Towards System oriented Innovation Policy Evaluation?

Evidence from EU28 Member States

Susana Borrás, Copenhagen Business School, Denmark

Mart Laatsit, Copenhagen Business School, Denmark

This is the pre-print version of the journal article published as:

Borrás, Susana and Laatsit, Mart (2019): "Towards system oriented innovation policy evaluation' Evidence from EU28 member states" in **Research Policy** Vol 48 (1), pp. 312-321
https://doi.org/10.1016/j.respol.2018.08.020

Abstract:

Many years after the introduction of the innovation system concept in innovation policy design, it is still not clear whether policy evaluation practices follow a system approach. Building on evaluation and innovation studies, this article develops the concept 'system oriented innovation policy evaluation' based on four attributes (coverage, interaction, temporality and sources). The attributes are used as analytical devices for gathering extensive empirical evidence on the actual practices of EU28 member states. The findings show that few countries have developed the most complete type of system oriented evaluation, the one we call "holistic". Another small group does not have any system oriented evaluation; while most other EU28 countries have some traits of it ("flexible") or are beginning to develop one ("starter"). The advent of a system approach to evaluation offers the opportunity of comprehensive, contextualized, and evidence-based innovation policy-making. However, there are still serious obstacles as such an approach requires important knowledge and organisational capacities. Overcoming these obstacles would need more decided evaluation capacity-building at the national level.

Keywords: Evaluation, innovation policy, innovation system, innovation indicators, evidence-based policy, European Union, innovation system, holistic.

Acknowledgements: The authors would like to acknowledge the Foundation Archimedes from Estonia, for generous funding of Maart Laatsit with grant number is 16-4.2/773.

1. Introduction

During the past decades there has been an increasing focus on the need to provide innovation policy-makers with more comprehensive and knowledge-based tools for policy-making. The complexity of innovation systems and the recent developments in new policy initiatives require more sophisticated and advanced intelligence skills and knowledge in policy-making. Further improvements in science and technology foresight, technology assessment and, innovation policy evaluation are largely seen as the three key areas for creating strong, comprehensive and strategic policy advice (Kuhlmann, Boekholt et al. 1999). This paper focuses on one of these three, namely, innovation policy evaluation. In particular, the paper examines the changes of evaluation practices in that direction, to determine whether they are moving in the direction of system oriented evaluation practices.

In their quest towards obtaining more sophisticated and comprehensive inputs for decision-making, several European countries have been introducing novel approaches to evaluating innovation policy and innovation systems' performance in what seems to be a move towards introducing system oriented innovation policy evaluation'. That is, evaluations which are system oriented by means of assessing policy initiatives that encompass the entire innovation system. Anecdotal evidence indicates that some countries might have started developing such an approach to evaluation. However, there is a lack of broad-based empirical evidence about how far (and how) countries are developing their evaluation practices. A system oriented approach to evaluation is naturally rather ambitious, because it aims at obtaining concrete clues to (re)designing innovation policy interventions according to what works or not, and to the specific needs and problems in the innovation system.

One possible reason for the increasing acceptance of the relevance of the system oriented approach in evaluation practices is the mounting dissatisfaction with the mismatch between the traditional narrow focus of program and policy instrument evaluation on the one hand, and the needs of policy-makers for evaluations to be more system oriented. Policy-makers seem to have embraced the notion of innovation system when defining innovation policy interventions (Edquist 2005) (Kuhlmann, Shapira et al. 2010). However this is not the case for most evaluation practices, which traditionally focus on individual programmes rather than on the system oriented-perspective (Arnold 2004).

As we will review in the next section, most scholarly publications have addressed this matter from a normative perspective, suggesting specific models for how policy-makers could go about it. Some of these normative models suggest the integration of different innovation policy evaluations to obtain a coherent overview (Edler, Ebserberger et al. 2008) (Magro and Wilson 2013). Others put more emphasis on integrating the results of policy evaluations with insights about specific

problems and bottlenecks in the innovation system (Arnold 2004) (Jordan, Hage et al. 2008) (Hage, Jordan et al. 2007).

In spite of the relevance of these normative models for how to conduct the evaluation, we still do not have empirical studies substantiating whether or not European countries are in fact organizing and conducting system approach evaluations, and if so, how they actually implement them. In other words, we still lack empirical evidence about the actual efforts currently made by policy-makers, and about the types of practices that are currently emerging across these countries (Martin, Nightingale et al. 2012). Building on the above-mentioned scholarly approaches, this article develops the concept 'system oriented innovation policy evaluation', which is based here on four distinct attributes: coverage, interactions, temporality and sources. These attributes are used as analytical dimensions to gather and characterize empirical evidence about the actual evaluation practices of EU's 28 member states. Hence, the leading research question of this paper is: How far, and if so how, are EU 28 member states developing system oriented innovation policy evaluation?

The paper proceeds as follows. After reviewing the literature on this topic in section 2, section 3 builds from there and provides a clear-cut definition of system oriented innovation policy evaluation based on the four above-mentioned attributes. Those attributes are operationalized in order to undertake an orderly empirical analysis, and the data sources and some important methodological considerations of the analysis are reflected upon. Sections 4 and 5 present the analysis, first looking at how the EU28 countries perform in terms of each of the four attributes (section 4), and thereafter looking at a four-fold typology characterizing each of the 28 EU countries. The conclusions summarize the findings, pointing out cross-national diversity, and discussing the need for capacity building as a way of overcoming actual obstacles.

2. Models in the Literature

During the past two decades the innovation system approach has gained substantial endorsement among scholars and policy-makers alike. This approach sees innovation as a complex social process of a cumulative nature, embedded in complex institutional and organizational national contexts (Lundvall 1992) (Nelson 1993). It brings forward the notion of innovation as the outcome of complex interactions and dynamics in the idiosyncratic socio-economic context of an economy. Yet, the more the innovation system approach has gained the upper hand, the more policy-makers have grown dissatisfied with the limitations of current evaluations of individual R&D programs. With the appearance of more complex policy systems and mixes (Smits and Kuhlmann 2004), the innovation system approach has made apparent the need for advanced tools that provide knowledge for evaluation at the system level. In particular, the system approach of innovation policies (Edquist 2011) with the focus on interaction and interactive learning between organizations (Lundvall and Borrás 1998) has raised awareness about the need for more

sophisticated tools to enable policy-makers to better grasp the complexity of the impact of policy instruments and decisions.

In his seminal paper about the new frontiers of evaluation studies, Irwin Feller (2007) reflected upon this growing dissatisfaction between the needs of policy-makers for more encompassing approaches stemming from the innovation system approach on the one hand; and the conventional praxis of research evaluation of individual R&D programs on the other. Single evaluations are increasingly perceived to be too limited to provide answers regarding the impacts of public initiatives in the wide framework of the economy. "Existing evaluations touch only lightly, however, on how the strategies, behavior, performance of the sectors or actors described in the national innovation taxonomy change as a result of the cumulative, long term impact of a cluster of programs" (Feller 2007).

