



Policy Monitoring in the EU: The Impact of Institutions, Implementation, and Quality

Jonas J. Schoenefeld · Kai Schulze · Mikael Hildén ·
Andrew J. Jordan

Published online: 28 November 2019
© The Author(s) 2019

Abstract Policy monitoring is often seen as a crucial ingredient of policy evaluation, but theoretically informed empirical analyses of real-world policy monitoring practices are still rare. This paper addresses this gap by focusing on climate policy monitoring in the European Union, which has a relatively stringent system of greenhouse gas monitoring but a much less demanding approach to monitoring policies. It explores how institutional settings, policy implementation, and the quality of information may impact the practices and politics of policy monitoring. Drawing on quantitative regression models and qualitative interviews, it demonstrates that policy monitoring has evolved over time and is itself subject to implementation pressures, but also exhibits learning effects that improve its quality. In further developing both everyday policy monitoring practices and academic understanding of them, there is a need to pay attention to their design—specifically, the impact of any overarching rules, the institutional support for implementation, and the criteria governing the quality of the information they deliver. In short, policy monitoring should be treated as a governance activity in its own right, raising many different design challenges.

Keywords Policy evaluation · Policy implementation · Climate change · Monitoring mechanism · European Union

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11615-019-00209-2>) contains supplementary material, which is available to authorized users.

J. J. Schoenefeld (✉) · K. Schulze
Darmstadt, Germany
E-Mail: schoenefeld@pg.tu-darmstadt.de

J. J. Schoenefeld · A. J. Jordan
Norwich, United Kingdom

M. Hildén
Helsinki, Finland

Policy-Monitoring in der EU: Der Einfluss von Institutionen, Implementation und Qualität

Zusammenfassung Obwohl Policy-Monitoring oft als ein wichtiger Beitrag zur Policy-Evaluation verstanden wird, sind theoretisch fundierte empirische Analysen bestehender Policy-Monitoringpraktiken noch immer eine Seltenheit. Dieser Beitrag widmet sich dieser Forschungslücke und betrachtet dazu das weniger anspruchsvolle und noch relativ junge Monitoring von Klimapolitiken in der Europäischen Union, welches neben einem viel stringenteren System für die Überwachung der nationalen Treibhausgasemissionen existiert. Die vorliegende Analyse beleuchtet insbesondere, welchen Einfluss Institutionen, Implementation und Informationsqualität auf die Praxis und die Politik des Policy-Monitorings haben. Quantitative Regressionsanalysen und qualitative Interviews zeigen, dass sich das Monitoring von Klimapolitik im Laufe der Zeit institutionell verändert hat, selbst unter Implementationsdruck steht, aber auch, dass Lerneffekte die Qualität des Policy-Monitorings verbessern. Zur Weiterentwicklung von Policy-Monitoringpraktiken und ihrer wissenschaftlichen Erforschung sollte daher ein besonderes Augenmerk auf das Design von Policy-Monitoringsystemen gelegt werden. Dazu gehören umfassende Regeln, institutionelle Unterstützung der Implementation und Kriterien der Informationsqualität. Kurz gesagt, sollte Monitoring als eine eigenständige Governance-Aktivität verstanden werden, die viele Herausforderungen an das Policy-Design stellt.

Schlüsselwörter Policy-Evaluation · Policy-Implementation · Klimawandel · Monitoring Mechanism · Europäische Union

1 Introduction

Many sources in the academic literature recognise “monitoring” as a crucial ingredient of governance and, notably, policy evaluation (e.g., Aldy 2014; Dunn 2018; Jensen 2007; Vedung 1997). This paper starts from the premise that the political dynamics, institutional questions, and, at times, technical issues that have often been identified in other policy making stages such as agenda setting and policy formulation, are also likely to unfold in relation to policy monitoring practices. It is thus somewhat surprising that so few researchers have explored existing practices of policy monitoring in detail (Tosun 2012). Analysing them may have been neglected because policy monitoring is typically portrayed as a rather mundane technical activity. Hence, difficulties that are encountered in policy monitoring may not be fully understood. Building on Carol Weiss’s famous argument, we assume that policy monitoring is unlikely to be “[a] neutral, antiseptic, laboratory-type entit[y]” (Weiss 1993, p. 94–95). This paper breaks new ground by exploring the extent to which policy monitoring can be subjected to the same types of political analysis that are routinely applied to other areas of politics and policy making.

Policy monitoring can be an important ingredient of policy evaluation, which has been defined as a broader set of approaches to “determining the merit or worth or value of [public policy]; or the product of that process” (Scriven 1981, p. 53). Ac-

According to the Organisation for Economic Co-operation and Development (OECD), policy monitoring may be understood as “a continuous process of collecting and analyzing data to compare how well a project, program, or policy is being implemented against expected results” (OECD-DAC 2002, p. 30). While Vedung (1997, chap. 9) argues that policy monitoring mainly focuses on the implementation process, leading from an intervention theory to data collection and assessment, we follow Rossi et al. (2018, chaps. 4 and 5), who distinguish between process and outcome monitoring. In line with the focus of this special issue (Stephenson et al. 2019), we concentrate on the European Union’s (EU) deeply institutionalised efforts to monitor climate policies in its member states. In this area of policy, process monitoring may, for example, centre on whether subsidies for renewable energies have been paid out and how, whereas outcome monitoring would refer to assessing whether efforts to expand renewable energy have actually reduced greenhouse gas emissions. This paper concentrates on outcome monitoring, which is normally conducted with the help of indicators in order to understand the effect of public policies.

In the area of climate change policy, scholars have highlighted the need for policy monitoring in order to ensure effectiveness and transparency (e.g., Aldy 2018), but emerging studies have also unpacked some of the challenges that emerge in policy monitoring practice—many of them, such as variations in data quality or power struggles between policy monitoring actors, often being linked to politics and institutions (Niederberger and Kimble 2011; Schoenefeld and Jordan 2017; Schoenefeld et al. 2018). More conceptual analyses have explored the extent to which policy monitoring includes (or should include) processes of data collection and/or data analysis (see Schoenefeld and Rayner 2019). Today, policy monitoring is often conducted within—but increasingly also beyond—nation states in order to achieve international/supranational goals, such as enforcing international agreements (see Schoenefeld and Jordan 2017) or ensuring the implementation of EU policies (Tosun 2012; Verdun and Zeitlin 2018).

It has been suggested that policy monitoring should be based on unambiguous procedures and variables that allow policy comparisons across countries and tracking over time (Öko-Institut et al. 2012). The EU has developed its legal base of climate monitoring through political negotiations and by providing guidance to member states in order to systematise common practices. It first created a monitoring mechanism for national-level greenhouse gas emissions with a view to reporting them to the institution of the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 (Bodansky 1993; Hildén et al. 2014; Hyvarinen 1999). The monitoring has since been extended to include requirements on monitoring climate public policies in specific policy areas (Schoenefeld et al. 2018).

We focus on the monitoring of what the United Nations (UN) and EU refer to as the “policies and measures” that member states have or expect to put in place in order to achieve emission reductions (“policy monitoring”).¹ Due to its central role

¹ Article 3 (14) of Regulation (EU) No. 525/2013 simply states that “‘policies and measures’ means all instruments which aim to implement commitments under Article 4 (2) (a) and (b) of the UNFCCC, which may include those that do not have the limitation and reduction of greenhouse gas emissions as a primary objective.”

in EU climate governance, the monitoring mechanism (EU Regulation 525/2013)² is a suitable case to start exploring the dynamics of policy monitoring in more detail. As Schoenefeld and Rayner (2019) explain, compared to the long-standing practice of monitoring greenhouse gases at the national level, monitoring public policies generated new challenges, starting with the difficulty of assigning emissions reductions to individual policies in complex and widely interconnected governance systems. The harmonisation of policy monitoring practice and ensuring sufficient levels of quality were also demanding, highlighting the importance of institutional complexity and multilevel dynamics.

