy transfers and the system adopted for implementing them is considerable smaller.

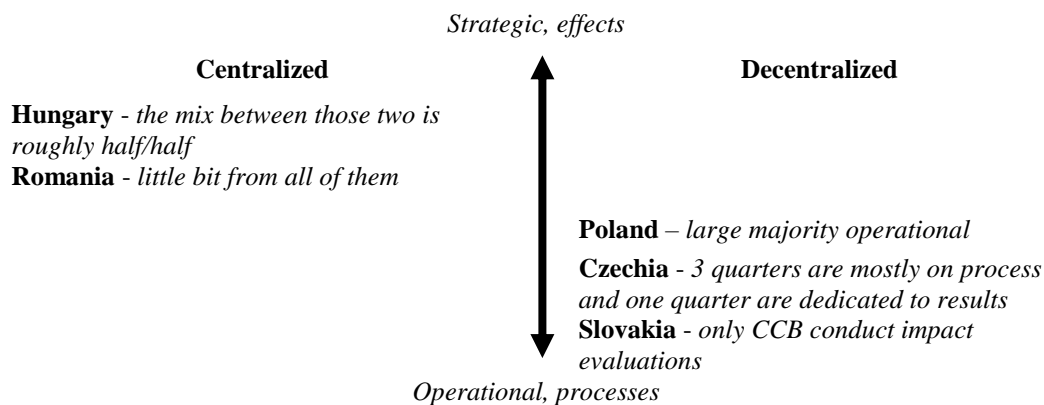
Figure 3 Number of completed evaluation studies



Source: IDIs

We have asked local experts (from central coordinating bodies and national evaluation societies) about the proportion of summative evaluation studies (focusing on effects of intervention), and formative (focusing on the process of implementation). To the extent we may rely on those opinions we can see that there is much more of strategic orientation in centralized systems of Hungary and Romania, while in the decentralized ones process evaluation dominates by large.

Figure 4 orientation of evaluation studies – perception of local experts



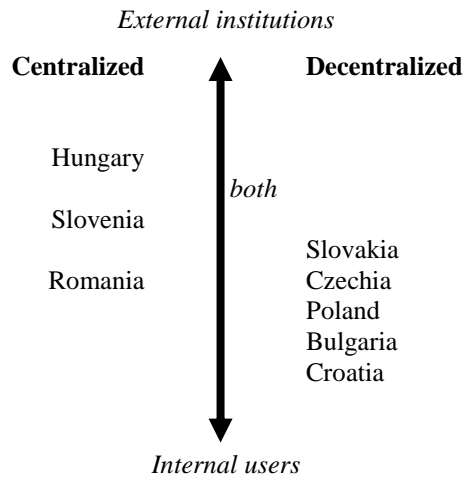
* Slovenia – no answer, Bulgaria & Croatia – no clear answer, but operational seems to dominate

Source: IDIs

The latter corresponds with the dominant target audience evaluation units focus on. For Hungarian evaluation unit external audience - other institutions in cohesion policy implementation system, domestic public institutions dealing with other policies, and institutions at EU level – are as much important as internal users - managers of other units and senior public administration staff in the same institution (e.g. department directors). In all other countries internal users are more important. The four countries where the focus on internal users is the strongest are the ones with decentralized system.

Order described above and presented in figure X is based on average values for all evaluation units in particular country. However the differences are small, and when we test the difference between samples of single evaluation units from centralized and decentralized systems it appears statistically insignificant (Mann-Whitney U Test).

