



The Politicisation of Evaluation: Constructing and Contesting EU Policy Performance

Paul J. Stephenson · Jonas J. Schoenefeld · Frans L. Leeuw

Published online: 16 December 2019
© The Author(s) 2019

Abstract Although systematic policy evaluation has been conducted for decades and has been growing strongly within the European Union (EU) institutions and in the member states, it remains largely underexplored in political science literatures. Extant work in political science and public policy typically focuses on elements such as agenda setting, policy shaping, decision making, or implementation rather than evaluation. Although individual pieces of research on evaluation in the EU have started to emerge, most often regarding policy “effectiveness” (one criterion among many in evaluation), a more structured approach is currently missing. This special issue aims to address this gap in political science by focusing on four key focal points: evaluation institutions (including rules and cultures), evaluation actors and interests (including competencies, power, roles and tasks), evaluation design (including research methods and theories, and their impact on policy design and legislation), and finally, evaluation purpose and use (including the relationships between discourse and scientific evidence, political attitudes and strategic use). The special issue considers how each of these elements contributes to an evolving governance system in the EU, where evaluation is playing an increasingly important role in decision making.

Keywords EU Institutions · Performance · Contestation · Learning · Evaluation use

P. J. Stephenson (✉) · F. L. Leeuw
Maastricht, The Netherlands
E-Mail: p.stephenson@maastrichtuniversity.nl

J. J. Schoenefeld
Darmstadt, Germany

Norwich, UK

Die Politisierung von Evaluation: Debattenbeiträge zur Konstruktion von EU-Policy-Performanz

Zusammenfassung Obwohl systematische Policy-Evaluation bereits seit vielen Jahrzehnten betrieben wird und insbesondere innerhalb der Institutionen der Europäischen Union (EU) und der EU-Mitgliedstaaten stark gewachsen ist, wird sie weiterhin größtenteils von der politikwissenschaftlichen Literatur vernachlässigt. Bestehende Arbeiten in der Politikwissenschaft und der Policy-Analyse konzentrieren sich meist auf andere Themen, wie Agenda-Setting, Policy-Entstehung, politische Entscheidungsprozesse oder die Implementierung öffentlicher Politiken. Während an einigen Stellen vereinzelte Forschungsarbeiten zur Evaluation in der EU erscheinen, die sich meist auf Policy-Effektivität konzentrieren (ein Kriterium unter vielen), fehlt bisher ein strukturierter Ansatz. Das aktuelle PVS-Schwerpunktheft stößt in diese Forschungslücke und konzentriert sich auf vier Schlüsselaspekte: Evaluationsinstitutionen (inklusive ihrer Regeln und Kulturen), Evaluationsakteure (Kompetenzen, Macht, Aufgaben, Interessen und Interaktionen), Evaluationsmethoden und -theorien (inklusive Folgenabschätzung und deren Einfluss auf das Policy-Design und Gesetzgebung) und zum Schluss Evaluationszwecke und -nutzen (inklusive der Beziehung zwischen Diskurs und wissenschaftlicher Evidenz, politische Einstellungen und strategischem Nutzen). Das Schwerpunktheft betrachtet, inwiefern diese Elemente zur Entwicklung des EU-Governance-Systems beitragen, in dem Evaluation eine immer wichtigere Rolle in Entscheidungsprozessen spielt.

Schlüsselwörter EU-Institutionen · Performanz · Evaluationsdebatten · Lernen · Evaluationsverwendung

1 Introduction

Despite new forms of governance becoming increasingly important for the production of collective goods, their effectiveness and efficiency seem to be limited (Grande 2012). As a result, “modern societies must be afraid of having been caught in a ‘governance trap’” (Grande 2012, p. 565). This debate is of particular relevance to the European Union (EU), where, beyond pure efficiency concerns, issues of quality, independence, objectivity and scope are crucial for the throughput and output legitimacy of its governance and policy making (Schmidt 2013). The quality of policy outputs therefore depends in good measure on the quality of governance. The end of “permissive consensus” (Hooghe and Marks 2009) on European integration and the rise of Euroscepticism make demonstrating the effectiveness and efficiency of EU policies politically more challenging. According to the Oxford English Dictionary, we understand a (EU-level) public policy to be “a principle or course of action adopted or proposed as desirable, advantageous, or expedient; esp[ecially] one formally advocated by a government, political party, etc.” Note, however, that there is a wide-ranging discussion on the meaning of public policy in the relevant policy literatures (Hill and Varone 2017, pp. 15–23).

How do we go about ascertaining how well a policy performed, or indeed, if it was a well-designed policy? Many have argued that evaluation—here understood as “the process of determining the merit or worth or value of something; or the product of that process” (Scriven 1981)—has a key role to play in the objective and systematic assessment of public policy. Yet while policy evaluation (hereafter “evaluation[s]” for brevity) has often been heralded as a key ingredient of successful governance, it nonetheless remains heavily contested by political actors who are prone to conduct and use evaluations strategically. Contrary to those who have sought to depict evaluation as a mere technical exercise, politics is inherent in any governance practice aimed at measuring and deliberating the performance legitimacy of public policy, and especially that of the EU (Bovens et al. 2006; Scharpf 2009; Vedung 1997). Political differences may emerge, for example, over the underlying “values” that evaluation uses in its assessment (see Majone 1989). There are many values and principles that can be used in an evaluation, ranging from economic ideas on efficiency to effectiveness against stated political objectives, as well as more processual approaches such as criteria of fairness, equity, or democratic legitimacy. Depending on the values applied, the results of an evaluation may differ starkly, generating interest-based preferences for some evaluation approaches and outcomes over others.