In this paper, we examine to what extent an analysis of institutional settings, policy implementation and quality of policy monitoring can help us understand member state policy monitoring practice in the context of mandatory reporting to the EU institutions. Our approach is novel in the way it combines several separate bodies of literature. We explore the impact of the institutional settings, policy implementation and quality in order to better understand the underlying patterns and drivers of policy monitoring. To do so, we triangulate different methods, including document and exploratory quantitative analysis, as well as qualitative key informant interviews with four staff members of the European Environment Agency (EEA), which manages the monitoring mechanism. The final section synthesises our findings, provides insights for practitioners, and concludes with an assessment of future research needs.

2 The Monitoring Mechanism and its Outputs

Modern-day greenhouse gas and climate policy monitoring centres on a UN framework based on internationally agreed monitoring standards developed by the Intergovernmental Panel on Climate Change (IPCC; see Eggleston et al. 2006). Historically, a focus on monitoring and its use to evaluate policies was not the preferred policy outcome for many EU countries as they prepared for the foundational 1992 Rio de Janeiro Earth Summit, which established the UNFCCC. On the contrary, it was only because EU countries could not agree on substantive climate policies that they opted to focus on monitoring instead (Haigh 1996). Monitoring was perceived to be more technical and thus politically more tractable to progress in a difficult diplomatic environment (see Bodansky 1993, p. 451). What has since emerged in the EU is a relatively stringent system of greenhouse gas monitoring along with a much less demanding and still emerging approach to monitoring of public policies, which countries started putting in place in order to reduce emissions—together known as the monitoring mechanism (Hyvarinen 1999). The EU has revised its monitoring mechanism several times—by and large in order to comply with international agreements. In 2013 (see EU Regulation No. 525/2013) more detailed reporting requirements were introduced, especially on policies and measures (Schoenefeld et al. 2018). The EEA compiles the policy monitoring data from the member states, checks their quality, and then submits them to the European Commission, the EU's main ex-

² The Monitoring Mechanism Regulation (MMR) EU 525/2013 will be replaced by the Energy Union Governance Regulation (EU 2018/1999) from 2021, but all elements of the MMR have been carried over.

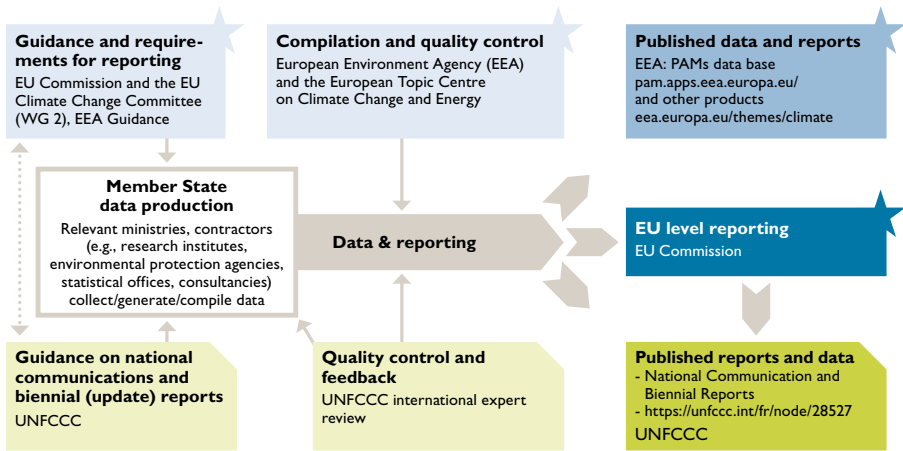


Fig. 1 Schematic of climate policy reporting in the European Union and the United Nations Framework Convention on Climate Change. European Union and United Nations monitoring and reporting occur in parallel and interact. The stars in the corners of the boxes indicate the EU process

ecutive body. The Commission uses the data in EU-wide reporting to the UNFCCC in accordance with international reporting guidelines. The member states also report the details of their individual policies and measures directly to the UNFCCC. Fig. 1 outlines the institutional actors involved in the climate policy monitoring, as well as the relationships between them.

The monitoring mechanism has centred on reporting *ex ante* (i.e., forward-looking) predictions of policy impact, whereas *ex post* (i.e., retrospective) reporting is optional since the mandatory national greenhouse gas inventories provide information on overall climate policy success. This means that the EEA Policies and Measures Database³ contains much more information on *ex ante* projections than on *ex post* policy-specific data (Schoenefeld et al. 2018). In addition, the policy-based data emerging from the mechanism since the late 2000s contain considerable inconsistencies. For example, the number of policies and measures reported has fluctuated widely, and the plausibility of some past or projected greenhouse gas reductions has been questioned (AEA et al. 2009; ETC and ACC 2012; European Environment Agency 2016; Farmer 2012; Hildén et al. 2014; Öko-Institut et al. 2012; Schoenefeld et al. 2018). These issues still remain, even though a quality assurance and control regimen was created in the mid-2000s to improve the quality of submissions (Schoenefeld et al. 2018).

Our analysis uses official policy monitoring data submitted by the EU member states across all climate policies and measures, and thus all relevant policy sectors (such as energy and transport) as well as instrument types (such as regulatory and information-based).⁴ We consider only policies outside the EU Emissions Trading

³ <http://pam.apps.eea.europa.eu>.

⁴ Because the monitoring mechanism data includes more fine-grained data (such as qualitative, country-based reports) that could not be considered here for space reasons, the discussion section contains a range

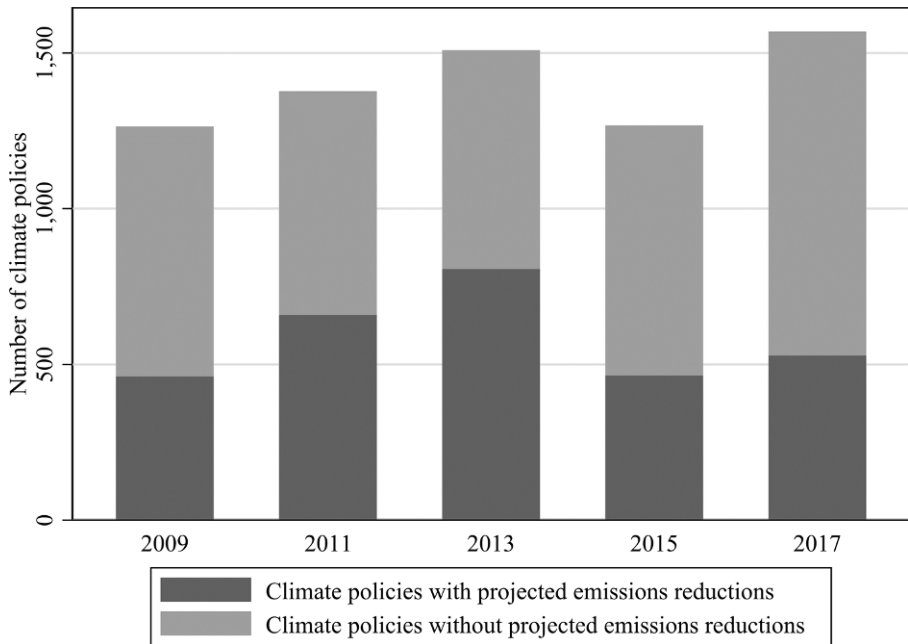


Fig. 2 Number of climate policies and measures reported in the monitoring mechanism (European Union Emissions Trading System excluded)

System (ETS) because the ETS is an EU-wide, quantity-based policy instrument, which effectively prescribes the reductions at the EU level. Fig. 2 shows that the grand total of climate policies (outside the ETS) reported each year is now around 1500 across all member states. The shading of the bars indicates the proportion of climate policies for which the member states provided quantitative projected emissions reductions relative to a counterfactual. Thus, the projection can be a positive number (i.e., emissions decreases), zero, or a negative number (i.e., emissions increases). The proportion of quantitative reporting increased steadily to about half in 2013 but dropped significantly in 2015 and has since continued at a lower (but again growing) rate.

