

Chapter 33*

* A previous version of this chapter was presented at the European Consortium of Political Research (ECPR) sixth biennial Regulatory Governance conference, Tilburg, 6-8 July 2016. We are grateful to Samuele Dossi, Erica Melloni, Lorna Schrefler, Stijn van Voorst for their comments on that earlier draft. Laurie Buonanno made additional suggestions at the editorial stage going beyond the call of editorial duty. Our research on policy instruments in the European Union (EU) is funded by the project PROTEGO, Procedural Tools for Effective Governance, European Research Council grant 694632.

If Evaluation is the Solution, What is the Problem?

Claire A Dunlop and Claudio M Radaelli,
Department of Politics, University of Exeter (UK)

Introduction

Evaluation has attracted so much attention from European Union (EU) policy-makers, who call for methods and standards of evaluation across the whole policy cycle, from inception to implementation (European Commission, 2015a). Through conditionality, the EU has diffused evaluation systems across public managers in regions and local authorities (Hoerner and Stephenson, 2012). EU-related evaluation practice has spawned communities of professional evaluators and consultancy firms in Member States previously alien to the culture of evaluation. And yet, there is a dearth of systematic social-science research on the subject: 'Policy evaluation in the EU has hardly been studied' (Smismans, 2015:6). 'Evaluation in the EU is an under-researched topic' (Hoerner and Stephenson, 2012:699). The few studies available discuss the methods of evaluation. Much less is known about its usages, the link between ex ante (pre-project evaluations such as feasibility studies, impact assessments, forecasting, and policy analyses) and ex post (unlike summative program evaluations, ex post are conducted some time after project termination in order to evaluate the project's long-term impact and sustainability), 'who is getting what' – or, bluntly, the politics of EU evaluation. Major textbooks on EU politics and policies do not even have a chapter on evaluation – for an exception: Versluis et al. (2011; chapter 9).

There have been some studies and reports on EU evaluation – see, for example, Smismans (2015). Yet even if there is (some) literature, it is divided internally, between those who have examined ex-post evaluation of expenditure programs and interventions, and

those who have traced the evaluation of regulation and legislation, mostly ex-ante but more recently ex-post too.

But what is evaluation in the EU, exactly? According to the recent better regulation toolbox, evaluation is 'a tool to help the Commission services assess the actual performance of EU interventions compared to initial expectations. By evaluating, the Commission services take a critical look at whether EU activities are fit for purpose and deliver, at minimum cost, the desired changes to European businesses and citizens and contribute to the EU's global role. Evaluation also provides a key opportunity to engage stakeholders and the general public, encouraging feedback on how EU interventions are perceived' (European Commission, 2015b: 253-254). And again: '*evaluation is not an assessment of what has happened; it considers why something has occurred (the role of the EU intervention), and, if possible, how much has changed as a consequence*' (European Commission, 2015b:254, emphasis in original). This is a notion that chimes with the literature on ex post evaluation as analytical activity to learn from the implementation of policies or projects (Shadish, Cook and Leviton, 1990). However, the Commission has also produced guidance on ex-ante evaluation of expenditure programs, with the 2001 guidelines, issued by DG Budget, on 'Ex ante evaluation: A practical guide for preparing proposals for expenditure programs' (European Commission, 2001a). Indeed, ex ante evaluation of expenditure is legally binding under the Financial Regulation applicable to the general budget of the EU.

The complexity does not end here. On the same page of the Commission's better regulation toolbox we encounter something that is – so the Commission says – 'like an evaluation'. In fact, 'A Fitness Check (FC) is like an evaluation of an individual intervention except that it covers a group of measures which have some relationship with each other justifying a joint evaluation' (European Commission, 2015b: 254). Interestingly, these definitions are contained in a toolbox dedicated to better *regulation* (not to expenditure or

specific interventions). Actually this guidance document identifies other tools that have some important evaluative dimensions, like impact assessment (IA) of proposed *legislation*.

IA is a tool to appraise the effects of policy proposals on a wide range of stakeholders and across different categories of costs and benefits. Whilst 80 per cent of ex-post evaluations are carried out by consultancy firms (the figure is cited by Højlund, 2015:36, who draws on a study by Sven Jacobsen in Danish), IA is mostly performed in house – thus differentiating the mechanisms of the two tools in fundamental ways. In one case the Commission is a user of knowledge produced elsewhere, in another it draws on IA to build analytical capacity in house. IA can cover a whole framework directive, like the thematic Air Directive of the EU (Meuwese, 2009), whilst ex-post evaluation in the context of cohesion funds may refer to single projects in a given region of Europe.