In their review of the literature Molas-Gallart and Davis argue that "the practice of policy evaluation continues to lag behind advances in innovation theory. Innovation theory has produced successive generations of more sophisticated conceptual models that seek to explain how the relationship between scientific and technological research and the market opportunities for innovation occurs." (Molas-Gallart and Davies 2006). However, much of the evaluation undertaken today is still performed at the project and program level, and is mainly based on simple models of impact assessment and accountability. Yet, how to aggregate and integrate findings relating to specific policies and programs into an overarching framework that evaluates the effects policies within the national innovation systems is not an easy task. The innovation system approach and the theoretical framework it implies "have proved difficult to use in the practice of evaluation, resulting in a gap between evaluation practice and Science Technology Innovation (STI) policy theory" (Molas-Gallart and Davies 2006).

In response to the need of evaluation to move beyond the myriad of isolated individual program-focused evaluations, a few models have been suggested different ways to provide more encompassing insights to guide policy-makers' decisions. In the earliest work about this theme, Arnold (2004) suggests how to develop research and innovation policy evaluation in a systems world. He proposes an approach to evaluation that considers "to a greater extent the interplay of these tools with their environments" p.2. His model combines three levels: the traditional program evaluation, whose scope needs to be expanded to aim at identifying regularities across programs through meta-evaluations; the evaluation of the health of the innovation system based on a series of system-wide dimensions (such as the innovativeness of the business sector, adequacy and provision of infrastructures, the regulatory framework conditions for innovation, etc); and sub-systems evaluations, which target specific possible bottlenecks at a meso-level (policy mixes, or institutions performance). (Arnold 2004).

A similar multi-level model for innovation system evaluation is proposed by Jordan and Hage. Mainly focusing on developing an epistemological and indicator-based model, these authors distinguish between the micro- meso-levels (Jordan, Hage et al. 2008), and macro-level (Hage,

Jordan et al. 2007). Building from Arnold, the authors aim at outlining "a theories-based innovation systems framework (ISF) of indicators for RTD evaluations that can aid government policy makers in policy formulation and *re*formulation. The indicators that are proposed suggest protocols for performance monitoring and evaluation" (Jordan, Hage et al. 2008) (p. 118). They take knowledge production as the starting point for a theory of innovation systems and of its evaluation: "The use of three levels responds to the call for a knowledge production theory with these three analytical levels and gives the opportunity to contribute to other theories and frameworks, such as organizational learning, knowledge communities, and econometric input-output evaluation models" (Hage, Jordan et al. 2007) p. 733.

Other approaches focus instead on the nature of the material upon which the evaluation is currently based, suggesting a model based on a novel use of evaluations that provides a better overview of policy performance in the innovation system. In this sense, Edler et al (2008) suggest "using existing evaluations to learn about policy performance and policy effects on the system level". Inside this frame they separate two concepts, namely, evaluation synthesis and meta-analysis, both of which serve as the basis for an overall framework for utilizing and analyzing existing evaluation data. Evaluation synthesis is understood as "an aggregated content analysis based on multiple evaluation reports on similar programs or projects" (Edler et al 2008). For its part, meta-analysis allows for "an improved comparison and understanding of interventions and their effects by taking into account the results of a large number of evaluations" (Edler et al 2008). Hence, whereas the former aggregates and synthesizes existing evidence, the latter provides the basis for contextualizing such evidence in a broader context, allowing for more strategic insight and overview.

A somehow similar approach has been suggested by Magro and Wilson (2013), who focus on "meta-evaluations or secondary analyses that build on individual evaluations in trying to capture the system oriented nature of policies; moving ahead from isolated, individual evaluations". In that respect, they share a common viewpoint with Edler et al (2008) focusing on the policy space, or, more concretely, on "the innovation policy system as the conjuncture of policy mix and multi-level dimensions" (p. 1647). They use this model in one case study, conducting an evaluation mix of the Basque Country innovation policy. The starting point of their model is the identification of individual policy rationales and their corresponding instruments. Hence, the evaluation mix protocol that they suggest is the practical articulation of how to conduct this evaluation in a way that brings together the focus of policy mixes with the recent calls for more system oriented approaches to evaluation.

From the above we can see two main approaches. Whereas Arnold and Jordan-Hage tend to take the starting point in the innovation system striving to identify specific indicators and contents for an encompassing system-focused policy evaluation framework, Edler et al. and Magro & Wilson for their part begin with pre-existing innovation policy evaluations and suggest active efforts and means to generate synthesis and meta-analysis from them to create a comprehensive system oriented perspective where policy mixes seem to be the core items. These two approaches are

relevant in terms of providing an initial response to dealing more effectively with this matter. Most importantly, they provide specific understandings of the ways in which policy-makers might address the complexity of both the evaluation of innovation policy and the assessment of innovation systems within a comprehensive analytical overview to help policy-makers make decisions.

However, in spite of its practical suggestions to policy-makers, the literature still needs to come to grips with a single conceptual apparatus that can provide a clear analytical framework for undertaking empirical investigations of the actual evaluation practices of countries. The next section does this.

3. Investigating the practices of 'System oriented Innovation Policy Evaluation'

3.1 Definition and Operationalization

For our current purpose of examining the empirical evidence on the practices of EU28 countries, the models reviewed in the previous section can provide initial conceptual basis. In particular, these models serve to distinguish between different forms of aggregating the results of evaluations, different forms of "analysis of impact of system health" on the bottlenecks of innovation systems, and of incorporating interactions between innovation policy performance and innovation system performance into the evaluations. However useful these normative models, there is still a need to develop the next step, namely, to develop an analytical framework for studying empirically the current country practices. More concretely we need to define the concept of system oriented evaluation in a way that allows an empirical analysis. This is important because we need to be able to identify clearly whether or not a system oriented innovation policy evaluation. The attributes that we assign to this concept aim to provide such clarity, since they are based on the theories in the innovation system and evaluation literature. A clear definition and its operationalization will then allow us to grasp the complexity of the empirical reality, while avoiding the classical problems in the social sciences of 'conceptual stretching' (Sartori 1970). Likewise, a clear concept is important for clarifying the specific attributes that define it, and hence the analytical dimensions required to undertake empirical studies, and to characterize the diversity of forms that have unfolded as countries have developed and organized their own evaluation practices.

To be sure "evaluations are used to inform policy-makers, program managers and other stakeholders about the effectiveness, efficiency, appropriateness and impact of policy interventions" (Edler, Ebserberger et al. 2008) p. 175. Following from this and the discussion above, we define 'system oriented innovation policy evaluation' as the set of evaluation practices which exhibit a wide coverage of evaluation elements, an integrative approach to horizontal and transversal interaction between different levels of policy and system performance, are conducted

regularly, and draw on different sources of expertise. By this definition we specify that the purpose of system oriented innovation policy evaluation is to provide an overall, critical and strategic overview of the performance of the system and of innovation policies therein. The ultimate purpose is to guide future strategic public intervention, which include policy interventions in the fields of research, science, technology and innovation.

