Environmental Policy and Governance
Env. Pol. Gov. (2017)
Published online in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/eet.1761

Reconsidering EU Compliance: Implementation performance in the field of environmental policy

Elena Bondarouk* and Ellen Mastenbroek
Radboud University, Institute for Management Research, the Netherlands

ABSTRACT

European Union (EU) environmental policy can only work in practice when it is implemented by and within the member states. Yet, despite its importance, we still lack a solid and cumulative understanding of the practical implementation of EU environmental policies, mainly because of the dominance of case-specific empirical insights and the dichotomous conceptualization of compliant implementation. This paper proposes a conceptual framework for analysing implementation performance, which is built around three dimensions: substance, scope and effort. The framework's relevance and analytical quality are substantiated by a systematic review of empirical studies on practical implementation of 18 EU environmental directives. We find evidence of three types of knowledge deficits: there is neglect of the 'scope' and 'effort' dimensions of implementation; disproportionate attention to the Water Framework Directive, and the Northern and Western European member states. The proposed conceptual framework aims to inform future research on EU environmental implementation. © 2017 The Authors. Environmental Policy and Governance published by ERP Environment and John Wiley & Sons Ltd

Received 27 April 2016; revised 20 March 2017; accepted 23 March 2017

Keywords: environmental policy; EU implementation; policy implementation; practical implementation; systematic review

Introduction

A KEY ISSUE IN ENVIRONMENTAL GOVERNANCE LIES IN ENSURING THAT AMBITIOUS POLICIES IN BOOKS ARE TRANSLATED INTO POLICIES IN action (Leventon and Antypas, 2012). This is no different for the European Union (EU), which is an active producer of environmental policies. The EU's central environmental goals of greening the EU economy, protecting nature and safeguarding health and quality of life across the EU (Knill and Lenschow, 2000) are clearly undermined if EU policies are not complied with by the member states.

Consequently, a number of scholars of EU environmental governance have studied compliance with EU environmental policy, i.e. the national implementation of EU environmental policy (e.g. Bennett, 1993; Knill and Lenschow, 1998, 2000; Börzel, 2000; Börzel and Risse, 2000; Haverland, 2003; Bugdahn, 2005; Laffan and O'Mahony, 2008; Jans *et al*., 2009; Liefferink *et al*., 2011; Morris, 2011). This area of research has closely examined to what extent and how the EU's member states have implemented the policy requirements laid down in various directives.

*Correspondence to: E. Bondarouk, Radboud University, Institute for Management Research, The Netherlands, Thomas van Aquinostraat 5.0.69, Postbus 9108, Nijmegen 6500, The Netherlands. E-mail: e.bondarouk@fm.ru.nl

However, the literature on EU environmental policies has not produced a complete picture of the state of EU environmental implementation (Tosun, 2012). The reasons for this lack of overall insight are 2-fold. First, most existing studies are highly case-specific, deriving their operationalization of implementation conformity from the provisions of particular directives (Treib, 2014). This approach leads to the drawing of idiosyncratic conclusions, which do not reach a broader audience and hinder the accumulation of knowledge and the drawing of general conclusions (Töller, 2010; Tosun, 2012; Engeli and Allison, 2014; Treib, 2014; Schaffrin *et al.*, 2015: 257).

Second, environmental implementation studies often evaluate implementation in dichotomous fashion – compliant or non-compliant implementation (e.g. Börzel, 2000; Bauer *et al.*, 2007; Morris, 2011; see Treib, 2014 for more examples). This conceptualization of responses to EU policies is of restricted use in the field of EU environmental policy, given the growing use of procedural provisions and open norms (Scott, 2000; Börzel, 2003; Knill and Lenschow, 2004; Hartlapp and Falkner, 2009; Liefferink *et al.*, 2011). The mere fact that a member state or local implementer lives up to a procedural obligation does not have any bearing on *the extent to which* the implementer really makes 'EU policy work' (Haverland and Romeijn, 2007). Accordingly, variation in implementation practices is not systematically assessed.

As a result, we are still in the dark about the extent to which and ways in which member states take EU environmental policy seriously (see Voermans, 2015). Answering this question requires a fully fledged conceptualization of the various aspects of implementation practices in response to EU legislation (Lange, 1999; Winter, 2006, 2012; Hupe, 2014; Thomann, 2015). To this end, the first goal of this paper is to develop and test a general 'systematized concept' (Adcock and Collier, 2001: 532) of implementation performance (Hill and Hupe, 2003: 475; Winter, 2006: 159; Hupe, 2011: 66) that goes beyond the details of specific directives, and captures variation in implementation.

This article uses a two-step deductive approach to construct the conceptualization. The first step is to derive aspects of implementation performance from the literatures on national policy implementation, analysis, design, evaluation and change. The second step is to compare the conceptual framework thus obtained with the insights from the literature on the practical implementation of EU environmental policies. We conduct a systematic literature review of case studies on practical implementation in the field of environmental policy published in the years 2010–2014, covering 18 different environmental directives. We check whether the framework is complete, and whether the categories are mutually exclusive. This second step also facilitates a second goal of this article: the isolation of knowledge gaps in the current research on practical implementation of environmental policies.

Finally, this study conveys lessons for the broader field of EU compliance (Treib, 2014). While the literature has repeatedly called for more research on the practical implementation of EU directives (e.g. Mastenbroek, 2005; Versluis, 2007; Treib, 2014), it has found itself in need of a useful concept for doing so systematically (Tosun, 2012; Treib, 2014). By developing and corroborating such a concept, this paper may step up compliance research in other policy sectors. We aim to pave the way for more systematic assessment of EU implementation performance in the field of environmental policy, thus facilitating knowledge accumulation (Tosun, 2012; Saetren, 2014; Sager *et al.*, 2014; Thomann, 2015). Ultimately, the paper is expected to allow for fuller diagnosis and understanding of the implementation deficit that has been claimed to haunt the EU (Hupe, 2014: 170; Treib, 2014; Voermans, 2015: 365).

Setting the Scene: Defining Implementation Performance

Before systematically conceptualizing implementation performance, we need to define the concept's core components: implementation and performance. Starting with the former, *implementation* refers to the stage between the transposition of EU directives and the enforcement of these directives by European or national actors. To differentiate it from transposition or legal transposition, this stage is also called practical implementation (Mastenbroek, 2005; Treib, 2014).