Then we have proper audit tools which intersect with the logic of evaluation activities: the Court of Auditors has shifted to performance and effectiveness audits – as opposed to more limited financial audits (Stephenson, 2015). On some occasions the Court of Auditors has evaluated the IA system of the Commission – this is a kind of meta-evaluation (ECA, 2010). But external evaluations of IA have also been commissioned directly by Brussels: The Evaluation Partnership (2007) provided its own evaluation of the Commission’s approach to IA in response to a tender launched by Brussels (under pressure by Member States to demonstrate the value for money of IA).

It’s enough to confuse *any* reader, unless one knows about the history of policy evaluation and how ideational politics and constellations of actors, tools, and methods have evolved during the last three decades. Evaluation is also an ideational frame and a reference for organizational activity where the Member States, the Commission as complex organization, the European Parliament (EP) and the stakeholders define who has control over EU policy, who defines ‘success’, who can exercise oversight on the quality of evaluation

(for example, the Commission or independent bodies). As such, the history of evaluation in the EU may look like an ambiguous set of odd pieces if we assume that the only problem is to create coherence and rationality across the policy cycle. But the same ambiguity allows the key actors to play their ideational and organizational politics, and constantly re-define what problems are associated with evaluation as solution.

To clarify matters, then, we have to take a broad, historical approach to evaluation. First, we do not necessarily make the assumption that its main purpose is to help the Commission to learn via evidence. There may be other institutions and stakeholders that benefit from evaluation. Or, to put it with different roles, evaluation may serve either the objective of learning or the objective of making an organization (or a unit within an organization) accountable to other organizations and therefore controllable. Or both, but possibly with some trade-offs between the two – compare Smismans (2015) and Højlund (2015). Or none – in which case evaluation is yet another perfunctory routine performed by a bureaucracy.

Second, we consider two dimensions of evaluation (see Table 33.1) and how they have become intertwined over the years, providing more (or less) coherence and creating new challenges, but also opportunities for evidence-based policy, learning, accountability and control. The two dimensions are whether an evaluation is performed at the stage of policy formulation or when policy is already on the road. The first dimension is called ex-ante, the other ex-post. In the Commission, the ex-post dimension is divided into formative, mid-term and summative evaluation. In the second dimension we distinguish between the broad categories of EU ‘expenditure’ and ‘regulation’ because these are two policy types well-known in the field of public policy analysis.

<<Insert Table 33.1 about here>>

In Table 33.1 we portray the names that these different evaluations have in the EU language. Behind the table's cells lie distinct communities of expertise, tools and constellations of actors for which evaluation is a solution to different problems. These constellations want something different from the practice of evaluation. This is our major theme, as indicated by the title we have chosen. Here we note that there is a separation in the literature between the columns of 'expenditure' and 'regulation' in Table 33-1. The evaluation community of expenditure programs such as structural funds and the Common Agricultural Policy has its own journals like *Evaluation* and *Evaluation Review*, and a consolidated professional network, the *European Evaluation Society*. It is not anchored to a particular discipline. By contrast, the literature supporting the 'regulation' column of our table 1 is supported by social scientists who see themselves as lawyers, economists, and political scientists working on governance and regulation. These experts publish in mainstream journals of law, political science, public economics and public management. They do not have their own large European-wide network, although there is a *Society for Benefit-Cost Analysis* <https://benefitcostanalysis.org/> (popular mostly with North-American scholars rather than Europeans) and in Europe the *European Consortium for Political Research* hosts a section on Regulatory Governance <http://regulation.upf.edu/> (the median member of the section, however, is not typically interested in regulatory evaluation).

Time is also fundamental, because these approaches to evaluation have not developed in synch, although at some important junctures they have met. To illustrate: 'regulation' or more precisely the regulatory paradigm pursued by the Commission has gradually become the focal point for the definition of the purpose of evaluation. In other words, the better regulation philosophy is today the dominant 'doctrine' in the Commission's thinking about what evaluation should be and for what purposes.