Our definition can be seen as an 'ideal type': a notion that defines the general traits of the expected phenomena, and which is used for analytical purposes (Goertz 2006). Ideal models are formed deductively from theorizing endeavors and aim at providing clear guidance for empirical analysis. However, because they are 'ideal' they might not be found in their 'purity' or 'entirety' in the real world. They are abstractions, and may not necessarily to be found 100% replicated in the empirical complexity of social phenomena.

For this reason, we rarely expect to find countries carrying out ideal types of system oriented evaluation, because it is very demanding given the complexity of the task. Instead, in our empirical analysis we expect to find only few countries which are conducting a rather strong form of 'system oriented policy innovation evaluation' in the sense or complying in an assertive manner with the four attributes that define our ideal model (see Table 1 below). Most likely, none of these few countries will be able to cover in its entirety and in totality the complexity of an innovation policy and innovation system as such. What we expect among them, however, is that they have developed ways to gather and manage a sufficiently large amount of information regarding innovation policy evaluation results in relation to the most significant and relevant problems that afflict their innovation system performance.

Table 1: The four attributes defining the concept "system oriented innovation policy evaluation", their operationalization and measurement.

Definition of the attributes	**Operationalization for empirical analysis**	**Measurement* scores**
Coverage: The extent to which the evaluation covers three most important elements (see the cell to the right)	We examine one-by-one whether countries are devising evaluation tools for the following three elements: - Innovation policy Instruments - Innovation policy mixes - Socio-economic assessment	Value 2: when there is a substantial number and sophisticated forms of evaluations Value 1: fewer numbers of evaluations and less sophisticated Value 0: very few or none of the above

Interaction: The extent of multi-level interaction between different levels of innovation policy evaluations (individual instruments and mixes), and of transversal interaction between innovation policy performance and innovation system performance	We examine whether or not countries have strategic reviews (reviews with an overview and dealing with interactions), and whether those strategic reviews include both multi-level and transversal interactions, or only one interaction	Value 2: Strategic reviews include both multi-level and transversal analysis Value 1: Strategic reviews only include one of the two (either multilevel or transversal) Value 0: No strategic reviews
Temporality: The extent of regularity in the evaluation in all the three coverage elements	We examine whether countries have conducted evaluations on a regular basis	Value 2: Evaluations are conducted with a high level of regularity Value 1: Some evaluations are conducted regularly, but others more sporadically Value 0: Evaluations are done sporadically and ad-hoc
Source: The extent to which different sources are involved in conducting evaluation of the three elements above	We examine whether countries use diversified sources of evaluation, particularly the combination of national and international, internal (ministerial/public) and external (private consultancies, universities, think-tanks, etc)	Value 2: when a country has a strong combination of national/international evaluations that are either internal/external to the government Value 1: when a country has significant record of only two of the above Value 0: when a country has only one or none of the above

*See section 3.2 on data and methodology, and Section 4 for more detailed operationalization of the different elements of "coverage".

The four attributes above define a system oriented innovation policy evaluation. The first attribute, **coverage**, refers to the extent to which the most important elements (areas) of evaluation are included. This attribute is inspired by earlier treatments in the literature that

examined how extensive the object of evaluation actually is (Dahler-Larsen 2012). In our study, we operationalize 'coverage' into three elements, namely, the evaluation of innovation policy instruments, of innovation policy mixes, and of socio-economic performance assessment.

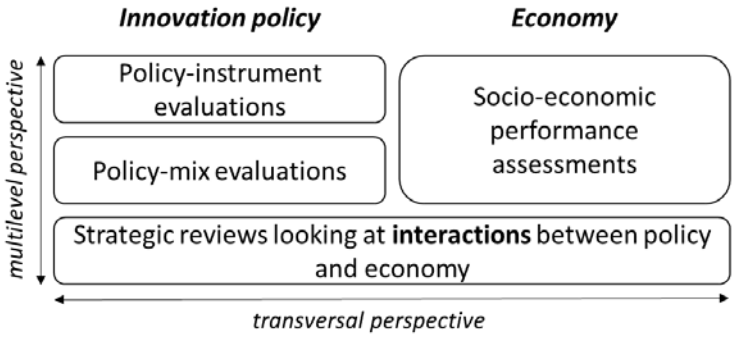
- By policy instrument evaluation we understand evaluation practices whose focus is to assess the impact of one particular innovation policy programme, for example, an instrument to assess the impact of an R&D program or of a tax incentive scheme.

- Policy mix evaluations, the second element, are the assessments of more than one policy instrument at once, and take into consideration their joint impact (additionality and complementarity). Policy-mixes have been considered of fundamental importance in understanding the performance of innovation policies (Flanagan, Uyarra et al. 2011) (Cunningham, Edler et al. 2016) and thus are highly relevant in the context of system oriented innovation policy evaluation.

- The third element, socio-economic performance assessments, refers to the appraisal of the impact of dynamics on the innovation system. These assessments typically include consideration of indicators such as business enterprise expenditure on research and development (BERD) or employment in knowledge-intensive activities, and assess their dynamics over time. They often discuss analytically the possible factors behind such developments. There is a wide variety of approaches to this kind of assessment, carried out with varying degrees of sophistication, ranging from simple reporting of indicators to far more sophisticated large-scale innovation performance assessments. It is important to note that merely collecting and publishing statistical data does not amount to a socio-economic performance assessment. Instead the 'raw' data has to be appraised in the national context to be considered a proper assessment.

The second attribute in system oriented innovation policy evaluations is the **interactions** among the elements of 'coverage'. This attribute is important for theoretical reasons. Theory holds that national systems of innovation are based on two dimensions, namely, the institutional set-up (formal and informal rules of the game and framework conditions) and the socio-economic dimension – the production sector (Lundvall 1992). Hence, innovation policy can be considered part and parcel of the innovation system. For this reason, evaluations that are system oriented would invariably include assessment of the interactions between innovation policy (and their multiple levels, from individual instruments to policy mixes), and innovation system performance. Therefore, in order to determine whether a country assesses such interactions, we must first determine whether that country has conducted any strategic reviews, that is, advanced and comprehensive reports of a strategic nature that address those interactions. More concretely, we must determine whether those strategic reviews study multi-level and transversal interactions, or only one of these interaction. Figure 1 represents these multi-level and transversal interactions.