We argue that practical implementation consists of two distinct substages, the first of which we refer to as *final policy formation* on the ground. Hupe and Hill (2016: 106–107) argue that even at the stage of practical implementation there is a lot of decision making yet to be done. Any attempt to implement a policy always brings

DOI: 10.1002/eet

new issues on the agenda and thus implementation and decision-making overlap (Lindblom and Woodhouse, 1993: 11; Hupe and Hill, 2016: 106). Even though generally overlooked in the EU implementation literature, this stage should be regarded as an important step in practical implementation, as it involves the local authority's efforts to operationalize national policy for practical purposes (Hill and Hupe, 2003: 479; Spicker, 2006: 44; Winter, 2006: 159; Versluis, 2007; Leventon and Antypas, 2012: 256; Steunenberg and Dimitrova, 2014: 7). During this substage, complex national policy – or transposed EU legislation, for that matter – is broken down into a series of tangible implementation tasks (Spicker, 2006: 43; Winter, 2006: 159). Furthermore, the responsible authority is assigned, a timescale within which the task is to be completed is established, and evaluation parameters are set (Spicker, 2006: 43; Winter, 2006: 159). If policy formation is completed, *policy delivery* (Versluis, 2007; Winter, 2012; Treib, 2014) commences. This second implementation stage is characterized by the actual putting in practice of the policy instruments. In environmental policy, typical examples are the enactment of physical measures or the provision of permits.

Having defined *implementation*, we move on to the definition of *performance*. The literature on policy analysis and implementation distinguishes two interpretations of performance. One focuses on policy *outputs*, e.g. the actions taken in response to law (Vedung, 1997; Hill and Hupe, 2003: 475; Winter, 2006: 159; Howlett *et al.*, 2009: 183; Hupe, 2011: 66). The second interpretation focuses on policy *outcomes*, impacts or effects, i.e. the question of whether a policy indeed resolved the problem it set out to solve (Barrett and Fudge, 1981; Mastop and Faludi, 1997; Vedung, 1997; Berke *et al.*, 2006; Winter, 2006: 159; Tosun, 2012: 440).

This study understands performance in the first way, i.e. in terms of *outputs*. It does so for a methodological reason, as an evaluation of EU policy impact is extremely challenging, because an isolation of the EU effect is practically impossible (Haverland, 2006; Tosun, 2012; Bauer and Knill, 2014). Any assessment of policy outputs presupposes a firm understanding of *policy instruments*: the techniques by which authorities attempt to change or maintain the policy status quo (Howlett *et al.*, 2009; May, 2003: 225; Schaffrin *et al.*, 2015). Being the building blocks of any policy (May, 2003: 225), policy instruments are the core of any policy output (Bauer and Knill, 2014; Schaffrin *et al.*, 2015: 260).

We propose to compare implementation performance both in a vertical and in a horizontal manner. Where the EU standards are provided a vertical comparison is necessary, i.e. a comparison of implementation performance with the objectives set out in the superior EU law. Horizontal comparison, in turn, concerns the differences in implementation performance among implementers at the same administrative layer, e.g. municipalities, using the same policy instruments (Winter, 2006; Hupe, 2011).

This horizontal evaluation is especially relevant in cases when EU legislation leaves discretion to implementers or in case of procedural provisions. The discretion may result in great variance in implementation, which is still within the boundaries left by EU directives (Hartlapp and Falkner, 2009; Sager *et al.*, 2014; Treib, 2014; Thomann, 2015). EU environmental law increasingly contains procedural provisions (Scott, 2000; Héritier, 2002; Börzel, 2003; Knill and Lenschow, 2004; Liefferink *et al.*, 2011). These are designed to indirectly affect the desired policy outcome through the manipulation of policy processes (Huber and Shipan, 2002; Howlett *et al.*, 2009). For example, implementers may have to develop an action plan to tackle an environmental problem. A mere vertical evaluation of implementation in this case would reduce a great variance in policy responses to a dichotomous notion of compliant versus non-compliant implementation. A horizontal evaluation in this case would allow for more benchmarking and thus also better insight into implementation practices.

These horizontal and vertical assessments of policy performance can be made on two dimensions: density and intensity (Knill *et al.*, 2012; Bauer and Knill, 2014; Schaffrin *et al.*, 2015). Policy *density*, firstly, refers merely to the number of policy instruments put in place to reach the policy objectives, i.e. the breadth and differentiation of legislative activity (Bauer and Knill, 2014: 33). Policy *intensity*, secondly, concerns the content of the policy instruments (Knill *et al.*, 2012; Schaffrin *et al.*, 2015), i.e. the breadth and differentiation of policy responses (Bauer and Knill, 2014: 33). To study to what extent local implementers make EU policy work, we specifically focus on *intensity*, as this sheds light on the policy commitment of local implementers (cf. Schaffrin *et al.*, 2015: 261).

In summary, this article defines implementation performance as the intensity of policy outputs undertaken by implementers in response to EU policy instruments – relative to the directive's objectives (vertical aspect) and to other implementers' outputs (horizontal comparison). Given the importance of open norms and procedural requirements in EU environmental law (Knill and Lenschow, 1998, 2000; Scott, 2000; Newig and Fritsch, 2009;

DOI: 10.1002/eet

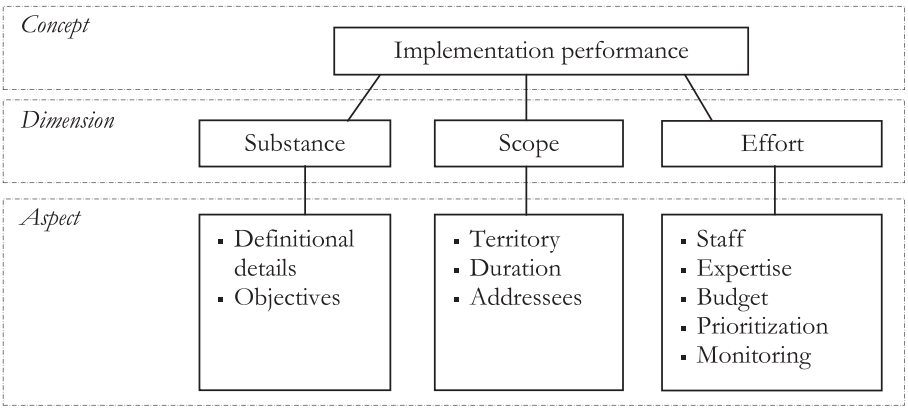

Figure 1. Conceptual framework

Newig and Koontz, 2014), the concept does not only have a vertical focus, i.e. aimed at comparing implementation with EU rules, but also a horizontal focus, i.e. aimed at comparing implementation practices between various implementing actors – either member states or units within these member states. The next section develops and specifies this concept by proposing three different analytical dimensions, together comprising 10 specific aspects.