We then offer our historical account, differentiating between ex post evaluation in the context of cohesion policy (Molle, 2008) and the rise of ex ante IA standards and activity, finding different constellations of actors and tools. We will then zoom-in on the new kids on the bloc in the field of ex post regulatory evaluation. In the last section we answer our research question about solution-problem associations and the analytical conclusions.

(Some) Evaluation Practice and the Early Days of the Evaluation System

Højlund (2015: 36) follows previous scholarship and distinguishes between evaluation *practices* and the more institutionalized evaluation *system*. In the latter there are organizational structures for evaluation, a relatively stable network of actors with some shared beliefs about evaluation, and a steady flow of evaluations produced regularly.

Obviously evaluation in the EU is part of broader waves of evaluation approaches that have spawned across Europe during different decades – for a recent historical account, see Pattyn et al. (2016). Taking a very broad approach, there is a well-developed literature on transposition, implementation, compliance & enforcement, and Europeanization (Falkner et al. 2005; König and Luetgert, 2009; Radaelli, 2017). While these fields deal with ex-post topics, but they do not have the aim of evaluating at the center of their intellectual enterprise.

In this chapter we focus on ex post evaluation of expenditure and programs, mostly in the field of cohesion policy (though for agricultural policy see Bozzini and Hunt [2015]). For the period 1980-1994, Højlund (2015) finds evidence of ex post evaluation practice but not a systematic practice; rather evaluation seemed to involve the ad hoc production of ex-post evaluations, especially around DG Development Aid and DG Research – both with substantial expenditure programs. Impact assessment and ex-post regulatory evaluations were unknown in this period. This situation changed in the 1990s the Structural Funds were more

than doubled with the second Delors package, introducing DG Regio firmly in the network of actors with a permanent interest and substantial expertise in this type of evaluation.

Initially, ad hoc, case-by-case evaluation of expenditure was a solution to the problem of learning about the effectiveness of development aid and research programs. Then, in the 1990s, evaluation became a solution to the problem of keeping public spending under control, limiting fraud, and producing financial accountability for the structural funds (Højlund, 2015). Towards the end of the 1990s an embryonic evaluation system emerged, in the wake of the implementation of the Financial Regulation of 1995. Evaluation evolved more decisively than before into a tool to address the problem of sound financial management. Accordingly, evaluation gravitated around DG Budget, DG Financial Control and the Secretariat General. The DGs were encouraged to develop evaluation units, although those who did not manage expenditure programs did not see the point (Højlund, 2015:41). In the end, expenditure-related DGs learned how to manage evaluations during the 7-year Commission's policy cycle. Here 'managing' means that officers learned how to commission evaluations, deal with the contractors, and set the timing of their tenders. Højlund reports that this activity was decoupled from the design of policy – also because legislation was not covered by these evaluations and most of the policy activity of the Commission was and is through lawmaking.

Additionally, in 1995 the Commission mobilized epistemic learning with a comprehensive evaluation of its guidelines and the production of state-of-the-art methodological standards – this was called the MEANS program. The legacy of MEANS is a six-volume collection on the evaluation of socio-economic programs (European Commission, 1999). Today, the ex post evaluation of non-expenditure programs, for example in DG Regio, benefits from advanced methodological thinking on counterfactuals, quasi-experimental research designs, and outcome indicators and targets (Barca and McCann, 2011; the portal

EVALSED, which appeared first in 2013, is the successor of MEANS; see EVALSED in European Commission 2008).

On Separate Tracks: Ex Ante and Ex Post

In the last 15 years or so, ex-post evaluation has become more closely connected to the objectives of financial accountability. Looking at the many reports and communications (see Højlund, 2015) on ex-post evaluation produced in the context of attempts to ‘reform’ the Commission, ‘make it slimmer’, ‘less bureaucratic’, ‘modern’ and ‘better managed’, we argue that ex-post evaluation has also become a pivotal tool in organizational reforms. In short, a solution to the problem of improving on management, as shown by the connection between the push for evaluation and the so-called Kinnock reforms (Kassim, 2008; European Commission, 2000). Højlund (2015: 43) reports that a 2002 survey showed the majority of unit staff looked at legal requirements as the main reasons why formative, mid-term and summative evaluations were carried out. At the same time, this standardization of evaluation of structural funds, with its mandatory requirements, allowed a community of professional evaluators to consolidate across Europe. Countries like Italy had not been exposed to evaluation before the arrival of EU-mandatory evaluations of cohesion policy. With enlargement to the East, Eastern Europe became familiar with evaluation via conditionality. Since 2009, the Secretariat General has taken on the role of pivotal actor for standard setting and the dissemination of guidance in evaluation – with DGs like Regio, however, still holding on to their analytical and methodological ‘competitive advantage’ on cohesion funds. This advantage emerged in a context where cohesion policy accounted for 347 billion euros in 2007-2013, generating pressure for associating strong evaluation as a solution to the problems of implementation, fraud, and, to the challenge of managing strict auditing obligations. For example, Member States were obliged to designate an audit authority to