Figure 1: Multi-level interactions (those between individual policy instrument evaluations and policy mix evaluations) and transversal interactions (those between policy evaluations and socio-economic innovation performance)



The third attribute that defines 'system oriented innovation policy evaluations' is **temporality**, namely, the extent to which there is a certain level of regularity in the evaluation in all three of the coverage elements. We examine whether countries have conducted evaluations on a regular basis. Temporality is an important attribute, because evidence-based policy- making requires not only that different parts of innovation policy are evaluated, but also that the body of information is regularly updated and improved by subsequent exercises. Admittedly, different types of evaluations might have a different temporality – for example, strategic reviews are often undertaken in relation to particular strategic events, such as after major policy overhauls; whereas, socio-economic performance assessment might take place regularly every year. To a large extent, the general level of evaluation activity of a country will determine the general temporality of the system oriented innovation policy evaluation practices.

Finally, the fourth attribute defining 'system oriented innovation policy evaluation' refers to the **sources** of the evaluations, namely, the different sources involved when conducting different evaluation elements. More specifically, we examine whether countries use diverse sources of evaluation, in particular, do they combine national and international sources (conducted by international organisations such as OECD, EU, World Bank), as well as internal (conducted by governmental units) and external (by private consultancies, universities, think-tanks, etc.). Recent studies about practices of instrument-level evaluation look at this (Edler, Berger et al. 2012); in addition, the theory of absorptive capacity stresses the importance of combining internal and external sources in organizational capabilities (Borrás 2011). In the context of our conceptualization of 'system oriented innovation policy evaluations' this attribute is particularly relevant because of the widespread competences needed to conduct the different elements of evaluations and to deal with the complexity of establishing a meaningful overview.

3.2 The Data and Methodology

Data about the system oriented nature of innovation policy evaluation practices in EU countries are not easily accessible. For this reason the strategy of the present study has been to use a sequential research design to collect different types of data as a means of obtaining solid empirical evidence. Firstly, we have conducted a total of 62 semi-structured interviews in all EU28 countries: 52 with high-ranked government experts, and 10 with academic/independent researchers. The interviews were conducted between January 2016 and June 2017, with at least 2 interviews per EU28 country (see Annex 1). They were based on a guideline with specific semi-structured questions related to the items conceptualized above. Open room for discussion allowed in order to gather additional relevant information. The second set of data used in this study was gathered from a number of directly relevant documents of each country's evaluation practices. The RIO database[1] (Research and Innovation Observatory) and the SIPER database[2] (Science and Innovation Policy Evaluation Repository) have been particularly valuable in this regard. Additional documents were sent by interviewees, or found by the authors on the Web. The information obtained from the interviews was triangulated with those documents. On the few occasions when there was a mismatch, we conducted additional interviews and searches.

The next step of the sequential research design was to assign specific values to each country's attributes (See Table 2). We assigned scores of 0, 1 or 2, according to the intensity in the data (see Table 1 with the codebook). The coding of the data was performed very carefully and repeatedly by the two authors, in an internal working procedure similar to inter-coder reliability practices. We aggregated these scores into a total score, which go from the highest score, 12, to the lowest, 0. Countries with total scores of 12, 11 or 10 belong to the "holistic type", those with scores 7,8 or 9 belong to "Flexible Adapters", those with 6,5 or 4 to "Late Starters", and the countries with 3,2,1 or 0 total scores are deemed to have no system oriented innovation policy evaluation (see section 5 for details).

After the full analysis of the data (assignation of scores and types), we verified the findings between September and October 2017 using feed-back from national experts in the field. The findings were subsequently checked by the authors. The verification focused on eliminating possible misunderstandings or misinterpretations of the data. Adjustments were introduced when needed (see Annex 1).

[1] https://rio.jrc.ec.europa.eu/en
[2] http://si-per.eu/

4. Empirical Evidence in EU28

This section provides empirical evidence about how EU28 countries are organising their evaluation practices, and provides the basis on which we determined the extent to which these practices can be defined as 'system oriented innovation policy evaluations'. We first analyse these data as they relate to the coverage attribute, as this will indicate the extent to which countries do in fact conduct evaluations of these different elements. Thereafter, we will examine the other three attributes - interactions, temporality and sources - giving a careful account of the current practices in the EU28.

Table 2. Scores related to the four attributes defining system oriented innovation policy evaluation

	Coverage			**Interactions**	**Temporality**	**Source (internal/ external)**	**Total score**	**Type**
	Instrument evaluation	Policy-mix evaluation	Socio-economic performance assessment					
The Netherlands	2	2	2	2	2	2	**12**	**Holistic**
Austria	2	2	1	2	2	2	**11**	**Holistic**
Finland	2	2	1	2	2	2	**11**	**Holistic**
Ireland	2	2	2	2	2	1	**11**	**Holistic**
Sweden	2	1	2	2	2	2	**11**	**Holistic**
Germany	2	1	2	2	2	1	**10**	**Holistic**
Denmark	2	2	1	1	1	2	**9**	**Flexible**
France	2	1	1	2	1	2	**9**	**Flexible**
Belgium	2	1	1	1	1	2	**8**	**Flexible**
Poland	1	1	1	2	1	2	**8**	**Flexible**
United Kingdom	2	1	1	1	2	1	**8**	**Flexible**
Estonia	1	1	1	1	1	2	**7**	**Flexible**
Lithuania	1	0	1	2	1	2	**7**	**Flexible**
Slovenia	1	0	1	2	1	2	**7**	**Flexible**
Latvia	1	0	1	1	1	1	**5**	**Starter**
Spain	1	0	1	1	1	1	**5**	**Starter**
Hungary	1	0	0	1	1	1	**4**	**Starter**
Czech Republic	0	0	1	1	0	1	**3**	**Starter**
Portugal	1	0	1	0	0	1	**3**	**Starter**
Bulgaria	0	0	0	1	0	1	**2**	**None**
Croatia	0	0	0	1	0	1	**2**	**None**
Luxembourg	0	0	0	1	0	1	**2**	**None**
Romania	0	0	1	0	0	1	**2**	**None**
Italy	1	0	0	0	0	0	**1**	**None**
Slovakia	0	0	1	0	0	0	**1**	**None**
Cyprus	0	0	0	0	0	0	**0**	**None**
Greece	0	0	0	0	0	0	**0**	**None**
Malta	0	0	0	0	0	0	**0**	**None**

4.1 Coverage

There is a wide diversity across EU28 countries in their extent of coverage of the three evaluation elements. Regarding ***policy instrument evaluations***, we have divided countries into three categories: countries where policy instruments are almost always evaluated, countries where policy instruments are sometimes evaluated, and countries where policy instruments are rarely evaluated or are simply monitored (not evaluated as such). In the first category we have the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, the Netherlands, Sweden and the United Kingdom. In these countries there is a strong tendency to evaluate every programme, and hence we assign them a score of 2 (see table 2). Some of these countries have rigid structures for evaluations, grounded in legal or quasi-legal acts. For example, in the Netherlands, evaluations of programmes are tied to the general budgetary framework and each instrument has to be evaluated at least once every seven years (The_Netherlands_Goverment 2014). In other countries, there is no specific legal obligation to evaluate every programme but they have a strong evaluation culture. For example, in Austria or the UK, there is a strong tradition of evaluating all innovation policy programmes, or a "general expectation" that all programmes should be evaluated (see interviewees 1, 62).