Fig. 3, in turn, presents the same information for the individual member states. Notably, there is significant variance within and across member states, suggesting some implementation issues. Some of this variation may originate from the unclear definition of what a climate policy actually is (see Footnote 1 above). For instance, Finland has quantified an increasing share of their climate policies, while the Netherlands has recently reported additional policies without estimates of their emission reductions. Others, such as Poland, have not reported any quantified data except for

of suggestions for further research. Going forward, we therefore use the total number of policies as well as the number of policies with ex ante projections of emissions reductions for 2020 that the monitoring mechanism generated in 2009, 2011, 2013, 2015, and 2017. Member states are legally obliged to report on climate change policies every other year. Earlier quantitative data are not available.

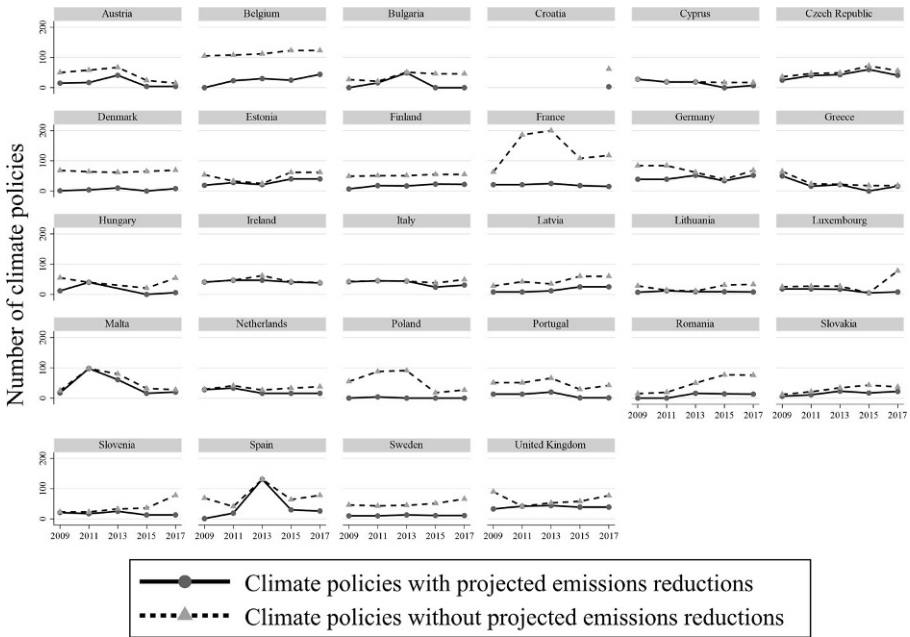


Fig. 3 Number of policies with and without quantified projected emissions reductions by member state

2011. In order to better understand this variance, the following section turns to three theoretical/conceptual approaches.

3 Understanding Policy Monitoring: the Impact of Institutions, Implementation, and Quality

This section unpacks three perspectives on policy monitoring, namely institutions, policy implementation, and quality, with a view to identifying patterns in ex ante policy monitoring practices. While the three perspectives have emerged from different literatures, this section demonstrates that they all hold considerable explanatory potential in the area of policy monitoring. This section reviews extant knowledge in each area, linking where appropriate with literature on reporting and policy evaluation.

3.1 Institutions

Policy monitoring typically takes place in specific institutional settings, often involving numerous actors, which may spread across different governance levels. It fits the description of March and Olsen (2009, p. 3), who note that “an institution is a relatively enduring collection of rules and organized practices, embedded in structures of meaning and resources that are relatively invariant in the face of turnover of individuals and relatively resilient to the idiosyncratic preferences and expecta-

tions of individuals and changing external circumstances.” From this perspective, it becomes important to understand to what extent the aforementioned “rules and organised practices” impact on policy monitoring systems, both in terms of providing stability but also allowing a certain amount of change over time. The tension between stability and change has long been noted in relevant literature, with Lindner (2003, p. 913) defining institutional change as “the introduction of new rules or rule interpretations that supplement or replace existing rules and interpretations.” It is thus relevant to uncover to what extent the variations among member state policy monitoring may originate from institutional differences and how rule changes impact policy monitoring practice. Doing so involves paying close attention to the legal and institutional features of the monitoring mechanism, such as relevant experience in the member states, as well as the evolution of such elements over time.

3.2 Implementation

Policy monitoring is often conceptualised as a means to improve implementation. However, policy monitoring systems, in and of themselves, are not necessarily self-implementing. Studies on environmental policy in the EU have highlighted several critical political aspects of policy implementation, especially related to the transposition of EU directives into national law (see Jordan and Tosun 2013; Knill and Liefferink 2007). Important factors in transposition (and, to a certain degree, also in the formulation of regulations) related to member states’ *willingness to comply* include party politics, institutional goodness of fit (or misfit), public opinion, and the presence or absence of interest groups. By contrast, states’ *capacity to comply* may be influenced by the number of veto players as well as administrative capacity (Treib 2014). More recent literature has also pointed to the important role of EU agencies in implementation (Thomann and Sager 2017). However, two shortcomings of this literature are that it has mainly focused on the transposition of EU directives (with a general neglect of the application of EU regulations), and it has also understood policy monitoring as an enforcement tool rather than another site of implementation (Treib 2014). The implementation of the monitoring mechanism may thus depend on its legal base as well as its specific institutional setting. We also expect that learning effects may generate improvements over time, in terms of both quantification and timely reporting, especially because climate policy monitoring requires both significant technical and administrative expertise and experience. The overall number of policies and measures with quantified emissions reductions may be related to the total number of reported policies (i.e., with and without quantifications).

3.3 Quality

Policy monitoring may have different levels of quality. Schwartz and Mayne (2005, p. 1) argue that while “there is now a large supply of evaluative information in the forms of evaluation, performance reporting and performance auditing, relatively little attention has been paid to assuring the quality of this information.” This may be in part because, as Stake and Schwandt (2006, p. 405) explain, “quality is multifaceted, contested, and never fully representable” (see also Dahler-Larsen 2019). Also, much

of the debate on quality has centred on policy evaluation rather than on monitoring. The concept of “quality” has both technical (e.g., needs are met) and cultural dimensions; that is, it has different meanings in different societies (Williams 2005). The quality of policy monitoring data also relates to its relevance for the policy at hand.⁵

What quality is (and for whom) will likely remain contested because it is intimately linked with various potential purposes of policy monitoring and evaluation, including accountability and learning functions, as well as politics (Schoenefeld and Jordan 2019; Schwartz and Mayne 2005). A policy monitoring system may originally be designed for checking policy goal attainment, but once the system starts producing information, this information may be used in many different ways. Therefore, one should pay attention to actors and groups, which may pursue different purposes and have linked interests (such as the EEA, the Commission, and the member states in the current case). Other debates on quality have centred on the nature of policy monitoring/evaluation methods (van Voorst and Mastenbroek 2019) as well as on standards—the latter a discussion and effort that has been around since at least the 1970s (Widmer 2004, 2012). These debates have been echoed by discussions on the quality of climate policy monitoring data (see Wettestad 2007) and concrete efforts to improve (e.g., Öko-Institut et al. 2012). In this paper, we follow what Dahler-Larsen terms “fixation” through quantification as one indicator of quality (2019, p. 142–147). Following this line of reasoning, aspects of quality may be reflected in the effort that countries put into policy monitoring, such as to what extent they seek to quantify policy effects, or the level of resources they invest monitoring policy results. Furthermore, we consider “temporalization”—that is, the timeliness of reporting—as an additional indicator of quality (Dahler-Larsen 2019, p. 139–142).