Such differences emerge especially because any number of different values might crucially also conflict with one another. The political challenge therefore involves forging political and societal agreement on the values that should, in turn, underlie evaluation (see Fischer 2006). Reflecting such dynamics, evaluation in the EU grew spectacularly in the 1990s and 2000s, driven by the quest for accountability and learning but also by a demonstrable desire of political actors to “manipulate political opportunity structures” (Schoenefeld and Jordan 2019; see also Stame 2003 and Stern 2009). More recently, the perceived and argued-for importance of policy evaluation has been exacerbated by a shift towards austerity in many European countries in the aftermath of the financial crisis that began in 2008 (Streeck and Schäfer 2013). As a result, demonstrating “value for money” has become increasingly important. Political pressure to demonstrate a “Europe of Results” goes hand in hand with political divergence over how to demonstrate results—and therein we have politics in the governance of evaluation, impacting not only upon the tendering and managing of policy evaluation (Schoenefeld and Jordan 2017) but also on determining both the decision making for conducting evaluation (or not) and the use (or non-use) of its findings.

Certain political motivations within public management have driven policy appraisal (ex ante evaluation), making it an “inherently political phenomenon” with actors shaping appraisal structures and practices to suit their interest-based preferences (Adelle et al. 2012). Conducting good policy assessments, which are then used to inform policy, depends largely on having sufficient institutional capacities and resources, as well as on leadership and political will. In other words, policy evaluation is intrinsically linked to deliberative processes within institutional settings—of designing policy and legislating (ex ante) and, thereafter, of implementing and monitoring, auditing, and scrutinising policy outputs (ex post) as part of the accountability regime, both normatively as a virtue and procedurally as a mechanism

(Bovens 2010; see also Stame 2006). Although some aspects of evaluation (such as impact assessment) have received greater attention in the political science literature than others, a more comprehensive treatment of evaluation in all its forms remains scarce. For example, research outputs in key public policy journals have tended to focus on the earlier stages of the policy cycle: agenda setting, policy formulation, policy decision making, and implementation (see Sabatier 2007). This has generated veritable gaps in scholarly debate and in knowledge on aspects of evaluation as one crucial aspect of policy making and the associated politics.

Why is there such a dearth of research into the political dynamics and tensions within the theory and practice of policy evaluation, especially in the field of political science? Some EU policy scholars may in fact be conducting aspects of evaluation research without realising it or without framing their work explicitly as such; for example, when writing on the impact of the Structural Funds in cohesion policy (Fratesi and Wishlade 2017; Hoerner and Stephenson 2012) or on the effectiveness of EU missions in external action (Batterbury 2006; Peen Rodt 2014). Beyond the immediate communities of European integration studies and EU public policy communities, certain public administration scholars have recently turned their attention to the efficiency of public sector organisations while a community of evaluation scholars has largely focused on methodological questions of evaluation practice.

This special issue addresses this fragmentation by combining articles from leading scholars of public administration, political science, public policy, and evaluation (for earlier related efforts, see Hoerner and Stephenson 2012; Vaessen and Leeuw 2010; Dahler-Larsen 2011) to explore four key themes concerning the state of policy evaluation in the EU:

1. *Evaluation institutions*—including the rules for evaluation, the contestation of new evaluation institutions and their fight for legitimacy, organisation and practices of active scrutiny at the policy level, and evaluation cultures.
2. *Evaluation actors and interests*—including the competencies, power, and transformation of public and private evaluation communities, as well as the effects of increased competition in evaluation on intra-institutional and inter-institutional politics, roles, and tasks.
3. *Evaluation design*—including the approach to, and purpose of, evaluation, as determined by research methods, theories and data collection, and their impact on policy design and legislation.
4. *Evaluation purpose and use*—including the relationships between discourse and scientific evidence, political attitudes and evaluation practice, as well as the strategic use of policy evaluation results, findings and recommendations.

2 Bridging Political Science and Evaluation Studies: Institutions, Actors, Design, and Use

So what is the existing divide across disciplines? What exactly is it that needs to be bridged? Both political science and evaluation studies stand to gain from engaging with one another in multiple areas of theory and practice. First, given the growing

interest in “evaluation systems” (Leeuw and Furubo 2008; Olejniczak 2013), there are potential links with debates on governance systems and institutions in political science; that is, structures of (overarching) rules and norms that set a context and operating space for evaluation. Second, evaluation scholars argue that evaluation systems require stable sets of evaluation actors (Leeuw and Furubo 2008). Debates on political actors and their constellations, interests, resources, and impacts are, of course, not new to political science, but have thus far only recently been applied to evaluation (see Schoenefeld and Jordan 2017). Third, both political science and evaluation use theoretical approaches as well as a range of methods. Building a strong analytical framework and applying rigorous methodological approaches is something that concerns both political scientists and evaluators alike, making for potentially productive cross-fertilisation. Fourth, evaluation scholars have specifically discussed evaluation purposes and uses, themes that link with debates on evidence-based policy making in political science (Patton 2008; Pawson 2005; Sanderson 2002; Strassheim and Kettunen 2014). Taken together, all four focal points hold considerable promise for expanding the political study of evaluation and, in doing so, generating productive linkages with evaluation studies. This section discusses extant scholarly debates related to each focal point with a view to the EU.

2.1 Evaluation Institutions (Political Science Perspective)

Institutional innovation in the EU’s evaluation architecture has brought an expansion of evaluation activity and capacity. For example, the European Commission and its Regulatory Scrutiny Board have endeavoured to harmonise evaluation cultures (across Directorates-General (DGs)), provide internal methodological support, aid data collection and comparison, stimulate learning, and encourage meta-evaluation (Radaelli 2018; Stern 2009). The directorates of the European Parliament and the new European Parliamentary Research Service (EPRS) are also engaging in new and increasing evaluation activity (see Schrefler 2016; Stephenson 2017). In addition, a wide range of nongovernmental organisations have also become involved either as evaluation contractors or as independent evaluators, as has been shown, for example, in the area of climate policy (Huitema et al. 2011; Schoenefeld 2018). The sum of such wide-ranging activities across a number of institutions may give rise to a broader evaluation system in the EU, assisted by rules such as a growing number of evaluation clauses in legislation (see Bussmann 2005). Scholars have also examined progress in the way that evaluation is institutionalised in various European countries from the perspectives of political, social and professional systems (Furubo et al. 2002; Jacob et al. 2015; Stern 2009; Stockmann et al. 2020).