Another group has less developed traditions and fewer legal requirements to evaluate programmes, but still conducts/ these countries still conduct a considerable amount of policy instrument evaluation; such countries include Estonia, Hungary, Italy, Latvia, Lithuania, Poland, Portugal, Spain and Slovenia (score 1 in Table 2). Many of these countries assess the impact of their innovation policy instruments following the rules of the EU Structural Funds. While the 'EU rules' only prescribe some minimum requirements regarding the rigor of the evaluations, the countries in this group have developed approaches that exceed these minimum requirements.

Finally, countries for which there is very little evidence of conducting policy instrument evaluations (i.e. received a score of 0 in Table 2) are Bulgaria, Croatia, Cyprus, Czech Republic, Greece, Luxemburg, Malta, Romania and Slovakia. These countries typically resort to either the bare minimum required by the EU Structural Fund regulations, or their practices are closer to descriptive monitoring rather than real analytical evaluations. For example, the Czech Republic has established procedures for the "evaluation of finished programmes"(Office of_the_Government_of_the_Czech_Republic 2013), but in practice only basic output data of the programmes are reported (Verification 1).

As the second element of the coverage attribute, we look at ***policy-mix evaluations***. Our data show that policy-mix evaluations, being a relatively new phenomenon, are not as widespread as policy instrument evaluations. We have defined three groups of countries according to the level of their policy-mix assessments. Firstly, there are countries that have carried out assessments on additionality and/or complementarity in their policy mixes. Secondly, some countries have treated the issue of policy interactions on a smaller scale, often within the framework of other types of

evaluations. While these countries do not apply policy-mix evaluations in a pure form, they are addressing the issues relevant to policy-mix and such endeavours should thus be recognized. Thirdly, there are countries with very weak or no signs of policy-mix evaluations taking place.

In the first group we find Austria, Denmark, Finland, Ireland and the Netherlands. For example, in Denmark, the Danish Agency for Science, Technology and Innovation commissioned two studies to assess the effects and interactions of different programmes on firm performance (DASTI 2014) (Daly and Christensen 2016). In Finland, different meta-analyses are bundled together to gain insight into the policy-mix performance (interviewee 20). In Ireland the analysis of the policy-mix forms an integral part of their comprehensive programme of evaluations (Department_of_Jobs 2015). In the Netherlands, a policy mix analysis assessing the interactions between instruments has been carried out for the so-called top-sector policy, a strategic initiative launched by the Dutch government aimed at boosting the competitiveness of priority sectors through a combination of policy measures (interview 45).

The second group consists of countries where we have detected some signs of policy-mix thinking without full scale policy-mix evaluations: Belgium (Flanders), Estonia, France, Germany, Poland, Sweden, the United Kingdom. For example, the innovation agency Enterprise Estonia has been carrying out a biannual evaluation of the impact of its policy mix, addressing also the additionality effects of the policies (interviewee 18). In France, some of the interactions between policies have been covered in the evaluation of the "Programme d'Investissement d'Avenir" (interview 22). The countries in the third group, those that do not seem to assess the interactive effects of their policy-mixes, are Bulgaria, Croatia, Cyprus, Czech Republic, Greece, Hungary, Italy, Latvia, Lithuania, Luxemburg, Malta, Portugal, Romania, Slovakia, Slovenia and Spain.

Regarding the coverage element ***socio-economic performance assessments***, the countries are as well divided into three groups: those conducting sophisticated exercises to assess their innovation performance; those who follow their innovation indicators analytically, but less rigorously; and those who merely resort to statistical reporting. The very few countries belonging to the first group have set up specific advanced formats for analytical assessments of their innovation performance, often maintained by non-governmental entities. Here we find Germany, Ireland, the Netherlands, and Sweden. In Germany, the scientific Commission of Experts for Research and Innovation (EFI) publishes an annual report on the structure and trends of Germany's innovation performance, also focusing on some of the key challenges (EFI 2017).

A large majority of the EU28 countries belong to the intermediate category, as they have developed some form of general analysis of their innovation indicators, often in association with the monitoring of national innovation strategies or similar. These countries typically assess their socio-economic performance by focusing on conventional analysis of general innovation indicators. This is the case for Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Latvia, Lithuania, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, the United Kingdom (with score 1 in Table 2). As an example, in Flanders, the Centre for Research & Development

Monitoring (ECOOM) has been set up to provide the Flemish government with information on the innovation performance and reports on a biannual basis on the development of the key innovation indicators (Koenraad and Veugelers 2015).

About one third of the EU member states do not have any specific practices for analysing their socio-economic innovation performance. Even if statistical data are collected, that is not supported by broader analytical efforts. These countries are Bulgaria, Croatia, Cyprus, Greece, Hungary, Italy, Luxemburg and Malta. As an example, in Cyprus the statistical data on innovation performance are reported to international organisations, such as the European Commission, but no specific analyses are conducted in the country.

4.2 Interactions

With regard to strategic reviews, which refers to the ***interactions*** between policy and the economy, our data show that the large majority of EU member states recognize the importance of paying attention to this aspect at some level. Following the three-scale measurement above, we found the following countries in the first group: Austria, Finland, France, Germany, Ireland, Lithuania, the Netherlands, Poland, Slovenia and Sweden. All of these countries have had one or several strategic reviews combining the policy dimension with the economic perspective regarding innovation. These reviews have often been conducted by the OECD or the World Bank, but there have also been nationally-led exercises conducted by other institutions. As an example of the latter is the Austrian "System Evaluation", carried out by a consortium of research institutes. It combined the analysis of Austrian innovation policy with insights into Austrian performance in productivity growth and innovation, its external competitiveness and the innovative performance of companies (Aiginger, Falk et al. 2009). Likewise, Germany's Expert Commission for Research and Innovation (EFI) has conducted extensive analysis of transversal issues which exhibit important shortcomings, such as the limited digitalization and entrepreneurship in the German innovation system and its policies.

In the intermediary group we have countries that have had strategic reviews with a strong focus on the policy dimension, but less on the interactions with the economy. Here we find Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Germany, Hungary, Latvia, Spain, Luxemburg and the United Kingdom. Most of these countries have had a European Commission facilitated peer-review (either CREST, ERAC or PSF), where the primary focus is on policy performance and less on the transversal aspects. It is worth noting that three of the countries in this group (Croatia, Bulgaria and Luxemburg) have ordered relatively sophisticated strategic reviews that focus on interactions from international sources. However, we argue that because of the lack of quality input from policy evaluations in these countries (virtually no "coverage" in all three), the basis for the assessments is rather limited in those strategic reviews as regards the interaction between policy and the economy. It is worth noting that although the UK is a strong performer in "coverage" and in policy and economic assessments, it does not seem to be fully exploiting this

potential, as it has a limited number of strategic reviews that truly examine interaction. Possibly, the sheer size of its economy and its complex innovation system represents a challenge in analytical terms.