Conceptual Framework

To capture the concept of policy performance in more depth and detail, this paper employs the approach of Adcock and Collier (2001). Accordingly, the concept of interest is broken down into different dimensions, to specify it in as detailed a fashion as possible.[1] In doing so, we use and integrate the existing literature on national policy implementation, policy analysis, policy design, policy change and policy evaluation (e.g. Baldwin and Cave, 1999; Huber and Shipan, 2002; Spicker, 2006; Winter, 2006, 2012; Howlett *et al.*, 2009; Hupe, 2011, 2014; Tosun, 2012; Bauer and Knill, 2014; Howlett and Cashore, 2014; Schaffrin *et al.*, 2015).

Synthesizing this literature, three main dimensions of implementation performance emerge: substance, scope and effort. As illustrated in Figure 1, we propose to divide these three dimensions each into more refined aspects. The scores for these aspects are to be aggregated to form one score per dimension, which can be aggregated to one overall score. In the case of horizontal comparison, scoring is the result of comparison of the different implementers under study, e.g. regions, agencies or municipalities. In this case the scores on implementation performance are established relatively (Hupe, 2014: 173). When evaluating implementation performance, it is best to present both scores: the scores on different dimensions and the total score on implementation performance.

Dimension 1: Substance

Substance, the first dimension of implementation performance, relates to the essence of what central issue is being regulated by the policy instrument (cf. Steunenberg, 2007; König and Mäder, 2013; Zhelyazkova, 2013). The literature suggests two aspects of this dimension: definitional details and objectives of the policy instrument.

The first aspect of substance relates to the *definitional details* used during practical implementation (cf. Huber and Shipan, 2002: 51; Spicker, 2006: 43). The directive and its transposing measures may contain ambiguous elements that will be subjected to interpretation at practical implementation (Beijen, 2011: 152). When vertical evaluation of implementation performance is impossible, implementation performance will depend on how restrictively or comprehensively these elements are defined during practical implementation relative to other implementing peers (Howlett *et al.*, 2009). The more specifically the definitional details are provided in the local statutes, the more

[1]The framework does not specify the indicators, as this would bring the policy context into the framework and preclude it from travelling across policy sectors and cases.

DOI: 10.1002/eet

specific and consistent implementation will be (cf. Baldwin and Cave, 1999: 43; Huber and Shipan, 2002: 50–51; Hupe, 2011: 69; Bauer and Knill, 2014: 33; Schaffrin *et al.*, 2015: 264). There are three elements of definitional details that have to be defined in the practical implementation.

First, most directives contain a list of definitions in one of their first articles (Beijen, 2011). Still, these definitions may be open to interpretation, which may give rise to differences in implementation performance on the ground. The definitions of 'waste', 'best available techniques' and 'discharge', for instance, have given rise to much case law and literature (Beijen, 2011: 152). The absence of clear definitions becomes particularly relevant if the literal wording of a directive finds its way into national legislation, because local implementers will have to refine these concepts themselves to make them operational for practical implementation (Scott, 2000: 45–46; Spicker, 2006: 43; Beijen, 2011: 152). Second, a directive often contains a provision that an appropriate body should be appointed to implement the policy. The practical implementers are often tasked to define the responsibilities of persons or bodies who are engaged in the delivery of policy outputs (cf. Scott, 2000: 45–46; Huber and Shipan, 2002: 50–51; Schaffrin *et al.*, 2015: 264). Lastly, definitional details relate to the exceptions when some tasks or actions should not be taken or are exempted from regulation (cf. Scott, 2000: 48). The more of such exemptions there are, the less comprehensive the practical implementation will be.

The second aspect of substance concerns the *objectives* of the policy instruments as adopted by the practical implementer. These objectives are 'specific on-the-ground policy requirements' (Knill *et al.*, 2012; Leventon and Antypas, 2012: 256; Bauer and Knill, 2014: 33; Howlett and Cashore, 2014: 21). If objectives are set, the purpose of the policy is made transparent to the public and other political actors, and thus the implementers can be held responsible for the achievement of these goals (Schaffrin *et al.*, 2015: 263). A directive contains two types of policy instruments, i.e. substantive and procedural, that have to be operationalized into practical objectives on the ground.

Substantive policy instruments include norms, standards or target values (Beijen, 2011: 154). Practical implementers may impose stricter or more lenient norms or standards in their own jurisdictions (Huber and Shipan, 2002: 50–55; Jans *et al.*, 2009; Beijen, 2011: 154). For example, local governments may aim for higher air quality standards then the EU directive prescribes. Procedural policy instruments concern matters such as public participation, formulation of policy plans/reports, or the designation and protection of areas (Knill and Lenschow, 1998, 2000; Scott, 2000; Beijen, 2011; Newig and Koontz, 2014). For example, practical implementers are likely to differ in how they set the objectives for public participation. Policy performance in this respect may range from mere information provision to full involvement with voting procedures installed (cf. Huber and Shipan, 2002: 58; Howlett *et al.*, 2009: 117–118). In the case of an obligation to formulate plans or reports, practical implementers will again set different objectives. These can vary in what types of measures are included in the plans, or what type of information is included in the reports (cf. Huber and Shipan, 2002: 58). As another example, practical implementers may set the criteria for designation and protection of areas in rather different ways (Beijen, 2011: 155).

Dimension 2: Scope

The second dimension of implementation performance concerns the ***scope*** of implementation (Beijen, 2011: 153; Bauer and Knill, 2014: 33; Thomann, 2015). This refers to the range of the policy: where, when and to whom does the policy task apply? If the scope is ambitious, it signals that an implementer takes the policy seriously (Schaffrin *et al.*, 2015: 263). The literature has suggested three aspects of scope on which practical implementation may differ: territory, duration and addressees (Huber and Shipan, 2002: 49; Jans *et al.*, 2009; Bauer and Knill, 2014: 33; Jenkins-Smith *et al.*, 2014: 189). Depending on whether a directive or national legislation specifies these three aspects of scope, implementation performance should be evaluated either vertically or horizontally.