4 Understanding the Monitoring Mechanism Outputs

4.1 Quantitative Analysis

We carried out exploratory quantitative analyses with a view to identifying potential drivers of policy monitoring, drawing on the three perspectives discussed above. We started by regressing the number of climate policies and measures for which member states reported quantified emissions reductions (projections for the year 2020) in 2009, 2011, 2013, 2015, and 2017 on a series of plausible drivers of policy monitoring practice derived from the literature. We interpret the generation of quantified emissions reductions for individual policies instead of a general aggregate or loose qualitative statements as a sign of serious commitment to policy monitoring and to the implementation of the monitoring decision/regulation.⁶ Moreover, we compiled data on the punctuality of reporting under the monitoring mechanism, our

⁵ We owe this point to one of our reviewers.

⁶ Our dependent variable includes only policies with projected reductions outside the EU Emissions Trading System.

second dependent variable. Member states must report on their climate policies by 15 March every other year, and we categorised countries that delivered their report in the month of March as on time, thus allowing for insignificant administrative delays (for the full data, see Table 4 in the online Appendix). Reporting punctuality can be seen as a crude proxy for the implementation of policy monitoring because states that fail to provide information on time probably face difficulties in organising the policy monitoring and/or lack resources for the task.

Our independent variables (i.e., drivers of policy monitoring) were chosen based on the review in Sect. 3 to reflect variation in institutional settings, policy implementation, and efforts to deliver high-quality policy monitoring. To capture institutional effects, we included a variable indicating previous experience with monitoring and evaluating national climate change policies (none, some, or extensive experience) based on data from a European Commission-funded report on quantifying the emissions effects of policies and measures (AEA et al. 2009). A dummy variable indicates the legislative shift from the Monitoring Mechanism Decision (MMD: 2004–2012) to the Monitoring Mechanism Regulation (MMR: 2013–2019). We expect that the number of policies with quantifications initially decreased with the introduction of the MMR (all else being equal), since it explicitly allows bundling of policy instruments that achieve their effects in combination. We model learning effects through a “trend” variable increasing linearly with each obligatory round of climate change policy reporting from 2009 until 2017. In terms of the capacity to comply with policy monitoring requirements, the regression models include governmental final consumption expenditures as a proxy for the national bureaucracy’s financial capabilities and a measure of political constraints (the expectation being that the more constraints, the less reporting). In terms of willingness to comply we further consider the seat share of green parties and the degree of climate change concern among the national population (both expected to increase quantitative reporting). Finally, we control for per capita CO₂ emissions of each country, its gross domestic product (GDP) per capita, and population size. All time-varying variables were entered with a 1-year lag because most data collection and preparatory work takes place in the year prior to submission. We estimated negative binomial regression models for the number of reported policies and measures with quantified emissions reductions and logit models for the punctuality of reporting. Because our observations within countries are not independent, we used standard errors adjusted for clustering within countries. Table 1 summarises our data, its sources, and the operationalisation of our variables.

Our first regression models (see Table 2) consider potential determinants of the number of policies with quantified, projected emissions reductions (for 2020). They reveal that the number of policies with quantified projections increases with the overall number of reported policies. More precisely, for every additional climate policy reported, the number of quantifications increases by 1.3; or almost 50% for a standard deviation increase in the number of reported policies. Put simply, if a country reports more climate policies, it also quantifies more compared to countries that report fewer policies and correspondingly generate fewer quantifications.

Second, greater experience with policy monitoring and evaluation in 2009 has translated into greater levels of reporting across our observation years, pointing to

Table 1 Summary statistics and data sources

Variable	Description	<i>N</i>	Mean	SD	Min.	Max.	Source
Number of climate policies with reduction projections	Number of reported policies and measures with quantitative estimates of emission reductions	135	21.63	19.05	0	131	EU MM
Number of climate policies (all)	Number of all reported policies and measures	135	52.25	31.24	5	200	EU MM
Report on time	First report submitted to the EEA within March	117	0.63	0.48	0	1	EU MM
Policy evaluation experience	Previous experience with ex post evaluation of climate change policies (as of 2009): none (0), some policies (1), comprehensive (2)	140	0.82	0.81	0	2	AEA et al. 2009
Monitoring Mechanism Regulation	Monitoring Mechanism Regulation	140	0.40	0.49	0	1	Own
Reporting trend	Count of reporting rounds, 2009–2017	140	3.00	1.42	1	5	EU MM
Government expenditure	State expenditures as a percentage of GDP	140	19.97	2.98	12.38	27.37	WDI
Political constraints	Number of independent branches of government with veto power and the distribution of preferences across and within those branches	140	0.48	0.10	0.18	0.71	Henisz (2000)
Green seats	Seat share of green parties in parliament	140	2.54	3.82	0	13.10	Armingeon et al. (2018)
Climate concern	Perceived seriousness of the climate change problem, not at all serious (1) to extremely serious (10)	138	7.46	0.64	5.65	9.40	Eurobarometer
CO ₂ emissions per capita (log)	CO ₂ emissions in metric tonnes per capita, logged	140	15.23	0.90	12.00	16.11	WDI
GDP per capita (log)	GDP per capita, PPP (constant 2011 international \$), logged	140	10.38	0.36	9.63	11.46	WDI
Population (log)	Total population, logged	140	15.88	1.40	12.92	18.23	WDI

SD standard deviation, *EU MM* EU monitoring mechanism, *WDI* World Bank World Development Indicators, *GDP* gross domestic product, *PPP* purchasing power parity

Table 2 Determinants of the number of reported climate policies with reduction projections

	(1)	(2)	(3)
Number of climate policies (all)	0.012*** (0.003)	0.012*** (0.003)	0.013*** (0.004)
Policy evaluation experience	0.275** (0.112)	0.262** (0.111)	0.288* (0.164)
Monitoring Mechanism Regulation	-0.766*** (0.163)	-0.806*** (0.172)	-0.923*** (0.231)
Reporting trend	0.169*** (0.051)	0.163*** (0.057)	0.208*** (0.066)
Government expenditure	-0.046 (0.035)	-0.056 (0.035)	-0.049 (0.045)
Political constraints	-	0.876 (1.063)	0.643 (0.841)
Green seats	-	-0.017 (0.021)	-0.048 (0.035)
Climate concern	-	-0.044 (0.137)	-0.014 (0.125)
CO ₂ emissions per capita (log)	-	-	-0.075 (0.072)
Gross domestic product per capita (log)	-	-	0.646 (0.403)
Population (log)	-	-	-0.057 (0.106)
Constant	2.863*** (0.770)	3.046* (1.587)	-1.967 (4.800)
Observations	135	135	135
Countries	28	28	28
Alpha	0.742	0.731	0.690
AIC	1091.04	1095.36	1094.71

Negative binomial regression models. Country-clustered standard errors are in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

AIC Akaike information criterion

a certain level of institutional path dependency in policy monitoring. In other words, countries that were ahead 10 years ago remain, by and large, ahead today.