Finally, the last group of countries with no significant strategic review exercises is formed by Cyprus, Greece, Italy, Malta, Portugal, Romania and Slovakia. These countries have not taken part in any of the peer-review exercises facilitated by the European Commission, OECD or any other international organisation and neither have we found evidence of any nationally-defined and conducted overall strategic reviews.

4.3 Temporality

The analysis of the next attribute, namely, temporality applies the above classification of the countries into three main groups according to the data collected. First we have countries with a high degree of temporality, where various kinds of evaluations are conducted rather frequently and routinely. Secondly we have countries with a medium degree of temporality, where some types of evaluations are performed frequently, but others much less so. Finally we have countries with a low degree of temporality. Naturally, the higher the level of evaluation activity and coverage, the higher the temporality. In the first group of countries we find Austria, Finland, Germany, Ireland, the Netherlands, the UK and Sweden. In these countries the temporality of evaluative activity is high, with different elements of the innovation system being evaluated frequently and consistently. In the second group we have Belgium, Denmark, Estonia, France, Hungary, Latvia, Lithuania, Poland, Slovenia, and Spain. These countries display a medium level of temporality, meaning that their evaluation practices are very frequent in some aspects, but less frequent in others. In the third group we have countries such as Croatia, Czech Republic, Cyprus, Greece, Italy, Luxembourg, Malta, Portugal, Romania and Slovakia. In these countries the overall level of temporality is low, with evaluations being conducted rarely and infrequently.

4.4 Sources

As the fourth attribute, we look at the sources of expertise used in the system oriented evaluations of innovation policy. More specifically, we look at the extent to which EU member states are combining national and international, internal (ministerial/public) and external (e.g. private consultancies, universities, think-tanks) sources in evaluating their innovation policies. We assign values to the countries according to the diversity of these sources. Firstly there are countries that make use of diversified sources of evaluation, where the sources are both internal and external to national government bodies and where international expertise is used in innovation policy evaluation. Secondly we have countries with less diversified sources of evaluation, where only two of the sources listed above are present. Finally, we find countries where only one of the sources mentioned is relied upon.

In the first group we have Austria, Belgium, Denmark, Estonia, Finland, France, Lithuania, the Netherlands, Poland, Slovenia and Sweden. These countries make use of all three sources in their evaluations. For example, Finland has strong evaluative capacities in its innovation-agency TEKES and its public research institution VTT, making as well strong use of external consultants and academic institutions. Furthermore, it has had two international reviews, an OECD innovation review (OECD 2017) and an earlier 'custom-made' international review (Veugelers, Aiginger et al.). As two other examples, both Lithuania and Poland have demonstrated the use of a variety of sources in assessing their innovation policies. Lithuania has had an OECD innovation review (OECD 2016) and a CREST review (Edler 2007), while a government think-tank MOSTA as well as private sector evaluators have contributed significantly to its evaluative activity. In Poland, the Polish Agency for Enterprise Development (PARP) is using in-house resources as well as external evaluators to assess the innovation policy. On the international side, the World Bank carried out a strategic review of the Polish innovation system (Kapil 2013).

The second group consists of Bulgaria, Croatia, Czech Republic, Germany, Hungary, Ireland, Latvia, Luxembourg, Portugal, Romania, Spain and the United Kingdom. They use a more limited range of sources, combining either internal/external to the government (both national sources), or internal to the government (national) and international sources, or external to the government and international sources of expertise. For example, Germany, Ireland and the United Kingdom have generally sophisticated evaluative activity, but all three use almost exclusively national sources for evaluating their innovation policy. The UK had a CREST review in 2007 (Cunningham 2007), but that was of limited scale and was not followed up since then.

In the third group we find Cyprus, Greece, Italy, Malta and Slovakia. These countries have a low evaluative activity in general and they typically make use of only of a single source for their few evaluations.

5. Types of System oriented Innovation Policy Evaluations in EU28

Having examined the attributes one-by-one, we are going to make sense of these findings by placing them in a typology. Following our previous definition, a 'system oriented innovation policy evaluation' is the set of practices with high scores in all of the four attributes, that is: extensive coverage of evaluation elements, integrated interaction of horizontal and transversal policy evaluation and innovation system assessments, high regularity, and broad sources of expertise. This definition corresponds to what we have identified as a "holistic" evaluation practice in Table 2, which is a 'system oriented evaluation practice' in *sensu stricto*. It is "holistic" because the evaluation practices are rather comprehensive and organized in a way that purposefully attempts to provide sophisticated system oriented analysis for policy-makers to make decisions about innovation policies.

A strict interpretation of our concept would exclude countries which have introduced some system oriented aspects in their evaluation practices, but which have not reached the required level. If we are to provide a *sensu lato* notion, then the evalution systems of countries deemed "Flexible" and "Starters" would also constitute types of system oriented innovation policy evaluation. This would constitute a 'family resemblance' comparative methodological approach, which considers the attributes to be complementary and substitutable to some extent, and hence differences across types would be more a question of degree, once a certain level is attained (Goertz 2006) p. 45. We follow this approach, while being aware of the trade-off between extension and intension of concepts (Sartori 1970). The present paper thus focuses on a limited number of attributes with clear operationalized criteria. In so doing it considers that the "Holistic" type is a full case of clear-cut and strong system oriented innovation policy evaluation, but includes the categories "Flexible" and "Starters" designating cases of system oriented innovation policy evaluations with a less complete form than the holistic types. Our study sets as well the boundary for what is not a type of 'system oriented innovation policy evaluation' to be those cases where there is no solid coverage, interaction, temporality, nor varied sources of expertise. This conceptual boundary is determined empirically by the total scores in table 2 (see section 3.2 about measurement).

From our analysis we find that Austria, Finland, Germany, Ireland, the Netherlands and Sweden have developed holistic practices of system oriented innovation policy evaluation. All of these countries demonstrate a steady performance across the different categories of our typology. For example, Austria has a strong routine for evaluating all its innovation policy programmes, it presents an annual report to the parliament on the performance in the research and technology field, has had both a CREST peer review and a national "system evaluation" (also covering its policy-mix). As another example, in the Netherlands innovation policy programmes are routinely evaluated, with a policy-mix perspective being added at seven-year intervals. Furthermore, an annual report is prepared for the parliament on innovation performance, and both OECD as well as CREST reviews have been conducted.