Leeuw and Furubo (2008) consider social systems in and around evaluation as “evaluation systems” when they are characterised by four criteria: First, the activities carried out have to be characterised in terms of a certain cultural–cognitive perspective (Scott 2001). There should be some agreement among the players involved about what they are actually doing and why they are doing it, or to put it differently, there must be a shared epistemology. The second criterion is that the evaluation activities must be carried out by organisations and institutions and not only (or largely) by “lonely” or sole-trading evaluators. Third, to be labelled as an evaluation system,

there should be a degree of permanence or history in the activities involved: they are part of something ongoing. This also means there will be a tendency to replace ad hoc initiatives and ad hoc organisations with activities and organisations planned in advance that have a more permanent character. The fourth criterion is that the information from evaluative activities has to be (institutionally) linked to decision and implementation processes. This can be the planning process of a government department, universities, or the World Bank, but it can also be the governmental budget process or adjustments of the curricula of schools and universities.

Several evaluation authors, including Raimondo (2018), conceive of “evaluation systems” as different from what political scientists would perhaps normally speak of as “polities” or “systems of governance.” They distinguish between the system of performance monitoring; the systems of performance audit, inspection, and oversight; the system of (quasi)experimental evaluations and the evidence-based policy movement; the accreditation and evaluation system; and the monitoring and evaluation system. In practice, there are certain risks inherent in evaluation systems with regards to how they are formalised and institutionalised, and how, in turn, they become predictable when it comes to procedures, participants and ways of doing. Evaluation systems can be “captured” in the same way, perhaps, that means of interest representation and channels of access to policy formulators can also be captured. When the subjects of evaluation gain too much control over evaluation questions or even evaluation methodology, it might be said that evaluation systems have become too integrated with the administrative–political world. Hellstern and Wollmann (1986) were among the first Europeans to draw attention to this problem. Such dynamics link closely with the legitimisation of evaluation. One assumption is that the more that evaluation activities are part of the administrative–social–political “system,” the more their legitimisation increases. However, with increased legitimisation, evaluation’s role in “speaking truth to power” may be diminished (see Hellstern and Wollmann 1986).

The latter arguments link with a concept introduced in the late 1990s by Michael Power: rituals of verification (Power 1997). Power referred to this in regard to his work on (performance) auditing, where the auditors produced rituals that the auditees to some extent valued/loved because they created procedures and ways of doing things that made life relatively easy (see also Stephenson 2015 with regards to the European Court of Auditors). A similar concept can be applied to evaluation, in particular when evaluations merely become administrative “machines” (Dahler-Larsen 2011) that produce (on and on) “similar types of knowledge”/similar languages or speech-acts. More fundamentally, we should be asking not only what the purpose of a particular evaluation is, but what purpose the evaluation serves in the polity/politics/policy process (that is, at a systemic level). This question should not merely be in terms of accountability or policy learning but should perhaps be connected more explicitly with notions of problems and solutions (Kingdon 2011). Questions include (i) Did the policy solve a perceived problem? and (ii) Can an evaluation be used in such a way to determine this? In sum, the institutional and more systemic perspectives raise a number of pertinent questions related to the theme of governance by evaluation.

2.2 Evaluation Actors and Interests (Political Science Perspective)

A growing debate related to evaluation concerns the actors involved in producing and conducting evaluations (as well as the users; see below). In principle, as Schoenefeld and Jordan (2017) have highlighted, both state and nonstate actors have become involved in policy evaluation. Actor attributes have consequences for evaluation—for example, nonstate actors may be more independent, but they may struggle to bring their evaluation results into the policy process (Weiss 1993). By the same token, governmental actors may have greater access both in terms of data for their evaluation and in terms of the usage of their findings, but they may be constrained in other, political ways.

The fact that many evaluations are commissioned, that is, paid for by one actor and conducted by another, has generated a vibrant debate on principal–agent relationships in evaluation, in part as a function of the differences explained above. For instance, Pleger and Sager (2018) have argued that such relationships may both improve and detract from evaluation. This is especially relevant because we have been observing a changing role and influence of actors as a function of evaluation. To what extent do political interests affect or underpin their work? It is key to address the interplay between evaluation providers, bureaucracies, and political actors and their needs/demands given the ongoing politicisation of evaluation practice (The LSE GV314 Group 2013). Key questions include (i) Is there variation in the way that evaluations are announced, framed, tendered, delivered, and used/not used? and (ii) Do organisations learn from evaluations? We can answer these questions by drawing on insights from the evaluation community (e.g. Leeuw 1994) and political science (Bennett and Howlett 1992; Radaelli and Dunlop 2013).

Extant literatures have often focused on the European Commission (EC) as a key (institutional) evaluation actor (see Camis o and Vila Maior 2019). But Mastenbroek et al. (2016, p. 1330) claim that systematic evaluation by the Commission “is not likely to materialize” because evaluations “may uncover critical problems in the actual working of legislation.” An extensive administrative reform in order to increase evaluation activities may lead to a harmful internal overload, as it did, for instance, in the case of the Prodi Commission (see Levy 2006). As for the external dimension of the EC’s actorness, first, the Better Regulation Agenda established the “evaluate first” principle, meaning that “for any existing intervention, an evaluation should be the starting point of any discussions on performance and possible (significant) change” (European Commission 2017, p. 327). In other words, before modifying a piece of legislation, the Commission has to assess the existing evidence (e.g. already conducted evaluations that relate to the issue). Ideally, this will foster learning from past experience and link ex post evaluations with ex ante evaluations (impact assessments), since the former have to be considered in the preparation of the latter.

But in practice, neither the Commission nor other evaluation actors operate in isolation. Evaluation has increasingly been understood as a “community of practice,” which Wenger (2011) defines as “groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly” (p. 1). Indeed, these communities now have their own journals (e.g. *Evaluation* and the

American Journal of Evaluation), conduct their own conferences, and have created their own outputs, such as evaluation standards (see Widmer 2004). A key question thus becomes how this community interacts with the policy community. These arguments suggest that a significant focus on evaluation actors, their relationships, and their contexts is warranted.