The first aspect of scope is the *territory* where the policy instrument applies (Jans *et al.*, 2009; Bauer and Knill, 2014). Typically, the practical implementers may choose to make the policy instrument applicable to a whole region or only specific areas in a city. For example, local government may choose to target only a specific area in a city with extra air quality measures, while another will target the whole city. The second aspect of scope is temporal in nature, and concerns the *duration* of the policy task (Huber and Shipan, 2002; Jans *et al.*, 2009; Bauer and Knill, 2014). For example, a specific policy plan may differ in terms of temporal span among the implementing authorities. The implementers may apply the environmental standards earlier or longer than determined in the national or EU policy instruments. The third aspect of scope captures how broad or specific the group of *addressees* targeted by the policy is

DOI: 10.1002/eet

(Jans *et al.*, 2009; Bauer and Knill, 2014; Jenkins-Smith *et al.*, 2014). Practical implementers may reduce or broaden the group targeted by the policy. For example, they may target a broader group of companies to perform an environmental impact assessment than the national or EU legislation prescribes. In the case of procedural policy instruments, this aspect of scope relates to, for instance, who is invited to the public consultation.

Dimension 3: Effort

The third and final dimension of implementation performance concerns the ***effort*** that implementers put into accomplishing a policy's goals (Winter, 2006: 160; Howlett *et al.*, 2009: 186; Bauer and Knill, 2014: 34). Effort refers to 'the factors affecting the probability that substantial requirements are effectively achieved' (Bauer and Knill, 2014: 34). The policy instruments that are characterized by higher intensity have more effort invested in them (Schaffrin *et al.*, 2015: 262). Five aspects of the effort dimension emerge from the literature.

The first three aspects concern the resources that implementers allocate to implementation. The first aspect is the number of *staff*, i.e. organizational resources, designated to support the implementation of policy instruments (Hartlapp, 2009: 475; Tosun, 2012: 442; Bauer and Knill, 2014: 34; Schaffrin *et al.*, 2015: 262). It relates to how many people in the organization are responsible for defining the policy tasks and carrying them out. The second aspect concerns the *types of expertise*, i.e. informational resources, involved to support policy implementation (Radaelli and De Francesco, 2007; Bauer and Knill, 2014: 34; Schaffrin *et al.*, 2015: 262), which relates to the type of knowledge consulted during policy implementation. An example could be whether policy implementers have different backgrounds to facilitate the synergy of expertise and create a sound and feasible policy. The third aspect of effort is the percentage of an implementer's *budget*, i.e. the financial resources, allocated to the implementation of policy goals (Bauer and Knill, 2014: 34; Schaffrin *et al.*, 2015: 262–263). Because these resources are typically not specified in national or EU legislation, the assessment should be horizontal in nature.

The fourth aspect of effort concerns the *prioritization* of goals and measures within one policy (Winter, 2006: 160). Given their limited resources, implementers typically prioritize some goals or measures over others within the same policy. For example, a policy task might contain measures which the implementers are highly unlikely to implement in total. Therefore, one can analyse which measures or goals take precedence and receive most attention. The practical implementers are likely to vary in terms of what policy norms or measures are prioritized. Similarly to the first three aspects, prioritization is not specified by national or EU legislation, which necessitates horizontal comparison.

The final aspect of effort is *monitoring* (May and Winter, 1999; Howlett *et al.*, 2009: 185; Beijen, 2011: 159; Tosun, 2012: 442; Bauer and Knill, 2014: 34; Schaffrin *et al.*, 2015: 264). It describes how the practical implementer will assess the quality of the delivered task as well as the consequences of a failure to act, or how the practical implementers envision controlling for policy adherence (cf. Vedung, 1998: 31; Hartlapp, 2009: 475). Huber and Shipan (2002: 52) refer to it as the 'quality assurance' mechanism, which aims at ensuring policy adherence (see also May and Winter, 1999). The presence of such an enforcement mechanism signals to what extent practical implementers really make a policy work (Howlett *et al.*, 2009; Schaffrin *et al.*, 2015: 264). Depending on whether a directive or national legislation prescribes local monitoring, implementation performance on these aspects should be evaluated vertically or horizontally.

Research Design

Having conceptualized implementation performance, several questions present themselves. First, is the framework complete when compared to existing empirical studies of implementation? Second, are the theoretical aspects proposed mutually exclusive within a dimension (Schreier, 2012: 75)? Hence the collective exhaustiveness and mutual exclusiveness of the framework's dimensions must be corroborated (Schreier, 2012: 75). And third, how complete is our understanding of implementation performance in the field of EU environmental policy?

To answer these questions, a deductive qualitative content analysis (Schreier, 2012) was conducted, comparing the conceptual framework with existing empirical studies of practical implementation of specific environmental

DOI: 10.1002/eet

directives. This analysis reduces the data by classifying the specific and concrete information from the earlier studies under the aspects of our conceptual framework (Potter and Levine-Donnerstein, 1999; Schreier, 2012).

These existing studies, published in journals, served as data for our analysis. The studies were selected using a systematic literature review method to avoid any intentional or unintentional bias in the selection of the data (Petticrew and Roberts, 2006). Web of Science (Thomson), one of the largest scientific databases for the social and environmental sciences, was selected for the collection of data. The data were selected in five steps in July 2014.

First, as this research is interested in implementation and compliance with EU directives, a Boolean search was carried out using the keywords '(EU OR directive) AND (implement* OR compliance)'. Web of Science yielded 7989 documents with these keywords in either the title and/or the abstract of the documents. Second, the analysis was limited to academic articles, because these enjoy peer review which safeguards quality. The sample was restricted to articles written in English and dealing with environmental policy. This selection resulted in 940 articles. Third, the titles and the abstracts were closely read, to include only those articles that deal with the practical implementation of EU directives. Articles focusing on policy outcomes, cost–benefit analysis of policy implementation, technical calculations or anything else except implementation performance were excluded. This selection step yielded 187 articles. To validate this sample, we checked whether the articles on practical implementation of EU environmental directives identified in more general EU literature reviews (Angelova *et al.*, 2012; Treib, 2014; Toshkov *et al.*, n.d.) were also included in our sample of 187 articles. This was indeed the case. Fourth, aiming to grasp current knowledge, articles were selected if published in the previous 5 years, i.e. 2010–2014. This resulted in a database of 112 articles. As a final selection step, we applied the same inclusion and exclusion criteria as in step 3 of the selection process to the full texts of the articles. This led to the final identification of 70 articles[2] to be subjected to the qualitative content analysis.