In the group of "Flexible" countries we find Belgium, Denmark, Estonia, France, Lithuania, Poland, Slovenia and the United Kingdom. What characterizes the countries in this group is that all of the attributes making a system oriented innovation policy evaluation are present, but with varying degrees of sophistication. In terms of coverage, while a large majority of the countries conduct evaluations in all the three main areas (policy instruments, policy-mixes and socioeconomic assessments), we find that some countries have strong instrument evaluation practices, but there is less activity in policy-mix evaluations and socio-economic performance assessments. We can also see that the countries in this group are relatively strong in employing a variety of sources for evaluation, though with some important variation. The UK, for example, is the only country in the "Flexible" group with medium level of sources in its evaluations. This is because it has not commissioned any international evaluation with a system perspective (typically conducted by OECD and EU). When looking at the temporality of evaluations in the "Flexible" group we see that it is almost uniformly lower than in the holistic group. Again, the UK is an outlier here, as it has

high regularity. Therefore, when looking at 'temporality' and 'sources' we can see that the UK has sophisticated evaluation frameworks and demonstrates outstanding practices on several other dimensions, but is not there yet in terms of all the key features of system oriented evaluation.

'Starters' are countries that have generally little diversity of content and a low frequency of evaluative activity. The countries in this group include Czech Republic, Hungary, Latvia, Portugal and Spain. These countries all have some evaluation activity, but not a uniform coverage regarding content – some elements of "coverage" are there, but others not at all. We can see that none of the countries is conducting evaluations on their policy-mix. At the same time, a large majority of the countries in this group are making some effort to assess the interactions in their policy system, having ordered either a CREST, ERAC, PSF or a national strategic review. The latter effort is also contributing to some variety of sources used in evaluations, adding an international dimension to a field mainly dominated by domestic actors. Similarly to the flexible group, the overall frequency of evaluative activity in these countries is relatively low.

Last, we have countries which do not have any true system oriented innovation policy evaluation. The countries in this group are Bulgaria, Croatia, Cyprus, Greece, Italy, Malta, Luxemburg, Romania and Slovakia. None of these countries has any considerable evaluation activity. While some evaluations have taken place over time, they have been isolated examples, not exhibiting any temporality. For example, Cyprus has had an ERAC peer review of its innovation system, but almost no other evaluations. On the other hand, Italy has carried out some evaluations on its policy instruments, but there is very scarce activity on the other levels. While several of these countries have made plans for developing their evaluation capacities in order to provide a better understanding of the innovation system,[3] these initiatives are yet to take effect.

6. Conclusions

This paper has provided new insights about an under-researched phenomenon in innovation and evaluation studies, namely, the actual practice in 'system oriented innovation policy evaluations'. In order to do so, it has conceptualized this term, identifying its four constitutive attributes, which have then been operationalized and measured. On this basis, the paper reports on thorough empirical research that has examined all EU28 countries' practices, charactering them within a four-fold typology.

The findings show that only six out of the EU28 countries have developed the most complete type of system oriented evaluation practices (The Netherlands, Austria, Finland, Germany, Ireland and Sweden). These countries have developed what we term "holistic", meaning that they fulfil with great intensity the four attributes that define system oriented innovation policy evaluation. That

[3] For example Malta has ordered a PSF study on the monitoring of the Maltese national research and innovation strategy (Interview 43).

is, a wide coverage of evaluations, analyses of vertical and of transversal interactions between policy performance and socio-economic performance, a high level of regularity of those evaluations, and broad and varied sources of expertise. Not surprisingly, there are only a few countries in Europe in the group that fulfils these high-level criteria, and therefore can be considered to have a holistic approach. A second group of countries are characterised as "flexible", because their evaluation practices have most of the previous attributes as those in the holistic group, but with a weaker performance. Eight out of 28 countries are found in this group: Denmark, France, Belgium, Poland, the UK, Estonia, Lithuania and Slovenia. While the countries in the flexible group are still relatively strong in instrument evaluations, the policy-mix evaluations and socio-economic performance assessments are less prominent. Also, the overall frequency of evaluations is visibly smaller. For this reason, they are not holistic because their evaluation efforts do not incorporate the complexity of the interactions between policy performance and socio-economic performance.

The third group of countries are those we name "Starters": Latvia, Spain, Hungary, Czech Republic, and Portugal. These are countries with an uneven regularity of evaluation activities and uneven variation of the sources of those evaluations. Their coverage is rather limited, and so is the analysis of interactions in the system. But these countries have made clear attempts to engage with available expertise and tap into the available knowledge, typically from international sources, and to comply with conditions slightly above the minimum required by external funders. These are countries which have taken the first steps towards creating some basic structures of what could in the future become a more solid system oriented approach.

Last, we find a relatively large group of countries in the European Union (9 out of 28) without any kind of system oriented innovation policy evaluation: Bulgaria, Croatia, Luxemburg, Romania, Italy, Slovakia, Cyprus, Greece and Malta. Our conceptual boundary is very clearly defined here, as these countries have none or extremely few of the attributes of coverage, interaction, temporality, and sources. From our data we could not find any reasonable evidence of evaluation activities being conducted in a systematic manner. However, it is worth mentioning that some countries in this group are planning to do so in the future.

The widespread advent of a system oriented approach to evaluation would undoubtedly be an opportunity for more comprehensive, contextualized, evidence-based innovation policy-making in the EU28 countries. System oriented evaluation provides unique overviews and strategic assessment of the effectiveness of policies in the context of the innovation system. They are a very valuable foundation for future decisions about what policymakers can do in the face of challenges and bottlenecks. However, there are still serious obstacles in this regard because such practices require substantial knowledge and organisational capacities. It is not easy to create a system oriented innovation policy evaluation. In fact, it is quite demanding. For that reason, overcoming these obstacles would need more decisive capacity-building at national level, aiming at creating competences within governmental structures that will allow the initiation of such endeavors.

Our findings pose as well a series of highly relevant research questions for future analysis. The most obvious one is how and how far system oriented innovation policy evaluations being are being used. Are they transformative in the sense of providing a basis for relevant changes in policymaking? Is there any policy learning and policy change taking place? Who are the learners in that process? Are they learning differently across the process? Some evidence in this regard has started to accumulate with individual case studies. A very recent publication about the Basque country provides evidence of such dynamics at a regional level (Aranguren, Magro et al. 2017). The findings in that case study show that "explicit demand for evaluation, decisions around the appropriate mode of knowledge generation, the existence of dialogue spaces where relevant policy and stakeholders (including researchers) frequently meet, and the development of trust and cognitive proximity within these meeting places, are all critical factors if evaluation is to be transformative".(p. 703). Likewise, a recent study at EU level investigated what different learners actually learn from evaluations (Borrás and Højlund 2015). Their findings indicate that two groups involved in the evaluation are actually learning something from it, employees in programme units and external evaluators, and that they learn rather different things. Some of them get a programme overview, whereas others learn from small-scale programme adjustments. These new findings at regional and European Union level are very promising in terms of understanding the context in which innovation policy evaluation is actually being used and how.

The more evaluation practices move towards the system oriented approach, the more complex the interactions among actors will be. New opportunities will emerge as well in regards to understanding the role of system oriented innovation policy evaluation practices in the context of other intelligence gathering areas, such as technology foresight and technology assessment. Bringing these three activities together is considered an important way of creating distributed intelligence in innovation policy making. Therefore the empirical question that still remains unanswered is: how far are EU28 countries building such a distributed intelligence.