Third, the policy monitoring “trend” variable shows a strongly significant positive association with the dependent variable, hinting at the presence of a learning effect in policy monitoring over time (for further qualitative evidence, see Sect. 4.2). Based on model 3, with every biennial reporting cycle, the number of quantifications increases by around 23% (all other factors held constant). This is natural, as countries that initiate quantification are likely to replicate the approach across policy domains in order to achieve coherent policy monitoring and reporting.

Fourth, the model demonstrates that the introduction of the MMR in 2013 (in effect for the reporting rounds of 2015 and 2017) decreased the expected number of quantified projections by around 60%. In fact, that institutional shift seems to be associated with a reduction both in the total number of reported policies and in the share of policies with quantified projections (Fig. 2). Because we control for the

total number of instruments, we interpret this finding as an adjustment effect related to the aforementioned explicit possibilities for reporting “bundles” of policies and their emissions reductions in the MMR. All other variables, including government expenditure, political constraints, green seats, climate concern, CO₂ emissions, GDP, and population returned nonsignificant results. The online Appendix features several robustness checks such as inclusion of a lagged dependent variable, a government’s environmental policy position, and country fixed effects (see Table 5). These changes do not affect our main findings.

Our second exploratory analysis models determinants of punctual reporting (Table 3). The results should be interpreted with caution, given very limited extant knowledge of the determinants of timely reporting. We observe that prior experience with policy evaluation and monitoring, as well as higher levels of government expenditure, are positively and significantly associated with punctual reporting compared to less experience and lower expenditures. The latter effect, even though statistically significant only at the 10% level, is quite substantial, with a 1% increase in government expenditure being associated with a 26% increase in the odds of punctual reporting. Furthermore, there is a marginally significant effect in model 1, indicating that the introduction of the MMR led to an almost 80% decrease in the odds of punctual reporting. Slower reporting may, at least in part, originate from adjusting to the new policy monitoring system (similar to the arguments on this independent variable above). All the other variables yielded nonsignificant results (see model 3 in Table 3).

4.2 Qualitative Analysis

We interviewed four staff members of the EEA in order to better understand the policy monitoring patterns from the perspectives of institutional settings, implementation, and quality. This section presents our findings. The main *institutional* change at the EU level was the 2013 shift from basing the monitoring mechanism on a decision (“Monitoring Mechanism Decision” [MMD]) to basing it on a regulation (“Monitoring Mechanisms Regulation” [MMR]). Although EEA staff report that the choice of legal act was not a main driver of change,⁷ the new rules do affect the nature of the data. As one EEA staff member put it, “[w]e see the clear improvement in the 2017 reporting in comparison to 2015 reporting cycle. Both quantity and quality of the reported information improved.”⁸ In line with this statement, our quantitative analyses also suggest that an immediate impact of the introduction of the MMR was a decline in quantitative and timely reporting in 2015, but then improvements appeared to materialise in 2017 (see Sect. 4.1). We do not have data on institutional changes in the member states, except that the shift from a decision to a regulation removed potential ambiguities in reporting obligations introduced by national laws.

As regards *implementation*, the Commission opted to implement the monitoring mechanism via a regulation rather than a directive in order to ensure it was applied

⁷ EEA staff member 3, interview 11 January 2019.

⁸ EEA staff member 4, interview 11 January 2019.

Table 3 Determinants of punctual submission of policy monitoring reports

	(1)	(2)	(3)
Number of climate policies (all)	-0.011 (0.009)	-0.012 (0.009)	-0.008 (0.009)
Number of climate policies with reduction projections	-0.004 (0.011)	-0.003 (0.011)	-0.006 (0.012)
Policy evaluation experience	0.710** (0.343)	0.705** (0.355)	0.863** (0.438)
Monitoring Mechanism Regulation	-1.528* (0.857)	-1.342 (1.099)	-1.354 (1.096)
Reporting trend	0.482 (0.331)	0.352 (0.448)	0.372 (0.455)
Government expenditure	0.224* (0.131)	0.226* (0.125)	0.232* (0.127)
Political constraints	-	-0.486 (3.186)	-0.810 (3.243)
Green seats	-	-0.026 (0.059)	-0.029 (0.085)
Climate concern	-	-0.508 (0.552)	-0.460 (0.544)
CO ₂ emissions per capita (log)	-	-0.018 (0.272)	-0.017 (0.273)
Gross domestic product per capita (log)	-	-	-0.063 (0.991)
Population (log)	-	-	-0.146 (0.220)
Constant	-4.572* (2.595)	0.110 (5.375)	2.428 (14.033)
Observations	115	115	115
Countries	28	28	28
AIC	144.36	150.11	153.64

Logit regression models. Country-clustered standard errors are in parentheses

$N = 115$; for the remaining countries/cases, the exact submission time could not be determined (see Table 4 in the Appendix on-line)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

AIC Akaike information criterion

throughout the EU as consistently as possible. In theory, the Commission oversees implementation. In practice, there is no precise technical definition of what counts as a “policy” that a member state is expected to report (see Footnote 1 above). Member states can thus make discretionary choices in reporting. One EEA staff member noted the concerted efforts by the EEA to assist the member states in reporting, for example through the provision of webinars, workshops, and other technical assistance. The new regulation has also strengthened the coupling (or a compliance cycle) between the effort sharing and the greenhouse gas reporting, where a member state’s progress is checked against the targets each year.⁹ European Environment Agency staff members stated that the monitoring system “functions

⁹ EEA staff member 3, interview 11 January 2019.

well” and highlighted that its implementation has helped them to detect (increasing) member state difficulties in reaching the 2020 targets.¹⁰ However, progress checking is done against national aggregate greenhouse gas statistics, and not against the disaggregated policy-based specific data on which this article centres.

Finally, the EEA staff identified four key drivers of policy monitoring *quality*: First, they discussed what demotivates member states. One of the challenges is that “the concrete impact and use of this information, as well as the actual benefits of such reporting, might not always be very explicit from member states’ perspective.”¹¹ So the EEA has attempted to explain the benefits (to them) of climate policy monitoring and reporting, while also conceding that managerial logics around complying with existing guidelines and legislation also continue to play an important role.¹² The second factor concerns the political willingness to engage in policy monitoring. Ex post reporting may be unattractive (see also Schoenefeld and Jordan 2019, p. 370) because, as one EEA staff member put it,

[...] politicians are not always very interested in ex post evaluation results, because this is just about looking at the past and might possibly identify certain failures. Policy makers prefer focusing on future measures and demonstrating that their plans will succeed, for example in reaching their country’s target.¹³

The third factor affecting policy monitoring quality concerns the willingness and ability to mobilise resources for policy monitoring exercises. In some states—especially in times of austerity—policy monitoring may simply not be a key priority, and it may be subject to budget cuts.¹⁴ Fourth, the availability of commonly agreed methodologies affects policy monitoring quality in the EEA’s view. “À la carte” standards and methods, as Schoenefeld and Jordan (2017) have put it, are generally unable to generate consistent data; the EEA staff members stressed the existence of considerable methodological guidance, albeit accessing and using these documents can sometimes challenge the member states.¹⁵

4.3 The Links Between Institutions, Implementation, and Quality

Our quantitative analyses and qualitative findings reveal that institutional settings, implementation, and quality are closely interlinked. This becomes especially apparent in the recent EU efforts to bring together its climate and energy policy in the so-called Energy Union (see Knodt 2019).¹⁶ The regulation on the governance of the Energy Union and climate action (2018/1999) that entered into force in December 2018 integrates two policy areas whose monitoring efforts had developed rather separately at the European level: climate change and environment have been

¹⁰ EEA staff member 3, interview 11 January 2019.

¹¹ EEA staff member 3, interview 11 January 2019.

¹² EEA staff member 1, interview 11 December 2018.

¹³ EEA staff member 3, interview 11 January 2019.