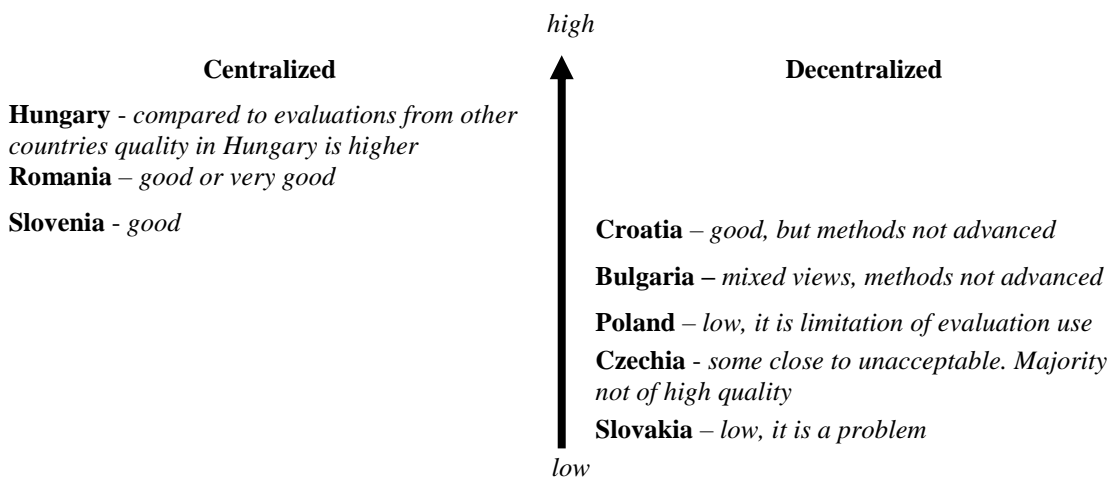
Figure 5 Target audience of evaluation findings



Source: CAWI

It is very risky to compare quality of studies based on subjective views of local experts, who certainly have different expectations and perceptions of quality depending on country. It is probable that in countries where the practice of evaluation is more mature (e.g. Poland, Czechia, Hungary) expectations are higher, and the experts are more demanding. At the same time in countries where the practice is less developed (e.g. Slovenia, Croatia, Bulgaria) expectations are lower, and assessment more lenient. Yet again we believe there is a difference between centralized and decentralized systems, with the quality higher in the former.

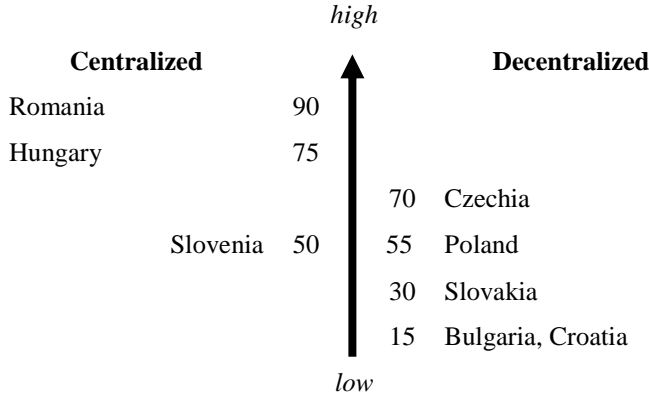
Figure 6 Quality of studies – perception of local experts



Source: IDIs

The quality of evaluation depends on many factors. Capacity of contractors is one of the most important. However the capacity of institutions commissioning, and then approving reports is also crucial. We could not measure how competent is staff of evaluation units. We could however compare the share of time spent on evaluation, which approximate to what extent civil servants dealing with evaluation may specialize in this subject.

Figure 7 Share of working time spent on evaluation (%)



Source: CAWI

In all studied countries units dealing with evaluation have also other responsibilities. This includes other analytical works, but also less related to evaluation like formulation and implementation of programs or information and communication tasks. The proportions vary greatly between countries. While other tasks amount to only 10% of working time in Romania, and ¼ in Hungary, they consume around 85% of time in Bulgaria and Croatia. We may observe that it is easier to specialize in evaluation in centralized systems, but again the difference tested on single units is not statistically significant (Mann-Whitney U Test).