References

Aiginger, K., R. Falk and A. Reinstaller (2009). Evaluation of Government Funding in RTDI from a Systems Perspective in Austria: Synthesis Report [reaching Out to the Future Needs Radical Change; Towards a New Policy for Innovation, Science and Technology in Austria: the Summary Report is Based on Nine Special Reports]. WIFO.

Aranguren, M. J., E. Magro and J. R. Wilson (2017). "Regional competitiveness policy evaluation as a transformative process: From theory to practice." Environment and Planning C: Politics and Space **35**(4): 703-720.

Arnold, E. (2004). "Evaluating research and innovation policy: a systems world needs systems evaluations." Research Evaluation **13**(1): 3-17.

Borrás, S. (2011). "Policy Learning and Organizational Capacities in Innovation Policies." Science and Public Policy **38**(9): 725-734.

Borrás, S. and S. Højlund (2015). "Evaluation and policy learning: The learners' perspective." European Journal of Political Research **54**(1): 99-120.

Cunningham, P., J. Edler, K. Flanagan and P. Larédo (2016). The Innovation Policy Mix. Handbook of Innovation Policy Impact. J. Edler, P. Cunningham, A. Gök and P. Shapira. Cheltenham, Edward Elgar.

Cunningham, P. e. a. (2007). Policy Mix Peer Reviews: Country Report. United Kingdom, a Report of the CREST Policy Mix Expert Group. Brussels.

Dahler-Larsen, P. (2012). The Evaluation Society. Stanford, Ca., Stanford University Press.

Daly, M. and M. L. Christensen (2016). The Effect of Multiple Participations in the Danish Innovation and Research Support System. Copenhagen, Centre for Economic Business Research (CEBR).

DASTI (2014). The Short-run Impact on Total Factor Productivity Growth of the Danish Innovation and Research Support System. Research and Innovation: Analysis and Evaluation No. 2. Copenhagen.

Department_of_Jobs, E. a. I. (2015). Evaluation of Enterprise Supports for Enterprise: Synthesis Report and Conclusions. Dublin.

Edler, J. (2007). Policy Mix Peer Reviews: Country Report. Lithuania, a Report of the CREST Policy Mix Expert Group. Brussels.

Edler, J., M. Berger, M. Dinges and A. Gök (2012). "The practice of evaluation in innovation policy in Europe." Research Evaluation **21**(3): 167-182.

Edler, J., B. Ebserberger and V. Lo (2008). "Improving Policy Understanding by means of Secondary Evaluation." R&D Evaluation **17**(3): 175-186.

Edquist, C. (2005). Systems of Innovation. Perspectives and Challenges. The Oxford Handbook of Innovation. J. Fagerberg, D. C. Mowery and R. R. Nelson. Oxford, Oxford University Press.

Edquist, C. (2011). "Design of innovation policy through diagnostic analysis: identification of systemic problems (or failures)." Industrial and Corporate Change **20**(6): 1725-1753.

EFI (2017). Report on research, innovation and technological performance in Germany 2017. Berlin, EFI.

Feller, I. (2007). "Mapping the frontiers of evaluation of public-sector R&D programs." Science and Public Policy **34**(10): 681-690.

Flanagan, K., E. Uyarra and M. Laranja (2011). "Reconceptualising the 'policy mix' for innovation." Research Policy **40**(5): 702-713.

Goertz, G. (2006). Social Science Concepts. A User's Guide. Princeton, Princeton University Press.

Hage, J., G. Jordan and J. Mote (2007). "A theory-based innovation systems framework for evaluating diverse portfolios of research, part two: Macro indicators and policy interventions." Science and Public Policy **34**(10): 731-741.

Jordan, G. B., J. Hage and J. Mote (2008). "A theories-based systemic framework for evaluating diverse portfolios of scientific work, part 1: Micro and meso indicators." New Directions for Evaluation **2008**(118): 7-24.

Kapil, N. (2013). Poland - Enterprise Innovation Support Review: From catching up to moving ahead. Washington DC., World Bank.

Koenraad, D. and R. Veugelers (2015). Vlaams Indicatorenboek. V. Overheid. Brussels.

Kuhlmann, S., P. Boekholt, L. Georghiou, K. Guy, J.-A. Heraud, P. Laredo, T. Lemola, D. Loveridge, T. Luukkonen, A. Moniz, W. Polt, A. Rip, L. Sanz-Menendez and R. E. Smits (1999). Improving Distributed Intelligence in Complex Innovation Systems. Munich Personal RePEc Archive.

Kuhlmann, S., P. Shapira and R. E. Smits (2010). Introduction. A Systemic Perspective: The Innovation Policy Dance. The Theory and Practice of Innovation Policy. An International Research Handbook. R. E. Smits, S. Kuhlmann and P. Shapira. Cheltenham, Edward Elgar**:** 1-22.

Lundvall, B.-Å., Ed. (1992). National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. London, Pinter.

Lundvall, B.-Å. and S. Borrás (1998). The Globalising Learning Economy: Implications for Innovation Policy. Brussels, European Commission.

Magro, E. and J. R. Wilson (2013). "Complex innovation policy systems: Towards an evaluation mix." Research Policy **42**(9): 1647-1656.

Martin, B. R., P. Nightingale and A. Yegros-Yegros (2012). "Science and technology studies: Exploring the knowledge base." Research Policy **41**(7): 1182-1204.

Molas-Gallart, J. and A. Davies (2006). "Toward Theory-Led Evaluation: The Experience of European Science, Technology, and Innovation Policies." American Journal of Evaluation **27**(1): 64-82.

Nelson, R. R., Ed. (1993). National Innovation Systems. A Comparative Analysis. Oxford, Oxford University Press.

OECD (2016). OECD Reviews of Innovation Policy: Lithuania 2016. Paris, OECD.

OECD (2017). OECD Reviews of Innovation Policy: Finland 2017. Paris, OECD.

Office_of_the_Government_of_the_Czech_Republic (2013). Methodology of Evaluation of Research Organizations and Evaluation of Finished Programmes (valid for years 2013 - 2015). Prague.

Sartori, G. (1970). "Concept Misformation in Comparative Politics." The American Political Science Review **64**(4): 1033-1053.

Smits, R. and S. Kuhlmann (2004). "The rise of systemic instruments in innovation policy." International Journal of Foresight and Innovation Policy **1**(1-2): 4-32.

The_Netherlands_Goverment (2014). Regulation on Periodic Evaluation, available at http://wetten.overheid.nl/BWBR0035579/2015-01-01 (Accessed: 09.09.2017).

Veugelers, R., K. Aiginger, C. Edquist, D. Breznitz, G. Murray, G. Ottaviano and M. Maliranta Evaluation of the Finnish National Innovation System - Policy Report. Helsinki, Ministry of Education and Ministry of Employment and the Economy.