¹⁴ EEA staff member 3, interview 11 January 2019.

¹⁵ EEA staff member 3, interview 11 January 2019.

¹⁶ <https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union>.

the competence of the Commission Directorate-General (DG) Climate Action and DG Environment, whereas energy policy has been the competence of DG Energy. As one EEA staff explained, “The Energy Union [...] integrates different pillars or dimensions that had been considered a bit separately until now.”¹⁷ Another EEA staff member vividly described the challenges of integrating monitoring activities across the two policy areas:

They’d [DG Energy] be organised by the oil sector, the nuclear sector, [...] the renewables sector, and they didn’t really talk to each other and [...] the different parts of the energy industry would have different inroads into DG Energy. But actually get[ting] them to start reporting on outcomes, or environmental outcomes or climate outcomes, is something that [...] has been a big change for them.¹⁸

These sentiments were echoed by another EEA staff member, who explained that “the idea to them [DG Energy] that a small agency in Copenhagen which has environment in its name would start messing around in the high politics of energy was just not OK for them.”¹⁹ The person went on to say that “it’s unbelievable how much time and effort is being spent on just having our agency working together with a DG.”²⁰ Furthermore, the EEA staff member emphasised that understandings of reporting differed significantly in the energy and the climate sectors: while reporting in the energy sector is often industry specific and sometimes based on data from existing sources, climate policy reporting is based on official country-based data that go through quality assurance procedures. Another EEA staff member also stressed the special nature of data emerging from the MMR, because “when data is officially reported to us, it has a different meaning.”²¹ Directorate-General Energy has, according to the EEA staff, had a different reporting approach, sometimes using consultants in order to harvest existing data.²² EEA staff described the institutional integration currently underway in the context of the Energy Union as “intensely political,” with “some terrible fights between ministries at national level.”²³ For example,

One of the kind of observed discussions was around what would happen to the Climate Change Committee in the Energy Union governance proposal where the Commission proposal was to have one committee at EU level, and the countries wanted to pick it apart again and say, no [...] we need an energy committee and we need a climate change committee.²⁴

¹⁷ EEA staff member 3, interview 11 January 2019.

¹⁸ EEA staff member 1, interview 11 December 2018.

¹⁹ EEA staff member 2, interview 17 December 2018.

²⁰ EEA staff member 2, interview 17 December 2018.

²¹ EEA staff member 1, interview 11 December 2018.

²² EEA staff member 1, interview 11 December 2018.

²³ EEA staff member 1, interview 11 December 2018.

²⁴ EEA staff member 1, interview 11 December 2018.

Asked about the potential integration of the Policies and Measures Database with other databases, such as MURE on energy efficiency,²⁵ one EEA staff member highlighted the “big institutional challenge in integrating these different reporting streams.”²⁶ In sum, institutional integration of policy monitoring across policy sectors and databases can generate significant obstacles, at least in the short term.

There are also key questions around who actively governs policy monitoring in the future (see Schoenefeld and Jordan 2017). One EEA staff member stressed the important role of public institutions in policy monitoring:

I think that in the next 5 years [...] this whole monitoring and data part of our work will go through a very rapid evolution, and I think if public institutions are not at the forefront of that for public purpose, it will be private institutions who take over part of the job based on free data, and they will then start to sell the information to others [...] also in the public sphere.²⁷

Taken together, our interviews illuminate the relevance of institutional settings, implementation, and quality for policy monitoring. These factors also affect the use of policy monitoring data in evaluations, where quality is highly multidimensional (Widmer 2012, p. 263). The significance of the factors is especially clear when institutions are modified through changes in legislation, such as in the case of the Energy Union.

5 Conclusions and Future Directions for Research and Policy

This paper started from the premise that while policy monitoring is often viewed as a key ingredient to achieve greater policy success, its existence and functioning should not be taken for granted. Our empirical results reveal that while climate policy monitoring is routinely advocated across the EU, it is performed in a highly differentiated manner. The fact that the European Commission does not provide clear technical definitions of what a climate “policy” and a “measure” are, likely contributes to some of this heterogeneity. The EU member states typically draw on different models, generating dissimilar quantitative ex ante estimates of policy effects. A part of the observed trends emerges from path dependency in policy monitoring practice, as well as from the relatively decentralised approach to implementation (i.e., data collection at the national level). These institutional variations in turn lead to significant differences in policy monitoring quality. The quality differences show up in the data reported to the EEA, which in turn struggles to provide

²⁵ The MURE (Mesures d’Utilisation Rationnelle de l’Energie) database provides information on energy-efficiency policies and measures that have been carried out in the member states of the EU. <http://www.mesures-odyssee-mure.eu/>. It has been developed as a separate project that is coordinated by ADEME and co-funded by EU research and development funding, but it is not officially linked to the EU Commission’s activities.

²⁶ EEA staff member 3, interview 11 January 2019.

²⁷ EEA staff member 2, interview 17 December 2018.

a cumulative climate policy database that would allow reliable comparisons across member states.

Actors are nonetheless learning how to monitor climate policy over time. But any improvements have emerged slowly, as climate policy monitoring is only to a minor degree a question of making the member states report in a standardised form. The main challenges emerge in the underlying process of shaping institutions and implementation, which, to be standardised, would require the standardisation of concepts, detailed methods and, ultimately, even standardised policy making and politics. Achieving such a high level of standardisation is highly unlikely (and potentially normatively undesirable) because, as Weiss famously argued, policy monitoring and evaluation activities typically “emerge [...] from the rough and tumble of political support, opposition, and bargaining” (Weiss 1993, p. 94–95). A more realistic objective may be to explore what modifications are needed to strengthen the nascent learning effects that are appearing across countries and in specific sectors.

Our initial exploratory analyses have clearly revealed the challenges of using available climate policy monitoring data. The current quantitative *ex ante* data cannot be used on their own to evaluate climate policies across the EU, and causal effects are therefore difficult to examine. For instance, summing quantitative predictions of policy impact and putting these in relation with the individual, national-level greenhouse gas reduction targets proved a futile exercise. Countries frequently use different models and different counterfactuals when producing projections for individual climate policies, generating serious risks of double-counting and inconsistent findings. We could not usefully analyse the quantified, policy-based projections across countries and sectors with regular statistical modelling, which raises concerns about the usefulness of quantitative, per-policy projected emission reductions for European-level climate policy making. The actual *ex post* assessment of policy effort could not even be attempted because the data are so scarce that meaningful statistical analysis proves impossible.

Policy monitoring is every bit as political as other aspects of governing. The politics of policy monitoring encompass the institutional settings, implementation dynamics, and ongoing debates about the (varying levels of) quality. The existence of monitoring is therefore not a binary affair (i.e., present or absent), and it is certainly not self-implementing. The impact of the institutional arrangements, the implementation, and the quality of policy monitoring are political because they affect the potential uses of policy monitoring data and, in consequence, the outcomes of policy making.

A potentially useful venue for future research would be to consider ways of exploring the data produced by the mandatory policy monitoring, for example by analysing the sectors where policy instruments are applied (such as agriculture, energy, and transport) and the type of instruments that have been used (e.g., were they regulatory, economic, or information-based?). Doing so would usefully extend the work begun by Hildén et al. (2014), who have unpacked which types of instruments are projected to generate the greatest emission reductions (based on the 2011 reporting data). Additionally, an exploration of the precise indicator mix used for policy

monitoring could shed new light on the EU's conceptualisation of climate policy.²⁸ It would also be useful to explore the link between reported ex post emissions and the ex ante projections. For example, are countries that overemitted in the past more likely to overproject in the future? The ex post outputs from the mandatory policy monitoring will have to improve significantly before such analyses can be conducted. Finally, it would be fruitful to explore potential determinants of policy monitoring institutions, implementation, and quality (as well as potentially their interactions in statistical terms) in other policy areas in order to test the generalisability of our findings.