2.3 Evaluation Design (Evaluation Perspective)

A third important focal point concerns evaluation design, including theory and methodology. In political science, theory is at the heart of evaluation and, indeed, is problematic in its own way, at least insofar as understanding of theory and its usage are fragmented. We can distinguish between theories of policy makers, stakeholders, and evaluators that underly their professional work in making policies and doing evaluations, and scientific theories capable of contextualising and explaining the consequences of policies, programmes, and evaluators' actions (Leeuw and Donaldson 2015). Much theoretical discussion has taken place over the years in the *Evaluation* journal, including suggestions to further stimulate the development of theoretical work in the evaluation profession: "theory knitting," "theory layering," and "theory-driven evaluation science" (Leeuw and Donaldson 2015, p. 467). Furthermore, the authors assert that there is no such thing as "a" or "the" evaluation theory currently applied in evaluation and by the evaluation profession.

Perhaps the essential difference between political science and evaluation is that evaluation is more instrumental and forward-looking, insofar as political scientists tend to do research and policy analysis of the recent past in order to convincingly reconstruct that past and thus better understand the actors/institutions and policy area, although they are often reluctant to predict the future (even if they commonly demand of theory that it have explanatory power and predictability) or to give recommendations. Evaluation scholars, by contrast, are more practically oriented, keen for their work to have real-world practical uptake. They often focus on theorising individual policies (e.g. programme theory) rather than overarching, societal mechanisms (see Rogers et al. 2000).

In spite of these differences, many normative but also practical concerns within political science research—validity, reliability, replicability, generalisability—can also be found in evaluation studies. There are disciplinary similarities and key concerns with regard to research design, data collection, and analytical frameworks for analysis. The quality and value of an evaluation, and indeed its authority and persuasiveness and the degree to which it can or cannot be contested, may ultimately be determined by the strength of its research design.

2.4 Evaluation Purpose and Use (Evaluation Perspective)

A fourth and final focal point of this special issue is the strategic use and political contestation of evaluation, which requires empirical research in order to examine the way in which methods, theory, data, and results are accepted or refuted. How does discourse/dialogue compete with scientific evidence in the deliberative practice of evaluation? Why is there such political interest in quantification, and how is this

linked to perceived legitimacy? Political attitudes and perceptions of evaluation's relevance affect its use, including in technical impact assessment and day-to-day politics. How do evaluations vary in scope, from impact assessments and ex post evaluations to policy reviews? What is the EU's degree of resilience to political interference, and how is it coping with misinformation, fake news, and/or poor data when it comes to assessing policy performance?

Engaging with such questions, evaluation scholars have made significant inroads into questions of evaluation use (Christie 2007; Henry and Mark 2003; Højlund 2014a, 2014b). Empirical investigations have often identified several challenges. As Højlund (2014b, p. 26) asserts, “administrators, politicians and citizens have an interest in knowing to what extent evaluations are used to improve policies,” but nonetheless, rarely do they change policies. He finds that “justificatory uses do not fit with evaluation’s objective of policy improvement and social betterment” (Højlund 2014b, p. 26). Evaluations may be commissioned to legitimise a course of action or as symbolic politics, or else there may be purposeful “non-use.” We can also distinguish between “findings use” (from evaluation findings) and “process use” (use during the evaluation process). Analysing the evaluating organisation and the factors “conditioning” the evaluation undertaken can often reveal a great deal about whether or not an evaluation will be used, and if so, how (see Alkin and Taut 2003).

King and Alkin (2018) have theorised evaluation use, recently looking back over 50 years of research on “use theory.” Drawing on work by Miller (2010) and Shadish et al. (1991), they put forward their own framework that captures those factors that scholars have focused their attention on when exploring *if* and *how* evaluations are in fact used: operational specificity (explicit details are given about how to foster evaluation use for studies in specific settings); range of application (explicit description is provided about where the theory is likely to increase use and where it is not likely to succeed); feasibility in practice (practitioners can easily and routinely conduct the activities); discernible impact (the prescribed activities do, in fact, lead to increased use); and reproducibility (different practitioners can reproduce the same outcomes at different times and places) (King and Alkin 2018, p. 436).

Some scholars have argued that we should “extend the narrow framing of use by adding a broader-based construct” (Alkin and King 2017, p. 443). Kirkhart (2000) suggests that we need a more inclusive understanding of the impact of evaluation—one that considers intention (intended or unintended), source (evaluation process or results), and time (immediate, end-of-cycle, long-term). She proposed the term “influence” as an addition to “use,” allowing for a better framework to capture effects that are “multidirectional, incremental, unintentional, and noninstrumental, alongside those that are unidirectional, episodic, intended, and instrumental” (Kirkhart 2000, p. 7). Given that the role of knowledge in politics has also captured political scientists’ attention, interaction between the relevant literatures holds significant promise.

3 Understanding Policy Evaluation in EU Governance: Contributions of the Special Issue

As a collective, the articles of this special issue contribute to all four focal points identified in Sect. 2. In Table 1 we plot the contributions of the special issue, grouping them by aspects that have traditionally been central to political scientists (focal points 1 and 2) and aspects on which evaluation scholars and practitioners have often focused (focal point 3 and 4). Multiple articles in this special issue address more than one focal point, demonstrating the overlaps and compatibilities that this special issue seeks to highlight. The remainder of this section introduces the contributions to each focal point in more detail.