For this analysis we coded these 70 articles based on the conceptual framework. We used a standard coding procedure (Schreier, 2012). Following this procedure, a codebook was developed, which contained operational definitions of the previously defined 10 aspects of implementation performance. Analysis of the first two articles yielded typical examples for all aspects, which were added to the codebook. The deductive analysis of the articles was performed using NVivo 10 software.

To examine the framework's quality, i.e. the collective exhaustiveness and mutual exclusiveness of the framework's dimensions, the following three steps were taken. First, to corroborate the concept's exhaustiveness, an additional code was created to keep track of any descriptions that would not fit the framework. Such instances were to be coded as 'emerging themes'. Second, to examine the mutual exclusiveness of aspects of one dimension, we checked whether the coding units had overlapping aspect-codes assigned to them (Schreier, 2012). And third, the coding was subjected to consistency control, i.e. reliability, by two additional coders (Kippendorf, 2004: 215; Schreier, 2012: 169). A strong measure of reliability and thus also quality assurance of the coding scheme, i.e. the conceptual framework, is to examine whether others can interpret the data by coding it in the same manner, or at least agree on the interpretation of the data (Kippendorf, 2004: 215; Schreier, 2012: 169).

This intercoder reliability examination was established in two ways. The articles were divided into two roughly equal and mutually exclusive subsets. The first subset ($n = 36$) was divided between the two coders and subjected to independent coding of the result sections of the articles. The second subset ($n = 34$) was also divided between these two coders, but instead of independent coding the coding of the main coder was checked by these two coders. The coders were instructed to keep track of any data that would not fit the framework, and code these as 'emerging themes'. These two different ways of performing intercoder reliability assessment are the most common ways of examining the reliability of coding (Schreier, 2012: 169).

To evaluate the completeness and coverage of our understanding of implementation performance in the field of EU environmental policy, the following three steps were taken. First, by coding 70 articles on the aspects of the conceptual framework we reflected on the state-of-the-art knowledge of implementation performance. Such analysis allowed us to see which aspects have received most empirical attention so far, and to identify any gaps in our understanding of environmental policy implementation. Second, we kept track of what directives were examined to see whether some directives have received more attention and whether there are systematic differences in how

[2]For this list, please see the online Appendix.

DOI: 10.1002/eet

EU directive	Frequency
Water Framework Directive (60/2000)	32
Multiple directives	7
Strategic Environmental Assessment Directive (2001/42/EC)	6
Natura 2000 (Wild Birds Directive (79/409/EEC) and the Habitats Directive (92/43/EEC))	5
Environmental Noise Directive (2002/49/EC), Habitats Directive (92/43/EEC)	3
Drinking Water Directive (98/83/EC), Floods Directive (2007/60/EC)	2
EU Biofuels Directive (2003/30/EC), Dangerous Preparations Directive (1999/45/EC), Environmental Impact Assessment Directive (85/337/EEC), Waste Framework Directive (Directive 2008/98/EC), Integrated Pollution Prevention and Control Directive (2008/1/EC), Landfill Directive (99/31/EC), Nitrates Directive (91/676/EEC), Wastewater Treatment Directive (91/271/EEC), Seveso II Directive (96/82/EC), Sustainable use of pesticides Directive (2009/128/EC)	1

Table 1. EU directives in the sample

implementation performance was evaluated based on the directive at stake. And third, for the same reasons, we also kept track of what countries were studied in these articles.

Analysis

As depicted in Table 1, the sample covers 18 EU environmental directives. In total, 32 articles (46%) focus on the Water Framework Directive (WFD). Seven articles cover multiple directives in their study. The EU Strategic Environmental Assessment Directive is examined by six articles, the Natura 2000 by five articles, and the Environmental Noise Directive and Habitats Directive by three articles each.

As can be seen in Table 2, the sample covers 19 different member states. Seventeen articles (24%) examined two or more countries. The top five most examined countries in the sample are UK, Germany, Ireland, Italy and the Netherlands.

Moving to the quality of the framework, the analysis revealed that the conceptual framework captures implementation performance, as defined by the existing directive-specific studies, rather well. In the first place, none of the coders assigned the 'emerging themes' code, implying that all aspects of implementation performance covered by the existing studies could be subsumed under one of the headings of our conceptual framework. Accordingly, it seems safe to conclude that this forms a complete representation of implementation performance.

Secondly, to assess the mutual exclusiveness of aspects within one dimension, we checked whether different framework aspect codes were assigned to the same coding unit. No such overlapping coding was found. Third, the two subsets of data displayed high intercoder reliability agreement: 75% agreement for the whole framework in the first subset (column A, Table 3), and 95% agreement in the second subset (column B).[3] It thus seems safe to conclude that others interpret the same data in the same way, using this conceptual framework.

The next step in the analysis was to evaluate the state of knowledge on implementation performance in the field of EU environmental policy. Here, the analysis reveals that our knowledge is fragmented in three ways. First, we see variant coverage of the 10 aspects of implementation performance in existing studies. Whereas none of the articles covered all the proposed aspects, all reported at least one aspect of the conceptual framework. In total, 19% of articles used eight of the aspects to examine implementation performance, another 19% used seven aspects, followed by 16% elaborating on five aspects, 13% on six aspects, 11% on four aspects and 9% on nine aspects. More specifically, Table 3 presents the relative frequency scores[4] of the conceptual framework aspects. Column C shows that the 'objectives' aspect received most attention, followed by 'expertise', 'definitional details' and 'monitoring'. The popularity of the 'objectives' aspect is not surprising: it is an intuitive first step in implementation research to

[3]This is not surprising as the first type of intercoder reliability test is a more critical reliability assessment (Schreier, 2012).
[4]As most articles paid attention to different aspects at the same time, the relative frequency scores do not add up to 100%.

DOI: 10.1002/eet

Country	Frequency
Multiple	17
UK	8
Germany	7
Ireland	5
Italy, The Netherlands	4
Spain, Denmark, Greece, Sweden	3
France, Belgium, Slovenia	2
Croatia, Cyprus, Czech Republic, Hungary, Poland, Portugal, Romania	1

Table 2. Countries in the sample

examine which of the prescribed policy instruments have been put in place and applied when studying practical implementation. Much less attention, however, has been paid to the 'scope' and 'effort' dimensions of implementation performance. This is surprising, given their theoretical importance for implementation performance. Only a few articles paid attention to 'staff' and 'prioritization' aspects. In sum, our analysis reveals important knowledge gaps in implementation performance in EU environmental policy.