oversee the implementation of the 2007-2013 cohesion funds, but the Commission can still carry out audits at short notice (Hoerner and Stephenson, 2012: 701).

The progress on ex post regulatory evaluation has been slower, with acceleration during the last part of the Barroso II Commission, and continuing in the Juncker Commission. The OECD (2014) published and promoted a new framework for regulatory evaluation across the policy cycle, from IA to ex post appraisal (see also the OECD expert paper by Cary Coglianese, 2012 on establishing causality in regulatory evaluation). The OECD's work has complemented the evolution of thinking about evaluation at the top of the Commission. For Brussels, however, the important point was to raise the stakes for better regulation, not what the OECD thought; for example, Jean-Claude Juncker (Commission president) and Frans Timmermans (Commission vice-president with the regulatory portfolio) claim that better regulation underpins (a) all the policy formulation work of the Commission (see Commission, 2015c), (b) the withdrawal of obsolete or burdensome initiatives, and (c) the ambition to draw on appraisals in order to close the policy cycle. In this connection, the Commission has taken a vow to 'evaluate first'. Thus, no policy formulation work should start absent systematic knowledge of what previous policy has delivered. More generally, the ambition is to connect ex ante and ex post regulatory evaluation. At the time of this writing in late 2016, we can only describe this as ambition: van Golen and van Voorst (2016) gathered 309 ex-post legislative evaluations and 225 IAs – finding that only 9 ex-post evaluations use IA and 33 IAs draw in some ways on ex-post evaluations.

The fact that the Secretariat General is today the key actor both in ex ante and ex post regulatory evaluation is the organizational expression of an impulse coming from the very center of the Commission. Perhaps more importantly, the political composition (center-right) of the Commission has generated this change, as well as the relative power of key Commissioners like Timmermans in the Juncker Commission.

If evaluation is a solution to the problem of integrating (analytically) and closing the policy cycle, then the old style evaluations of initiatives and *programs* offer little assistance to *policy* design. This explains why evaluation tools have somewhat mutated in the EU, with the arrival of new concepts like REFIT and tools like Fitness Checks and Cumulative Cost Assessments and a strong pressure towards systematic ex-post evaluation.

But there is another important connection, at the level of ideational politics. Ex post regulatory evaluation and IA, together with their integration across the policy cycle, are framed as key tools in the strategy of reducing regulatory burdens, keeping compliance and regulatory costs under control, and using regulatory policy to enhance the competitiveness of European firms. This regulatory paradigm has altered the conceptual framing of IA too, not just regulatory evaluation. This better regulation initiative of the EU emerged historically as a product of the European Commissions (2001b) White Paper on Governance, but it is also a response to a long-standing criticism of the quality and number of regulations produced by the EU. This criticism can be traced to the 1992 Edinburgh Council meeting, which marks the first summit in which EU heads of government discussed collectively actions to improve on the number of rules and quality of lawmaking. (In the 1990s this debate went under the name of ‘better lawmaking’). After a number of reports and relatively inconclusive pilot projects on business IAs and simplification (reviewed by Radaelli, 1999 and Gravey, 2016), the Commission launched its ‘Better Regulation Strategy’ in 2002. On the one hand, the strategy was based on Commission President Prodi’s ideas for new governance (systematic consultation, pluralistic usage of expertise, transparency, simplification and evidence-based policy proposals); on the other hand, this strategy was a response to Member States (led by Germany, the Netherlands and the UK). Some delegations in fact complained that the proposals for new EU legislation were not properly appraised and that the life-cycle of regulation was not managed. Absent rigorous evidence-based controls on new proposals and

existing regulations (this was the argument), the Commission could exploit its Treaty right to initiative legislation to engage in regulatory creep and task expansion.. Indeed, the UK and the Netherlands submitted ‘hit lists’ of regulations to be scrapped (see the historical account in Gravey, 2016). The then DG Enterprise and Industry was instrumental in creating bridges between these two views, although not powerful enough to create momentum and support from other DGs, the latter being more concerned about the possible negative impact of the ‘better regulation’ initiatives on the environment and the European Social Model. When the Secretariat General emerged as the better regulation pilot a compromise on what IA was for was struck (Allio, 2008).