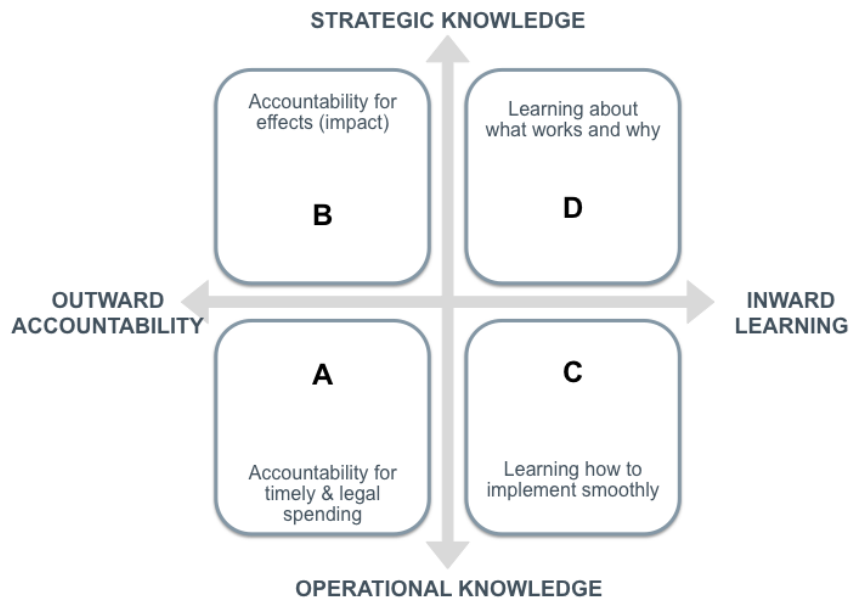
To conclude, we believe the decision to establish centralized or decentralized evaluation system has an impact on the operation of evaluation units constituting the system. It is an additional factor that differentiates the practice of evaluation and its outcomes, i.e. evaluation function and use. It may explain why organizations experiencing the same external pressure (regulations imposing evaluation), and the same internal propensity (dealing with the same type of tasks, implementing similar intervention), practice and use evaluation in different ways.

To illustrate the differences observed in the example of 8 national evaluation systems of cohesion policy we introduce the framework of four possible orientation of evaluation systems developed by Olejniczak et al. (2018). It is a resultant of two choices:

- the dominant type of knowledge delivered by evaluation unit;

- the primary purpose and target audience.

Figure 8 Orientation of evaluation system



Source: Olejniczak et al. (2018)

Summing up all observations presented above it is clear that decentralized systems are oriented on option C. The large majority of knowledge provided is operational, and the dominant target users are internal actors - managers of other units and senior public administration staff in the same institution. So the good news is the systems are focused on learning. The less optimistic is that this learning is mostly single loop, improvement of implementation process.

Centralized systems are offset in the direction to option B. There is more strategic knowledge, and external actors (other institutions in cohesion policy implementation system, domestic public institutions dealing with other policies, and institutions at EU level) are as much important as internal ones. That suggests that centralized systems are much more focused on accountability for effects. At the same time they are less concerned with internal users, which makes it more difficult to provide appropriate knowledge to support decision-making process.

Discussion

There are two issues we would like to discuss in this section. First are the factors determining the structure of national evaluation system and the second are the methodological limitations of our study.

a. Factors determining the structure of evaluation system

As we have presented above cohesion policy evaluation systems at national level varies greatly in structure, and role of central coordination body. We have shown that it also leads to differences in orientation of evaluation systems – provided knowledge, dominant target audience, orientation on accountability or learning. As a result although the intention of the European Commission is that evaluation should serve as an element of learning and improving the effectiveness of the implemented cohesion policy, this goal is clearly not achieved in the same way in individual Member States.

We believe that the way the evaluation systems are organized may result from specifics of the administrative system of the Member States, which in fact constitutes an immanent part, as well as from the attitude to evaluation, the internal tendency to evaluate and result in large variations in the types of use, and even the lack of use. In our opinion formal and informal characteristics of public administration systems should be considered in further exploration of the phenomena.

We propose to concentrate on four aspects of the national policy settings and internal institutional context, which – in general affects the effectiveness of public administration (see: Curristine, Lonti and Joumard 2007) – and may be responsible for differences in cohesion policy evaluation systems between EU countries.

The first one is related to the size of central government, which in fact determines the degree of decentralization in a given country. We may analyze total public expenditure as % of GDP, “core public administration” employment in total employment or number of ministries to define the model of organization of central power. The debate on the size of public sector has led some public policy analysts to conclusion countries with “small” public sectors on average report the highest scores for overall performance, and especially for administrative and economic performance (Afonso, Schuknecht, Tanzi, 2003). This may also refer to the effectiveness of evaluation systems.