Second, as already mentioned, a large proportion of the sample (32 articles) focused on the WFD (Table 1). Columns D and E of Table 3 show that our knowledge on WFD implementation performance is much more extensive than that on other directives, especially when it comes to the 'scope' and 'effort' dimensions. Another large proportion of the sample (12 articles) focused on more than one directive, i.e. multiple directives and Natura 2000 directives. Column F shows that many of these articles covered various aspects of the conceptual framework. However, little attention has been paid to the 'monitoring', 'territory', 'staff' and 'prioritization' aspects of implementation performance.

Third, the analysis shows that our insights on implementation performance mostly stem from case studies on Western member states, which are expected to have a relatively smooth practical implementation (Falkner and

Dimension	Aspect	A	B	C	D	E	F	G	H
		Agreement (%)		All (%)	Directive (%)			Worlds of compliance (%)	
		Subset I (n = 36)*	Subset II (n = 34)	Articles (n = 70)	WFD (n = 32)	Non-WFD (n = 38)	>1 (n = 12)	Group I† (n = 30)	Group II‡ (n = 19)
Substance	Definitional details	82	94	71	72	71	58	73	84
	Objectives	76	98	99	97	100	100	100	100
Scope	Territory	85	94	60	75	47	42	70	53
	Duration	71	92	53	66	42	58	60	42
	Addressees	68	97	63	69	58	67	70	53
Effort	Staff	92	98	19	19	18	42	13	21
	Expertise	78	92	74	78	71	67	80	79
	Budget	70	97	50	44	55	67	37	58
	Prioritization	75	100	39	53	29	33	47	37
	Monitoring	80	97	64	63	66	42	60	68

Table 3. Data and intercoder reliability assessment (%)
*Number of articles in the sample.
†World of law observance and world of domestic politics.
‡World of dead letters and world of transposition neglect.
Note: the percentages do not add up to 100% in columns C to H as a single article focuses on different aspects in the same study.

DOI: 10.1002/eet

Treib, 2008). According to Falkner and Treib's (2008) typology of the worlds of compliance, we divided the countries in the sample into two groups. Falkner and Treib (2008) argue that practical implementation of EU directives will run more smoothly in the countries of the 'world of law observance' and the 'world of domestic politics'[5] (column G, Table 3) than in countries of the 'world of dead letters' and the 'world of transposition neglect' (column H, Table 3). It seems that we know relatively less on the 'scope' dimension of the implementation performance in the countries where practical implementation was hypothesized to be more problematic (column H) than in the countries where implementation can be expected to be smoother (column G). The opposite seems to hold for the 'effort' dimension of implementation performance, where we know relatively more of this dimension of implementation performance in countries with hypothesized problematic implementation. This finding demonstrates the need to study implementation performance more systematically to draw better conclusions on how 'problematic' the implementation of EU directives actually is and how seriously the implementers take EU policies. As for now, due to the fragmented nature of our knowledge on the implementation performance important information might have been omitted from its evaluation.

Conclusions

This paper has proposed a three-dimensional conceptual framework for studying EU implementation performance that allows for systematic analysis of variation in practical implementation, while going beyond a conventional dichotomous understanding of compliance. The validity of the theoretically deduced conceptual framework was corroborated by an extensive qualitative content analysis of previous policy-specific empirical research on the implementation of EU environmental directives.

Even though environmental policy has been often claimed to be the most extensively researched policy field, the systematic literature analysis reported in this paper revealed that our knowledge of the practical implementation of EU environmental directives is fragmented in three ways. First, so far the various aspects of EU implementation performance have been examined neither equally nor systematically. While the 'substance' dimension has received most attention in the literature, much less attention has been paid to the 'scope' and 'effort' dimensions. Second, there is a need for the examination of practical implementation of directives other than the WFD, as other environmental directives have not been examined as frequently and as systematically as the WFD. When comparing the implementation of different directives, it is important to pay attention to 'monitoring', 'territory', 'staff' and 'prioritization' aspects, as these have received less attention up until now. And third, the analysis has revealed the need for more systematic research into countries where the practical implementation can be expected to be relatively problematic.

If the goal of research into EU environmental policy implementation, but also EU compliance, is to understand to what extent the member states really make EU policies work, there is a need for a more systematic approach to study implementation performance than has been practised up until now. If we continue to argue that there is a compliance deficit in the EU, we need to pinpoint exactly where, how and what is lacking in policy implementation. This conceptual framework can facilitate such research by offering a fully fledged conceptualization of implementation performance.

Acknowledgements

The authors would like to thank Eva Thomann, Jale Tosun and anonymous reviewers for their comments on the earlier versions of this paper.

[5]In this world of compliance, notwithstanding the fact that practical implementation is smooth, transposition with EU directives is contingent on domestic politics (Falkner and Treib, 2008).

References

Adcock R, Collier D. 2001. Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* **95**(3): 529–546. https://doi.org/10.1017/S0003055401003100.

Angelova M, Dannwolf T, König T. 2012. How robust are compliance findings? A research synthesis. *Journal of European Public Policy* **19**(8): 1269–1291. https://doi.org/10.1080/13501763.2012.705051.

Baldwin R, Cave M. 1999. Understanding Regulation: Theory, Strategy, and Practice. Oxford University Press: Oxford.

Barrett S, Fudge C. 1981. Policy and Action: Essays in the Implementation of Public Policy. Methuen: London.

Bauer MW, Knill C. 2014. A conceptual framework for the comparative analysis of policy change: measurement, explanation and strategies of policy dismantling. *Journal of Comparative Policy Analysis: Research and Practice* **16**(1): 28–44. https://doi.org/10.1080/13876988.2014.885186.

Bauer MW, Knill C, Pitschel D. 2007. Differential Europeanization in Eastern Europe: the impact of diverse EU regulatory governance patterns. *Journal of European Integration* **29**(4): 405–423. https://doi.org/10.1080/07036330701502431.

Beijen BA. 2011. The implementation of European Environmental Directives: Are problems caused by the quality of the directives? *European Energy and Environmental Law Review* **20**(4): 150–163.

Bennett G. 1993. The implementation of EC environmental directives: the gap between law and practice. *Science of the Total Environment* **129**(1–2): 19–28. https://doi.org/10.1016/0048-9697(93)90158-3.