Our results have a number of *policy implications*. The policy monitoring data we analysed are highly heterogeneous and therefore useful only to a very limited extent for broader analyses that seek to evaluate policy impact (especially because they are by and large ex ante). This state of affairs bears relevance for the emerging Energy Union and the Paris Agreement, both of which contain considerable policy monitoring provisions. Current EU policy-based data systems say little about the EU-wide climate policy effectiveness beyond emissions trading. The European Commission should in the first instance engage in a co-creation process with the member states to define and operationalise what “climate policies and measures” actually are in order to improve their evaluability (see also Dahler-Larsen and Sundby 2019). Furthermore, the learning effects regarding the implementation of policy monitoring that we detected provide some cause for optimism. If such improvements continue, perhaps in the future countries may be able to produce more consistent data. Supporting them should thus be a key priority. It may be necessary to provide further guidance or even stricter legally binding requirements on climate policy monitoring, as the current à la carte approach to quantified estimates of policy impact (see Schoenefeld and Jordan 2017) appears to have largely failed to produce sufficiently consistent data. The current usability of the policy monitoring data (e.g., for evaluation) thus remains limited.

The ambition of climate policy monitoring in the EU is to deliver data that can be used to judge the policies deployed in the member states and to evaluate their effectiveness. In practice, the limited current potential uses of these data include the following:

1. The ability to provide an overview of the kind of climate policies that member states employ;
2. The ability to trace policy effectiveness over time *within* individual member states based on the quantitative information;
3. Indicative insights on policy development for the Commission.

Beyond such uses, the data cannot, without substantial additional research effort, be mobilised for comparative policy evaluations of specific policies across member states in order to arrive at conclusions on the overall effectiveness of EU climate policies. With this in mind, evidence on the de facto utilisation of the policies and measures data in EU-level policy development remains patchy at best. Policy monitoring use is a significant area of future research because it may also shed

²⁸ We owe this idea to one of our reviewers.

further light on the interaction between policy monitoring and development at the member state and EU levels. In sum, the EU is still struggling to develop its own climate policy monitoring system and resolve associated political challenges. Other countries with less experience and lower capabilities may therefore face even greater challenges in the context of global climate governance.

Climate monitoring in the EU has so far mostly focused on the relatively stringent reporting of national greenhouse gas emissions according to IPCC guidelines (Eggleston et al. 2006). By contrast, the climate policy monitoring approach at the centre of this analysis appears a second-order priority compared to assessing national-level greenhouse gas emissions (low priority has also been demonstrated in other areas of monitoring and evaluating environmental policies; see Potluka 2019). One way to address this may be to strengthen the link between the national greenhouse gas inventories with the reporting on climate policies and measures. Doing so may also help address the glaring imbalance between *ex ante* and *ex post* reporting on policies (see Schoenefeld et al. 2018). In this way, policy-based data may become a helpful ingredient in explanations of changes in national-level greenhouse gas emissions and thus contribute to growing efforts at decomposition analyses, which seek to explain patterns of national-level emissions (e.g., Kisielewicz et al. 2016). Our results point to the fact that investing more resources into producing comparable results and incentivising learning among the member states is crucial for the future development of new policies in Europe and, ultimately, the mitigation of climate change.

After more than 25 years of development, climate policy monitoring in the EU remains a work in progress. Similar analyses of other policy areas (Radaelli 2003; Tholoniati 2010) have stimulated demands for policy monitoring to be combined with sanctioning mechanisms, such as in the European Semester (Verdun and Zeitlin 2018) and in the emerging Energy Union (Energy Systems of the Future 2019; [EU]2018/1999). There are, however, no guarantees that raising the political stakes on policy monitoring with sanctioning will quickly and easily address the fundamental institutional, implementation, and quality issues that we have identified. Doing so could even be counterproductive if it discourages candid reporting.

Acknowledgements We thank the special issue editors as well as the participants at the 2018 Evaluation workshop in Brussels for helpful feedback. The participants of the 2017 ECPR General Conference and the 2017 UACES Annual Conference made useful comments on an earlier version. J.J. Schoenefeld thanks the Technische Universität Darmstadt and the Tyndall Centre for Climate Change Research at the University of East Anglia for institutional support. K. Schulze acknowledges institutional support from the Technische Universität Darmstadt and the Institute for Housing and the Environment (IWU). We thank the EEA staff members for generously volunteering their time and Johannes Koch for his valuable research assistance. Two anonymous peer reviewers provided constructive comments, which strengthened the manuscript.

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.

Funding The first ideas for this work emerged from debates in the context of COST Action INOGO (IS1309), where J.J. Schoenefeld, M. Hildén and A.J. Jordan participated. J.J. Schoenefeld acknowledges the German Federal Ministry of Education and Research (Reference: 03SFK4P0, Consortium ENavi, Kopernikus). K. Schulze thanks the Institut Wohnen und Umwelt (IWU) for financial support. M. Hildén

acknowledges the support of the Strategic Research Council at the Academy of Finland (Grants 314325 and 314350). A.J. Jordan's contribution was also supported by the ESRC CAST—Centre for Climate Change and Social Transformations (ES/S012257/1).

References

- AEA, ECOFYS, Fraunhofer, and ICCS. 2009. *Quantification of the effects on greenhouse gas emissions of policies and measures: Final Report*. No. ENV.C.1/SER/2007/0019. Brussels: European Commission.
- Aldy, Joseph E. 2014. The crucial role of policy surveillance in international climate policy. *Climatic Change* 126(3–4):279–292.
- Aldy, Joseph E. 2018. Policy surveillance: Its role in monitoring, reporting, evaluating and learning. In *Governing climate change: polycentricity in action?*, ed. Andrew Jordan, Dave Huitema, Harro van Asselt, and Johanna Forster, 210–227. Cambridge: Cambridge University Press.
- Armingeon, Klaus, Virginia Wenger, Fiona Wiedemeier, Christian Isler, Laura Knöpfel, David Weisstanner, and Sarah Engler. 2018. *Comparative political data set 1960–2016*. Berne: University of Berne.
- Bodansky, Daniel. 1993. United Nations framework convention on climate change: a commentary. *Yale Journal of International Law* 18(2):451–558.
- Dahler-Larsen, Peter. 2019. *Quality: from plato to performance*. London: Palgrave MacMillan.
- Dahler-Larsen, Peter, and Anna Sundby. 2019. 'If it ain't evaluated, don't fix it!' The politics of evaluability in occupational health and safety. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00190-w>.
- Dunn, William N. 2018. Monitoring observed policy outcomes. In *Public policy analysis*, 6th edn., ed. William N. Dunn, 250–319. Oxon: Routledge.
- Eggleston, Simon, Lenadro Buendia, Kyoko Miwa, Todd Ngara, and Kiyoto Tanabe (eds.). 2006. *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Hayama, Kanagawa: IPCC, National Greenhouse Gas Inventories Programme. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol1.html>.
- Energy Systems of the Future. 2019. *Governance for the European Energy Union: Options for coordinating EU climate and energy policy up to 2030*. München, Halle, Mainz: Deutsche Akademie der Technikwissenschaften e. V.; German National Academy of Sciences; Union of the German Academies of Sciences and Humanities e. V.
- ETC, and ACC. 2012. *Assessment of the member states' policies and measures submitted under the EU Monitoring Mechanism in 2011*. (No. ETC/ACM Technical Paper 2011/19). Bilthoven: European Topic Centre on Air Pollution and Climate Change Mitigation. http://acm.eionet.europa.eu/reports/docs/ETCACM_TP_2011_19_assessment_GHG_PAM.pdf.
- European Environment Agency. 2016. *Environment and climate policy evaluation*. Copenhagen: European Environment Agency.
- Farmer, A. M. 2012. Monitoring and limiting greenhouse gases. https://ieep.eu/uploads/articles/attachments/3c5e75f4-7292-42a7-b8f4-26473c5a9f3d/3.4_Monitoring_and_limiting_greenhouse_gases_-_final.pdf?v=63664509872.
- Haigh, Nigel. 1996. Climate change policies and politics in the European Community. In *Politics of climate change: A European perspective*, ed. Timothy O'Riordan, Jill Jäger, 155–185. New York: Routledge.
- Henisz, Witold J. 2000. The institutional environment for economic growth. *Economics & Politics* 12(1):1–31.
- Hildén, Mikael, Andrew J. Jordan, and Tim Rayner. 2014. Climate policy innovation: developing an evaluation perspective. *Environmental Politics* 23(5):884–905.
- Hyvarinen, Joy. 1999. The European Community's Monitoring Mechanism for CO₂ and other greenhouse gases: the Kyoto Protocol and other recent developments. *Review of European Community & International Environmental Law* 8(2):191–197.
- Jahn, Detlef, Nils Düpont, Thomas Behm, Christoph Oberst, and Martin Rachuj. 2017. *PIP—parties, institutions & preferences: ASPM replication [Version 2017-02]*. Greifswald: University of Greifswald.
- Jensen, Christian B. 2007. Implementing Europe: a question of oversight. *European Union Politics* 8(4):451–477.
- Jordan, Andrew J., and Jale Tosun. 2013. Policy implementation. In *Environmental policy in the EU: Actors, institutions and processes*, 3rd edn., ed. Andrew J. Jordan, Camilla Adelle, 247–266. London & New York: Earthscan from Routledge.