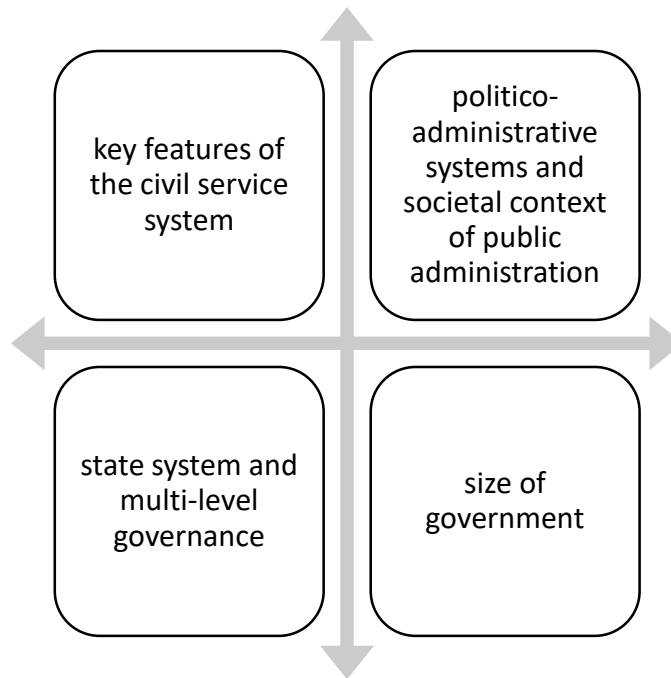
The second aspect concentrates on scope and structure of government. To explain the differences in evaluation systems, it may be important to identify core features of the state system, especially the multi-level governance of allocating government responsibilities and competencies to different tiers of government (distribution of power between the different government levels related to different policy areas). As previous studies shown the way the “machinery of government” is structured, affects policy capacity. Some countries around the world – like Australia and Canada for example – by reducing the number of separate ministries (departments) by consolidating previously separate but closely related ministries, has greatly improved policy co-ordination (Keating 2001). Within this area we should consider number of administrative tiers, distribution of

competences between government tiers, number of local governments or structure of executive of central government (consensual/intermediate/majoritarian).

The third aspect concerns features of the Member States' civil service systems. Analyses of the status and categories of government employees (like share of civil servants in public employment system) but also identification of the national systems based on categories such as type of HR system in public service sector (career-based/position-based/dual), openness of civil service (open/hybrid/closed), coherence of human resources management across different levels of government (high/medium/low), employee turnover rate in public sector (high/medium/low) may help to understand the differences in organization of cohesion policy evaluation system. It is also important to examine key characteristics of the central government human resources management system such as compensation level in public sector compared to the private sector, professionalism of civil service or overall assessment of civil service systems and HRM.

The last area is related to the broadest concept of the influence of the politico-administrative system and the societal context of public administration and the administrative tradition and culture. In this field the feature that may cause the differences in evaluation systems functioning are related to the process of decision-making including sources of policy advice, coordination and fragmentation of policy-making, managerial vs. procedural public administrations, regulatory density (red tape), transparency and accountability. Other important characteristics derive from the administrative tradition and cultural background in which the administration is embedded. Hence, administrative culture indicators by Hofstede, perception of corruption, trust in government should be analyzed.

Figure 9 Four areas of formal and informal characteristics of public administration systems to be analyzed



Source: own elaboration

The data needful to analyze these four areas will be gathered from several sources (i.e. Eurostat, EUPACK project final report, “A comparative overview of public administration characteristics and performance in EU28” European Commission Directorate-General for Employment, Social Affairs and Inclusion Report 2018).

b. Methodological limitations of our study

There were a few issues that made our analysis difficult and may rise doubts about what we have actually observed.

First is the number of evaluation units we studied. The sample amounts to $n=74$ and it is actually very close to total population of evaluation units in studied countries ($N=80$). What causes the problem is the great imbalance between units from centralized (5) and decentralized (69) systems. This imbalance is obvious. While in centralized systems we have one evaluation unit, in decentralized there is a number of them. It does however complicate the analysis of differences. Although we observe differences in country averages, we have no means to test their significance comparing single units. 5:69 sample seems too small to rely even on non-parametric tests like Mann-Whitney U Test.

Our second doubt are the compounding / extraneous variables. We have focused on the degree of decentralization as our independent variable, but there are others characteristics of evaluation systems that with high probability influence our dependent variable – orientation of evaluation system. One of them is described in the study. It is the existence and activity of Central Coordinating Body. The other one but probably even more important is the maturity of evaluation system. In less matured systems of Croatia, Slovenia less studies are completed. What is more important we believe there is substantially different perception of quality of studies, different understanding of what is strategic and operational, different approach to the role of evaluation unit in the process of commissioning, conducting and approving evaluation study. Immaturity might also enhance orientation toward accountability (that is in line not only with our intuition but also the literature on the subject [e.g. Raimondo, 2018]). However to eliminate this variable we would have to reduce our analysis to 4 countries – Poland, Czechia, Slovakia and Hungary, which would in turn even increase the problem of sample imbalance.