3.1 Evaluation Institutions: their Emerging Cultures and Systems

At an institutional level, Jankauskas and Eckhard (2019) demonstrate how the Juncker Commission's Better Regulation Reform (BRR) has redefined the "tools and rules" of evaluation within the Commission. The reform has introduced and strengthened norms such as the "evaluate first" principle, and it uses public consultations as well as the newly created Regulatory Scrutiny Board to further deepen and legitimise evaluation practice within the Commission. In a similar vein, Pattyn et al. (2019) demonstrate how the introduction of new public management in

Table 1 Towards a categorization of policy evaluation in EU governance based on the existing literature and including the contributions of this special issue

Political Science concerns/perspective		Evaluation concerns/perspective	
Focal Point 1— Evaluation institu- tions	Focal Point 2— Evaluation actors and interests	Focal Point 3— Evaluation design (methods and theory)	Focal Point 4— Evaluation purpose and use
<i>Area of potential contestation/evaluation activity</i>			
<ul style="list-style-type: none"> – institutional innova- tion – organisation of and mandate for evaluation – scrutiny of evalu- ation conduct and quality – adoption of techni- cal and professional norms in evaluation practice 	<ul style="list-style-type: none"> – writing rules and evaluation tenders – role and influence of administration – management of evaluation findings – input of evaluation into policy design 	<ul style="list-style-type: none"> – evaluation frame- works and method- ologies – data collection and interpretation – beliefs/logics of intervention/cause and effect 	<ul style="list-style-type: none"> – strategic use (or not) of findings by executive and legislature at EU, nat/reg/local level – symbolic politics – rejection and dele- gitimization
<i>Relevant author contribution</i>			
Jankauskas & Eckhard Pattyn, Brans & De Peutter Schoenefeld, Schulze, Hilden & Jordan	Jankauskas & Eckhard Hoerner Pattyn, Brans & De Peutter	Dahler-Larsen & Sundby Potluka Schoenefeld, Schulze, Hilden & Jordan De Francesco	Pattyn, Brans & De Peutter Hoerner Dahler Larsen-Sundby

Source: own compilation, based on existing literature

Belgium has increased the number of times that ministers have formally called for evaluation. Their work demonstrates how new rules—in part driven by the wider evidence-based movements across Europe—changed governmental behaviour.

Institutional innovation has also appeared in other areas of EU policy making, as the example of the EU's Monitoring Mechanism for Greenhouse Gases and Policies and Measures demonstrates (Schoenefeld et al. 2019). The authors unpack how policy monitoring—potentially a key ingredient to evaluation—is itself subject to institutional dynamics and change, implementation issues, and debates on quality. Monitoring should not simply be assumed as a feature of implementation (as many have done), but rather it should be subject to careful study and (eventually) policy design. The empirical example of climate policy monitoring in the EU reveals that institutional path dependencies loom large, but there are also learning effects in monitoring that can be identified in both quantitative and qualitative analyses (Schoenefeld et al. 2019).

3.2 Evaluation Actors and Interests: their Evolving Role and Influence

Construing the European Commission as an evaluation actor, Jankauskas and Eckhard (2019) demonstrate how the Better Regulation Agenda, in addition to deepening institutionalisation of evaluation in the Commission (see above), has also strengthened its “strategic actorness.” More specifically, multiple tools mentioned above have streamlined evaluation within the Commission, making it more internally coherent and, in addition, externally able to justify its policy decisions by the means of evaluation. There is thus evidence that the activism of the Commission to institute evaluation is generating effects, such as a considerably lower number of legislative proposals. Likewise, Hoerner (2019) demonstrates how political parties have used evaluations in order to pursue their political interests.

But actors with an interest in evaluation may also be constrained by their substantive policy field, as Pattyn et al. (2019) demonstrate in the case of Flanders in Belgium. Some policy fields, such as the environment or mobility and public works, drew a large number of demands for evaluations by ministers. Furthermore, the evidence suggests that ministers cite more evaluation demand in policy fields where there is substantial EU funding, pointing to multilevel interactions.

3.3 Evaluation Design: the Challenges in Evaluation and Impact Assessment

There are multiple ways in which politics and evaluation design, that is, its theory and methods, are intertwined. For starters, Dahler-Larsen and Sundby (2019) remind us that “hope is not the same as reality” (see article in this special issue) when it comes to producing high-quality evaluations. They argue that low evaluability can at times be viewed as politically propitious if the goal is to leave a policy unchanged. Drawing on the example of EU occupational health and safety regulation and its implementation in Denmark, they demonstrate that the nature of the legislation can make evaluation difficult, for example when there are no clear targets to evaluate against, when there is no clear programme theory, and when there is little focus on effectiveness and little data available. In sum, significant methodological challenges

may lie in the wait long before an evaluation begins; for those interested in keeping the status quo and avoiding potentially uncomfortable evaluation results, this may be a desired outcome (Dahler-Larsen and Sundby 2019).

A second theme emerging from this special issue is a focus on data quality and collaboration between different data-generating or data-sharing institutions. For example, Potluka (2019) demonstrates that although more data are becoming available in the area of EU environmental cohesion policy in the Czech Republic, sometimes their comparability is limited, and it can be difficult to obtain data from different public institutions. Such issues generate significant challenges for evaluators. Problems with comparability and quality have also long bedevilled climate policy monitoring in the EU, as Schoenefeld et al. (2019) demonstrate in their analysis. The authors were unable to compare quantitative *ex ante* policy estimates across sectors with national targets; however, increases in quantification over time demonstrate learning effects among the member states.

Finally, de Francesco (2019) demonstrates a certain level of stability of evaluation practice, in terms of its evaluation approaches and methods, in a meta-analysis of 52 evaluations focusing on EU railway policy. Even though new public management and efforts at evidence-based policy making had long arrived in the EU, he demonstrates that this has not driven significant changes in evaluation practice. Therefore, institutional changes (see Sect. 3.1 above) do not necessarily drive changes in evaluative practice, necessitating careful empirical analysis.

3.4 Evaluation Purpose and Use: the Strategies of Political Actors

Policy evaluations are neither merely technical tools nor are they unpolitical. The contributions to this special issue identified several ways in which the purposes and uses of evaluation intertwine with politics. In the case of Flanders in Belgium, Pattyn et al. (2019) analysed to what extent government ministers focused on policy planning, accountability, and policy learning in their evaluation announcements and found that about one-third of the evaluations were orientated towards policy planning, two-thirds towards learning, and only a small number towards accountability. Therefore, the professed claims that new public management reforms generate greater accountability do not appear to be true in this case, although a focus on learning and policy improvement can point to potentially positive impacts of evaluation.