Berke P, Backhurst M, Day M, Ericksen N, Laurian L, Crawford J, Dixon J. 2006. What makes plan implementation successful? An evaluation of local plans and implementation practices in New Zealand. *Environment and Planning. B, Planning & Design* **33**(4): 581–600. https://doi.org/10.1068/b31166.

Börzel TA. 2003. Environmental Leaders and Laggards in Europe. In Why There Is (Not) a "Southern Problem". Ashgate: Aldershot, UK.

Börzel TA, Risse T. 2000. When Europe hits home: Europeanization and domestic change. *European Integration Online Papers (EIoP)* **4**(15). Available at http://ssrn.com/abstract=302768 (accessed 20 April 2016).

Börzel TA. 2000. Improving compliance through domestic mobilisation? New instruments and effectiveness of implementation in Spain. In Implementing EU Environmental Policy: New Directions and Old Problems, Knill C, Lenschow A (eds). Manchester University Press: Manchester; 222–250.

Bugdahn S. 2005. Of Europeanization and domestication: the implementation of the environmental information directive in Ireland, Great Britain and Germany. *Journal of European Public Policy* **12**(1): 177–199. https://doi.org/10.1080/1350176042000311961.

Engeli I, Allison CR. 2014. Comparative Policy Studies: Conceptual and Methodological Challenges. Palgrave Macmillan: Basingstoke, UK.

Falkner G, Treib O. 2008. Three worlds of compliance or four? The EU-15 compared to new member states. *JCMS: Journal of Common Market Studies* **46**(2): 293–313. https://doi.org/10.1111/j.1468-5965.2007.00777.x.

Hartlapp M. 2009. Implementation of EU social policy directives in Belgium: what matters in domestic politics? *Journal of European Integration* **31**(4): 467–488. https://doi.org/10.1080/07036330902920051.

Hartlapp M, Falkner G. 2009. Problems of operationalization and data in EU compliance research. *European Union Politics* **10**(2): 281–304. https://doi.org/10.1177/1465116509103370.

Haverland M. 2003. The impact of the European Union on environmental policies. In The Politics of Europeanisation, Featherstone K, Radaelli CM (eds). Oxford University Press: Oxford; 203–221.

Haverland M. 2006. Does the EU cause domestic developments? Improving case selection in Europeanisation research. *West European Politics* **29**(1): 134–146. https://doi.org/10.1080/01402380500389331.

Haverland M, Romeijn M. 2007. Do member states make European policies work? Analysing the EU transposition deficit. *Public Administration* **85**(3): 757–778. https://doi.org/10.1111/j.1467-9299.2007.00670.x.

Héritier A. 2002. New modes of governance in Europe: policy making without legislating? In The Provision of Common Goods: Governance Across Multiple Arenas, Héritier A (ed.). Rowman & Littlefield: Boulder, CO; 185–206.

Hill M, Hupe P. 2003. The multi-layer problem in implementation research. *Public Management Review* **5**(4): 471–490. https://doi.org/10.1080/1471903032000178545.

Howlett M, Cashore B. 2014. Conceptualizing public policy. In *Comparative Policy Studies: Conceptual and Methodological Challenges* Engeli I, Allison CR (ed.). Palgrave Macmillan: Basingstoke, UK; 17–33.

Howlett M, Ramesh M, Perl A. 2009. Studying Public Policy: Policy Cycles and Policy Subsystems, 3rd edn. Oxford: Oxford University Press.

Huber JD, Shipan CR. 2002. Deliberate Discretion?: The Institutional Foundations of Bureaucratic Autonomy. Cambridge University Press: Cambridge, UK.

Hupe P. 2014. What happens on the ground: persistent issues in implementation research. *Public Policy and Administration* **29**(2): 164–182. https://doi.org/10.1177/0952076713518339.

Hupe P. 2011. The thesis of incongruent implementation: revisiting Pressman and Wildavsky. *Public Policy and Administration* **26**(1): 63–80.

Hupe PL, Hill MJ. 2016. 'And the rest is implementation'. Comparing approaches to what happens in policy processes beyond Great Expectations. *Public Policy and Administration* **31**(2): 103–121. https://doi.org/10.1177/0952076715598828.

Jans JH, Squintani L, Aragão A, Macrory R, Wegener BW. 2009. "Gold plating" of European environmental measures? *Journal for European Environmental and Planning Law* **6**(4): 417–435. https://doi.org/10.1163/161372709X12608898676797.

Jenkins-Smith H, Nohrstedt D, Weible C, Sabatier A. 2014. The advocacy coalition framework: foundations, evolution, and ongoing research. In *Theories of the Policy Process* Sabatier A, Weible C (ed.). Westview Press: Boulder, CO; 183–224.

Kippendorf K. 2004. Content Analysis: an Introduction to Its Methodology, 2nd edn. Thousand Oaks, CA: SAGE.