References

1. Christie C.A. (2007), “Reported Influence of Evaluation Data on Decision Makers’ Actions: An Empirical Examination”, *American Journal of Evaluation*, 28(1), 8–25.
2. Mark M.M., Henry G.T. (2004), „The Mechanisms and Outcomes of Evaluation Influence”, *Evaluation*, 10(1), 35-57.
3. Van der Knaap, P. (1995). *Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation?* *Evaluation*, 1(2), 189–216.
4. Weiss, C.H. (1972) *Utilization of evaluation: Toward comparative study*, in C.H. Weiss (Ed.) *Evaluating action programs: Readings in social action and education*. Boston, MA: Allyn & Bacon.
5. Weiss, C.H. and Bucuvalas, M.D. (1980) *Social science research and decision-making*. New York: Columbia University Press.
6. Wojtowicz D., Kupiec T. (2018). Reluctant to Learn? The Use of Evaluation to Improve EU Cohesion Policy Implementation in Polish and Spanish Regions. *Evidence & Policy: A Journal of Research, Debate and Practice*, v14 n1 p.103-120.
7. Alkin, M.C., Daillak, R., White, P. (1979) *Using evaluations. Does evaluation make a difference?* Beverly Hills, CA: Sage.
8. Alkin, M.C. and Coyle, K. (1988) *Thoughts on Evaluation Utilization, Misutilization and Non-Utilization*. *Studies in Educational Evaluation* 14(3): 331-340.
9. King, J. A., & 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*.
10. Huberman, M., M. Gather Thurler (1991) *De la recherche à la pratique: Eléments de base*. Berne: Peter Lang.
11. Shulha L.M., Cousins J.B. (1997), “Evaluation use: Theory, research, and practice since 1986”, *Evaluation Practice*, 18(3), 195–208.
12. Johnson, K., Greenesid, L.O., Toal, S.A., King, J.A., Lawrenz, F., Volkov, B. (2009) *Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005*. *American Journal of Evaluation* 30(3): 377–410.
13. Cousins, J.B. and Leithwood, K.A. (1986) *Current empirical research on evaluation utilization*. *Review of Educational Research* 56(3): 331-364.

14. Brandon PR and Singh JM (2009) The strength of the methodological warrants for the findings of research on program evaluation use. *American Journal of Evaluation* 30(2): 123–57.
15. Ledermann, S. (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178.
16. Balthasar A. (2006), „The effects of Institutional Design on the Utilization of Evaluation. Evidenced using Qualitative Comparative Analysis (QCA)”, *Evaluation*, nr 12(3), 353–371.
17. Balthasar A. (2009), „ Institutional Design and Utilization of Evaluation A Contribution to a Theory of Evaluation Influence Based on Swiss Experience”, *Evaluation Review*, nr 33(3), 226-256.
18. Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26–41.
19. Rist R., Stame N. (eds.) (2006) *From Studies to Streams: Managing Evaluative Systems*, London, Transaction Publishers.
20. Leeuw F.L., Furubo J.E. (2008), „Evaluation Systems: What Are They and Why Study Them?”, *Evaluation*, 14(2), 157-169.
21. Hanberger A. (2011), “The real functions of evaluation and response systems”, *Evaluation* 17(4), 327–349.
22. Højlund, S. (2014) Evaluation use in the organizational context - changing focus to improve theory. *Evaluation* 20(1): 26–43.
23. Eckerd A, Moulton S (2011) Heterogeneous roles and heterogeneous practices: understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation* 32: 98–117.
24. Boswell C (2008) The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy* 15: 471–88.
25. Wojtowicz D., Kupiec T. (2018). Reluctant to Learn? The Use of Evaluation to Improve EU Cohesion Policy Implementation in Polish and Spanish Regions. *Evidence & Policy: A Journal of Research, Debate and Practice*, v14 n1 p.103-120.
26. Williams, B. and I. Imam , eds (2007) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes: EdgePress of Inverness.
27. Olejniczak, K., Raimondo, E., Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation*, 22, 168–189.