Strategic use of evaluation results also becomes very clear when considering parliaments. Studying six national parliaments over 20 years, Hoerner (2019) demonstrates that members of parliament from Eurosceptic parties were particularly prone to citing evaluations of EU policies, while their counterparts from Europhile parties were much less inclined to do so. Furthermore, the more Eurosceptic the environment, the more evaluations were cited. According to Hoerner (2019), this suggests a high level of politicisation of evaluations. These findings chime with those of Dahler-Larsen and Sundby (2019), who similarly suggest that low evaluability can be a tool to drive certain types of (non-)use of evaluation. In sum, there are multiple purposes and uses of policy evaluation; the contributions in this special issue have

focused on providing concrete, empirical evidence in this area, where politics have often been cited but rarely been demonstrated.

4 Conclusions and Future Directions

Policy evaluation is a growing area of policy-making, conceptual debate, and practice in the EU. Related activities take place in various spaces, contexts, and communities, a situation that has, in turn, led to considerable fragmentation. This special issue addresses this problematic state of affairs by bringing together insights from political science and evaluation studies. It structures these debates around four focal areas, namely (1) evaluation institutions, (2) evaluation actors, (3) evaluation design (methods and theory), and (4) evaluation purpose and use. Policy evaluation is clearly a growing field of conceptual development and practice in the EU, but understanding its totality necessitates more systematic approaches and work across different communities. We hope that this special issue contributes to this endeavour.

One concluding observation is that institutional dynamics of policy evaluation generally remain under-researched, as evidenced not only in this special issue but also more broadly in relevant public policy journals. Evaluation has so far grown organically and often in an ad hoc fashion, driven by demands in particular substantive policy fields, by political pressures, or by the skilful activities of policy entrepreneurs. This has generated a veritable patchwork of activity, with systematising efforts so far largely focusing on evaluation design as driven by national and European evaluation organisations. The nascent efforts by the European Commission, the European Environment Agency, and several national governments to systematise evaluation in various ways deserve more attention (for an international review in the area of energy policy, see Schoenefeld and Rayner 2019). Efforts such as the BRR (Jankauskas and Eckhard 2019), with its specific expression of the Regulatory Scrutiny Board (Radaelli 2018), represent profound efforts at reshaping EU politics and policy making by the means of evaluation and defining of the overarching rules of such processes. To remain at the forefront of developments in the field of evaluation, scholars not only need to understand them better, but they also must develop an ability to articulate governance alternatives (Schoenefeld and Jordan 2017). Comparative perspectives may be a useful approach to explore the rule changes enacted at various governance levels.

Greater cross-fertilisation between the political science and the evaluation communities will be necessary to achieve such progress. Many of the evaluation design advances, discussed in journals such as *Evaluation* and the *American Journal of Evaluation*, are likely to bear relevance for political scientists seeking to understand public policies, their underlying mechanisms, and their effects. Evaluation scholars and evaluators stand to gain from insights into political science research on institutions; actors and their characteristics, including political behaviour; and also novel theoretical lenses and conceptual frameworks. Decades ago, Carol Weiss highlighted the political nature of policy evaluation (Weiss 1970), but conceptual developments that address the mounting fragmentation in knowledge on policy evaluation have been few. We hope that this special issue takes one step in the right direction of

rectifying this state of affairs. We would like to invite scholars to move beyond our categorisation, addressing issues such as the relationship between power and evaluation, questions of agency in politics, and the longer-term role of evaluation as a governance tool in the EU.

Furthermore, a range of rapid technological developments, including digitalisation, big data and artificial intelligence are likely to become new opportunities for evaluators in terms of assessing policy impact, as well as for those seeking to understand politicisation. Hoerner's (2019) contribution in this special issue demonstrates the valuable insights that automated, large-scale quantitative analyses can deliver, and future scholars should extend such approaches to other areas of social life beyond formal political institutions, such as social media.

At a time when the EU institutions have come under intense pressure in these times of austerity, rising populism, Euroscepticism, and growing political and societal demands to deliver in substantial policy fields such as climate change, a greater focus on evaluation is paramount. Evaluation could be the key to accountability and learning, and in time to more efficient, cost-effective, and sustainable policy, as well as to achieving greater impact. Our combined knowledge on such efforts—as well as potential alternatives—is only just emerging in a more systematic fashion. Whether these efforts will be successful remains the stuff of future research—endeavours to which, we hope, this special issue contributes a solid foundation.

Acknowledgements We would like to thank all the contributors for initially presenting their ideas at a dedicated 2-day workshop in Brussels in January 2018 and then for staying with us until the completion of this special issue. We are grateful to the PVS editorial board for their trust in our special issue idea, and to Editor-in-Chief Prof. Dr. Michèle Knodt, Managing Editor Katharina Kleinschmitzer and Dr. Christian Adam for their encouragement and patience throughout the review and production process. The Brussels workshop was financed by funds raised by Paul Stephenson from the Faculty of Arts and Social Sciences at Maastricht University, CERIM (Centre for European Research in Maastricht), and SWOL (Stichting Universiteitsfonds Limburg), and funds raised by Frans Leeuw from the Faculty of Law at Maastricht University. The authors are grateful to all four bodies whose financing made it possible to bring the authors of the special issue together. Jonas Schoenefeld would like to acknowledge support from the German Federal Ministry of Education and Research (Reference: 03SFK4P0, Consortium ENavi, Kopernikus).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adelle, Camilla, Andrew Jordan, and John Turnpenny. 2012. Proceeding in parallel or drifting apart? A systematic review of policy appraisal research and practices. *Environment and Planning C: Government and Policy* 30(3):401–415.
- Alkin, Marvin C., and Jean A. King. 2017. Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation* 38(3):434–450.
- Alkin, Marvin C., and Sandy M. Taut. 2003. Unbundling evaluation use. *Studies in Educational Evaluation* 29:1–12.
- Batterbury, Sarah C. E. 2006. Principles and purposes of European Union cohesion policy evaluation. *Regional Studies* 40(02):179–188.
- Bennett, Colin J., and Michael Howlett. 1992. The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences* 25(3):275–294.