Knill C, Lenschow A. 1998. Coping with Europe: the impact of British and German administrations on the implementation of EU environmental policy. *Journal of European Public Policy* **5**(4): 595–614. https://doi.org/10.1080/135017698800000041.
Knill C, Lenschow A. 2000. Implementing EU Environmental Policy: New Directions and Old Problems. Manchester University Press: Manchester, UK.
Knill C, Lenschow A. 2004. Modes of regulation in the governance of the European Union: towards a comprehensive evaluation. In The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance, Jordana J, Levi-Faur D (eds). Edward Elgar Publishing: Cheltenham, UK; 218–244.
Knill C, Schulze K, Tosun J. 2012. Regulatory policy outputs and impacts: exploring a complex relationship. *Regulation and Governance* **6**(4): 427–444. https://doi.org/10.1111/j.1748-5991.2012.01150.x.
König T, Mäder L. 2013. Non-conformable, partial and conformable transposition: a competing risk analysis of the transposition process of directives in the EU15. *European Union Politics* **14**(1): 46–69. https://doi.org/10.1177/1465116512447703.
Laffan B, O'Mahony J. 2008. Ireland and the European Union. Palgrave Macmillan: Basingstoke, UK.
Lange B. 1999. Compliance construction in the context of environmental regulation. *Social and Legal Studies* **8**(4): 549–567. https://doi.org/10.1177/a010430.
Leventon J, Antypas A. 2012. Multi-level governance, multi-level deficits: the case of drinking water management in Hungary. *Environmental Policy and Governance* **22**(4): 253–267. https://doi.org/10.1002/eet.1590.
Liefferink D, Wiering MA, Uitenboogaart Y. 2011. The EU Water Framework Directive: a multi-dimensional analysis of implementation and domestic impact. *Land Use Policy* **28**(4): 712–722. https://doi.org/10.1016/j.landusepol.2010.12.006.
Lindblom CE, Woodhouse EJ. 1993. The Policy-Making Process, 3rd edn. Englewood Cliffs, NJ: Prentice Hall.
Mastenbroek E. 2005. EU compliance: Still a 'black hole. *Journal of European Public Policy* **12**(6): 1103–1120. https://doi.org/10.1080/13501760500270869.
Mastop H, Faludi A. 1997. Evaluation of strategic plans: the performance principle. *Environment and Planning. B, Planning & Design* **24**(6): 815–832. https://doi.org/10.1068/b240815.
May P, Winter S. 1999. Regulatory enforcement and compliance: examining Danish agro-environmental policy. *Journal of Policy Analysis and Management* **18**(4): 625–651. https://doi.org/10.1002/(SICI)1520-6688(199923)18:4<625::AID-PAM5>3.0.CO;2-U.
May PJ. 2003. Policy design and implementation. In Handbook of Public Administration, Peters BG, Pierre J (eds). SAGE: London, UK; 223–233.
Morris RKA. 2011. The application of the Habitats Directive in the UK: compliance or gold plating? *Land Use Policy* **28**(1): 361–369. https://doi.org/10.1016/j.landusepol.2010.04.005.
Newig J, Fritsch O. 2009. Environmental governance: participatory, multi-level - and effective? *Environmental Policy and Governance* **19**(3): 197–214. https://doi.org/10.1002/eet.509.
Newig J, Koontz TM. 2014. Multi-level governance, policy implementation and participation: the EU's mandated participatory planning approach to implementing environmental policy. *Journal of European Public Policy* **21**(2): 248–267. https://doi.org/10.1080/13501763.2013.834070.
Petticrew M, Roberts H. 2006. Systematic Reviews in the Social Sciences: A Practical Guide. Blackwell Publishing: Oxford, UK.
Potter WJ, Levine-Donnerstein D. 1999. Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research* **27**(3): 258–284. https://doi.org/10.1080/00909889909365539.
Radaelli CM, De Francesco F. 2007. Regulatory Quality in Europe: Concepts, Measures and Policy Processes. Manchester University Press: Manchester, UK.
Saetren H. 2014. Implementing the third generation research paradigm in policy implementation research: an empirical assessment. *Public Policy and Administration* **29**(2): 84–105. https://doi.org/10.1177/0952076713513487.
Sager F, Thomann E, Zollinger C, Mavrot C. 2014. Confronting theories of European integration: a comparative congruence analysis of veterinary drug regulations in five countries. *Journal of Comparative Policy Analysis: Research and Practice* **16**(5): 457–474. https://doi.org/10.1080/13876988.2014.960244.
Schaffrin A, Sewerin S, Seubert S. 2015. Toward a comparative measure of climate policy output. *Policy Studies Journal* **43**(2): 257–282. https://doi.org/10.1111/psj.12095.
Schreier M. 2012. Qualitative Content Analysis in Practice. SAGE Publications: London, UK.
Scott J. 2000. Flexibility in the implementation of EC environmental law. In Yearbook of European Environmental Law, Somsen H (ed.). Oxford University Press: Oxford, UK; 37–60.
Spicker A. 2006. Policy Analysis for Practice: Applying Social Policy. Policy Press: Bristol, UK.
Steunenberg B. 2007. A policy solution to the European Union's transposition puzzle: interaction of interests in different domestic arenas. *West European Politics* **30**(1): 23–49. https://doi.org/10.1080/01402380601019639.
Steunenberg B, Dimitrova AL. 2014. Too difficult to handle? Compliance with EU policy in a multi-level context. Paper presented at the 2014 Annual APSA Conference, Washington. Available at http://ssrn.com/abstract=2454678 (accessed 20 April 2016).
Thomann E. 2015. Customizing Europe: transposition as bottom-up implementation. *Journal of European Public Policy* **22**(10): 1368–1387. https://doi.org/10.1080/13501763.2015.1008554.
Töller AE. 2010. Measuring and comparing the Europeanization of national legislation: a research note. *JCMS: Journal of Common Market Studies* **48**(2): 417–444. https://doi.org/10.1111/j.1468-5965.2009.02058.x.
Toshkov D, Knoll M, Wewerka L. n.d.. Compliance Database: An Online Database of Case Study Research of EU Law Application, in cooperation with the Institute for European Integration Research at the Austrian Academy of Sciences.
Tosun J. 2012. Environmental monitoring and enforcement in Europe: a review of empirical research. *Environmental Policy and Governance* **22**(6): 437–448. https://doi.org/10.1002/eet.1582.
Treib O. 2014. Implementing and complying with EU governance outputs. *Living Reviews in European Governance* **9**(1): 1–47.

Vedung E. 1997. Public Policy and Program Evaluation. Transaction Publishing: New Brunswick, NJ, USA.
Vedung E. 1998. Policy instruments: typologies and theories. In Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation, Bemelmans-Videc ML, Rist RC, Vedung E (eds). Transaction Publishing: New Brunswick, NJ, USA; 21–58.
Versluis E. 2007. Even rules, uneven practices: opening the 'black box' of EU law in action. *West European Politics* **30**(1): 50–67. https://doi.org/10.1080/01402380601019647.
Voermans WJM. 2015. Implementation: the Achilles heel of European integration. *The Theory and Practice of Legislation. SSRN Electronic Journal* **2**(3): 343–359. https://doi.org/10.2139/ssrn.2510310.
Winter SC. 2006. Implementation. In Handbook of Public Policy, Peters BG, Pierre J (eds). SAGE Publications: London, UK; 151–166.
Winter SC. 2012. Implementation perspectives: status and reconsideration. In Handbook of Public Administration, Peters BG, Pierre J (eds), 2nd edn. SAGE Publications: London, UK; 255–264.
Zhelyazkova A. 2013. Complying with EU directives' requirements: the link between EU decision-making and the correct transposition of EU provisions. *Journal of European Public Policy* **20**(5): 702–721. https://doi.org/10.1080/13501763.2012.736728.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.

Appendix. The list of articles reviewed for the systematic literature review.