IA emerged as the main pillar of these initiatives, defining the EU approach to ex-ante evaluation of non-expenditure proposals. But what type of ‘evaluation’ does IA carry out? The official line of the Commission is that IA is a single template for the analysis of different types of impacts around the pillars of economic, environmental, and social costs and benefits. This is a peculiar take on the task of an evaluation, clearly influenced by the internal politics of the Commission. In fact, as mentioned, in the early 2000s the Commission defined IA with a compromise among its main three internal stakeholders: the DGs tasked with promoting enterprise and the economy (DG Enterprise and Industry, today DG Grow: Internal Market, Industry, Enterprise and SMEs), the environmental DG(s), and DG Employment (Allio, 2008). Thus, on the one hand IA in the EU follows the more general trend towards evidence-based appraisals of new regulation: most OECD countries have IA systems, and many of them emerged in the years 1995-2005. At the same time, the precise incarnation of IA with its three dimensions of economic, social and environmental analysis refracts the organizational politics within the Commission.

In 2005, however, the EU redefined better regulation more decisively in the direction of the reduction of regulatory burdens. At the same time, better regulation became a major

component of the so-called Lisbon strategy for growth and jobs, the flagship agenda of the EU. Ironically, better regulation scaled up in ambition and ranked high on the Lisbon agenda, but at the cost of becoming more narrowly defined in terms of de-regulation rather than regulatory quality (Radaelli, 2007). This change in the ‘meaning’ and ‘practice’ of evaluation in the Commission can be, at least partly, attributed to some of Member State finance ministers (particularly Finance Ministers, at that time, Gerrit Zalm in the Netherlands and Gordon Brown in the UK). Significantly, at the same time these countries were experimenting at home with plans for the reduction of administrative burdens. Thus, the 2005 re-definition of the evaluation problem for which better regulation and IA in particular were proposed as a solution was clear: the problem was un-necessary (regulatory) costs and more generally excessive, burdensome regulation. This ideational paradigm has always been absent from ex post evaluation, where the evaluation of cohesion policy evolved around different problems, as we have seen above.

Interestingly, this turn towards de-regulatory ideas did not necessarily make IAs more narrowly defined. Data published by Fritsch et al. (2013) show that the three-pillar structure (that is, economic, environment and employment-social) of ex-ante regulatory evaluation (IA in short) was somewhat preserved over the years. The main tool adopted to reduce ‘red tape’ at the stage of policy formulation (the so-called standard cost model) does not show up systematically in the IAs reviewed by Fritsch and colleagues. Essentially, we have to distinguish day-to-day reality from (more or less ideological) statements about the purpose of IA.

Thus, at the end of our historical journey, we find that ex post evaluation of projects started early and evolved with its actors and tools around problems of learning, and more recently accounting, implementation and oversight. IA emerged from a different constellation of actors and problems, tentatively in the 1990s and in more robust ways in the next decade.

IA was initially associated with the problem of regulatory *quality* across economic, social and environmental costs and benefits, but during the last ten years it has been associated more decisively with the problems arising from regulatory *quantity*. Ex post legislative evaluation arrived even later on the scene. Gradually, the ideational politics of better regulation started to dominate justifications for all evaluation activity. With the recent advent of the era of ‘closing the policy cycle’ better regulation made proprietary claims on the whole of evaluation activities. To make the proprietary claim on evaluation somewhat realistic, however, the custodians of EU better regulation needed new tools that would allow them to connect ex ante IA with ex post evaluation of legislation. This is the topic of the next section.