- Bovens, Mark. 2010. Two concepts of accountability: accountability as a virtue and as a mechanism. *West European Politics* 33(5):946–967.
- Bovens, Mark, Paul 't Hart, and Sanneke Kuipers. 2006. The politics of policy evaluation. In *The Oxford handbook of public policy*, ed. Michael Moran, Martin Rein, and Robert E. Goodin, 319–335. Oxford: Oxford University Press.
- Bussmann, Werner. 2005. Typen und Terminologie von Evaluationsklauseln. *LeGes: Gesetzgebung & Evaluation* 16(1):97–102.
- Camisão, Isabel, and Paulo Vila Maior. 2019. Failure or success: assessing the European Commission's new strategy to foster EU's economic recovery. *Journal of European Integration*. <https://doi.org/10.1080/07036337.2019.1658750>.
- Christie, Christina A. 2007. Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation* 28(1):8–25.
- Dahler-Larsen, Peter. 2011. *The evaluation society*. Palo Alto: Stanford University Press.
- Dahler-Larsen, Peter, and Anna Sundby. 2019. 'If it ain't evaluated, don't fix it!' The politics of evaluation in occupational health and safety. *Politische Vierteljahresschrift* 60(4). <https://doi.org/10.1007/s11615-019-00190-w>.
- European Commission. 2017. 'Better Regulation Toolbox', to complement the better regulation guideline presented in SWD(2017) 350—Commission Staff Working Document, Brussels, 7 July 2017. https://ec.europa.eu/info/sites/info/files/better-regulation-toolbox_2.pdf. Accessed 15 October 2019.
- Fischer, Frank. 2006. *Evaluating public policy*. Mason: Cengage Learning.
- de Francesco, Fabrizio. 2019. Following the evidence practice: An analysis of evaluation studies on EU railway policy. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00191-9>.
- Fratesi, Ugo, and Fiona G. Wishlade. 2017. The impact of European Cohesion Policy in different contexts. *Regional Studies* 51(6):817–821.
- Furubo, Jan E., Rist, Ray C., and Sandahl, Rolf. 2002. *International atlas of evaluation*. Transaction Publishers: New Jersey.
- Grande, Edgar. 2012. Governance-Forschung in der Governance-Falle? – Eine kritische Bestandsaufnahme. *Politische Vierteljahresschrift* 53(4):565–592.
- Hellstern, Gerd-Michael, and Hellmut Wollmann. 1986. The contribution of evaluation to administration. In *City planning: administration and participation*, ed. Chiranj S. Yadav, 68–91. New Delhi: Concept Publishing.
- Henry, Gary T., and Melvin M. Mark. 2003. Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation* 24(3):293–314.
- Hill, Michael, and Frédéric Varone. 2017. *The public policy process*. Oxon: Routledge.
- Hoerner, Julian M. 2019. Adding fuel to the flames? Politicisation of EU policy evaluation in national parliaments. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00213-6>.
- Hoerner, Julian, and Paul Stephenson. 2012. Theoretical perspectives on approaches to policy evaluation in the EU: The case of cohesion policy. *Public Administration* 90(3):699–715.
- Højlund, Steven. 2014a. Evaluation use in evaluation systems—the case of the European Commission. *Evaluation* 20(4):428–446.
- Højlund, Steven. 2014b. Evaluation use in the organizational context—changing focus to improve theory. *Evaluation* 20(1):26–43.
- Hooghe, Liesbet, and Gary Marks. 2009. A postfunctionalist theory of European integration: From permissive consensus to constraining dissensus. *British Journal of Political Science* 39(1):1–23.
- Huitema, Dave, Andrew Jordan, Eric Massey, Tim Rayner, Harro Van Asselt, Constanze Haug, Roger Hildingsson, Suvi Monni, and Johannes Strippel. 2011. The evaluation of climate policy: Theory and emerging practice in Europe. *Policy Sciences* 44(2):179–198.
- Jacob, Steve, Sandra Speer, and Jan-Eric Furubo. 2015. The institutionalization of evaluation matters: Updating the 10 years later. *Evaluation* 21(1):6–31.
- Jankauskas, Vytautas and Steffen Eckhard. 2019. International bureaucracies as strategic actors: How the better regulation reform strengthens the European Commission. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00189-3>.
- King, Jean A., and Marvin C. Alkin. 2018. The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation* 40(3):431–458.
- Kingdon, John W. 2011. *Agendas, alternatives, and public policies*, 2nd edn., London: Longman.
- Kirkhart, Karen E. 2000. Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation* 2000(88):5–23.
- Leeuw, Frans L. 1994. *Can governments learn? Comparative perspectives on evaluation and organizational learning*. New Brunswick: Transaction.