- Kisielewicz, Jerome, Martine Sobey, Yann Verstraeten, Angelica Marino, Lucia Lavric, Sebastian Voigt, and Jurga Tallat-Kelpsaite. 2016. Decomposition analysis of the changes in GHG emissions in the EU and Member States. https://ec.europa.eu/clima/sites/clima/files/strategies/progress/docs/dca_report_en.pdf. Last access: 18 November 2019.
- Knill, Christoph, and Duncan Liefferink. 2007. Implementation effectiveness of EU environmental policy. In *Environmental politics in the European Union: policy-making, implementation and patterns of multi-level governance*, 145–161. Manchester: Manchester University Press.
- Knodt, Michèle. 2019. Multilevel coordination in EU energy policy: a new type of “harder” soft governance? In *Configurations, dynamics and mechanisms of multilevel governance*, ed. Nathalie Behnke, Jörg Broschek, and Jared Sonnicksen, 173–191. Cham: Palgrave Macmillan/Springer Nature.
- Lindner, Johannes. 2003. Institutional stability and change: two sides of the same coin. *Journal of European Public Policy* 10(6):912–935.
- March, James G., and Johan P. Olsen. 2009. Elaborating the “New Institutionalism”. In *The Oxford handbook of political institutions*, ed. Sarah A. Binder, Roderick A. W. Rhodes, Bert A. Rockman, James G. March, and Johan P. Olsen, 3–20. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199548460.001.0001/oxfordhb-9780199548460-e-1>.
- Niederberger, Anne Arquit, and Melinda Kimble. 2011. MRV under the UN climate regime: paper tiger or catalyst for continual improvement? *Greenhouse Gas Measurement and Management* 1(1):47–54.
- OECD-DAC. 2002. *Glossary of key terms in evaluation and results based management*. Paris: DAC Network on Development Evaluation, OECD.
- Öko-Institut, Cambridge Economics, AMEC, Harmelink Consulting, and TNO. 2012. *Ex-post quantification of the effects and costs of policies and measures*. No. CLIMA.A.3/SER/2010/0005. Berlin: Öko-Institut.
- 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>.
- Radaelli, Claudio M. 2003. *The Open Method of Coordination: a new governance architecture for the European Union?* Stockholm: Swedish Institute for European Policy Studies.
- Rossi, Peter H., Mark W. Lipsey, and Gary T. Henry. 2018. *Evaluation: a systematic approach*, 8th edn., London: SAGE.
- 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 Timothy Rayner. 2019. Monitoring energy policy. In *Handbook of energy governance in Europe*, ed. Michèle Knodt, Jörg Kemmerzell. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-73526-9_43-1.
- Schoenefeld, Jonas J., Mikael Hildén, and Andrew J. Jordan. 2018. The challenges of monitoring national climate policy: Learning lessons from the EU. *Climate Policy* 18(1):118–128.
- Schwartz, Robert, and John Mayne. 2005. Assuring the quality of evaluative information: theory and practice. *Evaluation and Program Planning* 28(1):1–14.
- Scriven, Michael. 1981. *Evaluation thesaurus*. Newbury Park: SAGE.
- Stake, Robert E., and Thomas A. Schwandt. 2006. On discerning quality in evaluation. In *The SAGE handbook of evaluation*, ed. Ian Shaw, Jennifer C. Greene, and Melvin M. Mark, 404–418. Thousand Oaks: SAGE.
- Stephenson, Paul, Jonas J. Schoenefeld, and Frans Leeuw. 2019. The politicization of evaluation: constructing and contesting EU policy performance. *Politische Vierteljahresschrift*. <https://doi.org/10.1007/s11615-019-00212-7>.
- Tholoniati, Luc. 2010. The career of the Open Method of Coordination: lessons from a ‘soft’ EU instrument. *West European Politics* 33(1):93–117.
- Thomann, Eva, and Fritz Sager. 2017. Moving beyond legal compliance: innovative approaches to EU multilevel implementation. *Journal of European Public Policy* 24(9):1253–1268.
- Tosun, Jale. 2012. Environmental monitoring and enforcement in Europe: a review of empirical research. *Environmental Policy and Governance* 22(6):437–448.
- Treib, Oliver. 2014. Implementing and complying with EU governance outputs. *Living Reviews in European Governance* 9(1). <https://doi.org/10.12942/lreg-2014-1>.
- Vedung, Evert. 1997. *Public policy and program evaluation*. New Brunswick: Transaction Publishers.
- Verdun, Amy, and Jonathan Zeitlin. 2018. Introduction: The European Semester as a new architecture of EU socioeconomic governance in theory and practice. *Journal of European Public Policy* 25(2):137–148.

- van Voorst, Stijn, and Ellen Mastebroek. 2019. Evaluations as a decent knowledge base? Describing and explaining the quality of the European Commission's ex-post legislative evaluations. *Policy Sciences*. <https://doi.org/10.1007/s11077-019-09358-y>.
- Weiss, Carol H. 1993. Where politics and evaluation research meet. *Evaluation Practice* 14(1):93–106.
- Wettestad, Jørgen. 2007. Monitoring and verification. In *The Oxford Handbook of International Environmental Law*, ed. Daniel Bodansky, Jutta Brunnée, and Ellen Hey, 974–994. Oxford: Oxford University Press.
- Widmer, Thomas. 2004. The development and status of evaluation standards in Western Europe. *New Directions for Evaluation* 2004(104):31–42.
- Widmer, Thomas. 2012. Evaluation quality in the context of sustainability. In *Sustainable development, evaluation and policy-making*, ed. Anneke von Raggamby, Frieder Rubik, 261–278. Cheltenham: Edward Elgar.
- Williams, Bob. 2005. Quality. In *Encyclopedia of evaluation*, ed. Sandra Mathison, 350. Thousand Oaks: SAGE.