- Leeuw, Frans L., and Stewart I. Donaldson. 2015. Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation* 21(4):467–480.
- Leeuw, Frans L., and Jan-Eric Furubo. 2008. Evaluation systems what are they and why study them? *Evaluation* 14(2):157–169.
- Levy, Roger P. 2006. European Commission overload and the pathology of management reform: Garbage cans, rationality and risk aversion. *Public Administration* 84(2):423–439.
- Majone, Giandomenico. 1989. *Evidence, argument and persuasion in the policy process*. New Haven: Yale University Press.
- Mastenbroek, Ellen, Stijn van Voorst, and Anne Meuwese. 2016. Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy* 23(9):1329–1348.
- Miller, Robin L. 2010. Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation* 31(3):390–399.
- Olejniczak, Karol. 2013. Mechanisms shaping an evaluation system—a case study of Poland 1999–2010. *Europe-Asia Studies* 65(8):1642–1666.
- Patton, Michael Quinn. 2008. *Utilization-focused evaluation*. Los Angeles: Sage.
- Pattyn, Valérie, Bart de Peuter, and Marleen Brans. 2019. Why do ministers ask for policy evaluation studies? The case of the Flemish government. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00211-8>.
- Pawson, Ray. 2005. *Evidence-based policy: A realist perspective*. London: SAGE.
- Peen Rodt, Annemarie. 2014. *The European Union and military conflict management: Defining, evaluating and achieving success*. London: Routledge.
- Pleger, Lyn, and Fritz Sager. 2018. Betterment, undermining, support and distortion: A heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning* 69:166–172. <https://doi.org/10.1016/j.evalprogplan.2016.09.002>.
- Potluka, Oto. 2019. Overcoming resistance to evaluating the environmental impacts of EU cohesion policy. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00210-9>.
- Power, Michael. 1997. *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Radaelli, Claudio M. 2018. Halfway through the better regulation strategy of the Juncker Commission: What does the evidence say? *JCMS-Journal of Common Market Studies* 56(S1):85–95.
- Radaelli, Claudio M., and Claire A. Dunlop. 2013. Learning in the European Union: Theoretical lenses and meta-theory. *Journal of European Public Policy* 20(6):923–940.
- Raimondo, Estelle. 2018. The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1):26–41.
- Rogers, Patricia J., Anthony Petrosino, Tracy A. Huebner, and Timothy A. Hacsí. 2000. *Program theory in evaluation: Challenges and opportunities*. San Francisco: Jossey-Bass.
- Sabatier, Paul A. 2007. *Theories of the policy process*. Boulder: Westview Press.
- Sanderson, Ian. 2002. Evaluation, policy learning and evidence-based policy making. *Public Administration* 80(1):1–22.
- Scharpf, Fritz W. 2009. Legitimacy in the multilevel European polity. *European Political Science Review* 1(2):173–204.
- Schmidt, Vivien A. 2013. Democracy and legitimacy in the European Union revisited: Input, output and ‘throughput’. *Political Studies* 61(1):2–22.
- Schoenefeld, Jonas J., and Andrew J. Jordan. 2017. Governing policy evaluation? Towards a new typology. *Evaluation* 23(3):274–293.
- Schoenefeld, Jonas J., and Andrew J. Jordan. 2019. Environmental policy evaluation in the EU: Between learning, accountability, and political opportunities? *Environmental Politics* 28(2):365–384.
- Schoenefeld, Jonas J., and Tim Rayner. 2019. Monitoring energy policy. In *Handbook of energy governance in Europe*, ed. Michèle Knodt, Jörg Kemmerzell. Cham: Springer. https://doi.org/10.1007/978-3-319-73526-9_43-1.
- Schoenefeld, Jonas J., Kai Schulze, Mikael Hildén, and Andrew J. Jordan. 2019. Policy monitoring in the EU: The impact of institutions, implementation and quality. *Politische Vierteljahresschrift* 60(4). <https://doi.org/10.1007/s11615-019-00209-2>.
- Schoenefeld, Jonas J. 2018. *Evaluation in polycentric governance systems: Climate change policy in the European Union*. (PhD Dissertation). University of East Anglia, Norwich. <https://ueaeprints.uea.ac.uk/id/eprint/69535/1/2018SchoenefeldJPhD.pdf>, accessed 26 Nov 2019.
- Schrefler, Lorna. 2016. *Evaluation in the European Commission: Rolling check-list and state of play*. Brussels: European Parliamentary Research Service.
- Scott, Richard W. 2001. *Institutions and organizations*, 2nd edn., Thousand Oaks: Sage.

- Scriven, Michael. 1981. *Evaluation thesaurus*. Newbury Park: Sage.
- Shadish, William R., Thomas D. Cook, and Laura C. Leviton. 1991. *Foundations of program evaluation: Theories of practice*. Newbury Park: Sage.
- Stame, Nicoletta. 2003. Evaluation and the policy context: The European experience. *Evaluation Journal of Australasia* 3(2):37–43.
- Stame, Nicoletta. 2006. Governance, democracy and evaluation. *Evaluation* 12(1):7–16.
- Stephenson, Paul. 2015. Reconciling audit and evaluation? The shift to performance and effectiveness at the European Court of Auditors. *European Journal of Risk Regulation* 1:79–89.
- Stephenson, Paul. 2017. Why better regulation demands better scrutiny of results. *European Journal of Law Reform* 19(1–2):97–120.
- Stern, Elliot. 2009. Evaluation policy in the European Union and its institutions. *New Directions for Evaluation* 2009(123):67–85.
- Stockmann, Reinhard, Meyer, Wolfgang, and Lena Taube (eds.). 2020. *The Institutionalisation of Evaluation in Europe*. Palgrave Macmillan: London.
- Strassheim, Holger, and Pekka Kettunen. 2014. When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy: A Journal of Research, Debate and Practice* 10(2):259–277.
- Streeck, Wolfgang, and Armin Schäfer. 2013. *Politics in the age of austerity*. Cambridge: Polity Press.
- The LSE GV314 Group. 2013. Evaluation under contract: Government pressure and the production of policy research. *Public Administration* 92(1):224–239.
- Vaessen, Jos, and Frans L. Leeuw. 2010. *Mind the gap: Perspectives on policy evaluation and the social sciences*. New Brunswick: Transaction.
- Vedung, Evert. 1997. *Public policy and program evaluation*. New Brunswick: Transaction.
- Weiss, Carol H. 1970. The politicization of evaluation research. *Journal of Social Issues* 26(4):57–68.
- Weiss, Carol H. 1993. Where politics and evaluation research meet. *Evaluation Practice* 14(1):93–106.
- Wenger, Etienne. 2011. Communities of practice: A brief introduction. <https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/11736/A%20brief%20introduction%20to%20CoP.pdf>. Accessed 3 September 2019.
- Widmer, Thomas. 2004. The development and status of evaluation standards in Western Europe. *New Directions for Evaluation* 2004(104):31–42.