Ex-post Regulatory Evaluation: The Politics of REFIT

REFIT (Regulatory Fitness and Performance) is much more political than the evaluation of projects within the structural funds. It is a component of a broader attempt of the Barroso II and Juncker Commissions to re-define and scale up the ambition of the European Commission’s previous initiatives to reduce administrative burdens and produce better regulation ‘across the policy cycle’. Since the launch of REFIT in 2012, the ambition has been to review the entire *acquis*. In the history of evaluation at the Commission, ex-post evaluation has typically been confined to single initiatives (e.g. structural funds) or, more recently, pieces of legislation. REFIT has allowed the Commission to experiment with novel tools like the fitness checks which cover the whole of a policy field. The first wave of Fitness Checks included the politically sensitive areas of employment and social affairs – such as the discipline of information and consultation with workers. It also covered the domains of enterprise (the regulatory framework for the approval of motor vehicles) and the civil aviation sector. The Commission’s 2016 work programme contained 27 REFIT initiatives, of which at least 17 have a strong evaluative dimension (see [European Commission 2016c](#),

[Annex ii](#)). The main objective of the 27 initiatives, however, is to reduce regulatory costs; hence, these evaluative dimensions respond to this over-arching objective of REFIT. As such, they are narrow in scope: this evaluation activity only focuses on some cost categories. Furthermore, REFIT is designed as a ‘rolling programme’ and therefore serves as a springboard for regularly withdrawing proposals. The Commission argues that since 2006 almost 400 proposals have been withdrawn, – 73 in the 2015 Work Programme alone (see European Commission 2015d).

Starting in 2013, another tool—the cumulative cost assessment—appeared under the REFIT umbrella. The cumulative cost assessment considers a single sector and attempts to measure the effects of different EU rules on a sample of representative facilities. The first two applications of the cumulative cost tool were in the steel and aluminium sectors (Schrefler et al., 2015). Yet again, we see the emphasis on evaluation as tool to mitigate regulatory cost problems. Every other dimension is excluded.

The distinctive properties of REFIT are its transparency, the heavy (or heavier, when compared to the typical policy evaluation of the Commission) drive provided by consultation, the lead of the Secretariat General, the springboard nature of the initiative, and the coalitions in support and against it.

First, under REFIT the Commission has tried to improve on transparency by reporting regularly on the results of its regulatory evaluations of the *acquis*. Transparency has been used also to avoid blame for the state of EU regulation: in fact the REFIT scoreboards are built in such a way as to differentiate the regulatory costs created by EU regulation from the responsibility of the Member States. Thus, the transparency game is also a component of attributing responsibility and avoiding blame.

Second, REFIT consultation goes beyond the simple provision of information by the stakeholders. Indeed, business stakeholders become co-responsible for the initiatives carried

out by the program. REFIT has drawn on direct complaints from stakeholders through various channels. It has also promoted a major ‘listening’ exercise dedicated to business and small and medium enterprises in particular that led to the list of legislative items targeted as most burdensome (the so-called top-ten initiative, see Commission 2013).

Third, with REFIT – and more generally with the push for ex-post regulatory evaluation- the Secretariat General has emerged as the organizational lead in driving the Commission towards the systematic evaluation of the *acquis* – thus making this body the most important political locus of activity for evaluation. In the meantime, the Sec Gen has become the pivotal actor in IA (Radaelli and Meuwese, 2010). This consolidates the presence of this actor within evaluation across the policy cycle. In the end, for the Sec Gen evaluation is a solution to the problem of creating policy leadership. Even with respect to evaluation methods, the Secretariat General has raised its profile by producing an integrated framework for the evaluation of regulation across the whole policy cycle (European Commission, 2015b). Therefore, the Sec Gen is today the ultimate custodian of guidance on evaluation, both ex ante and ex post.

The fifth REFIT characteristic is coalitional politics. For those willing to push for regulatory culture change in the EU, there are essentially two options. One is to be openly and unashamedly for de-regulation, and demand a reduction of EU regulation across the board. This has been the position of the British coalition government and, even more so, of the former Cameron government. (Indeed, a more systematic reliance on IA and regulatory cost compression targets were included in Cameron’s deal on Brexit). The other option is to pursue objectives of de-regulation by adopting a positive posture – positive in relation to the nature of European integration, as well as with reference to the necessity of regulation. REFIT has offered such as option: although it is a channel for de-regulatory demands, it is not framed as anti-EU or anti-regulation per se. This explains why countries such as Germany

and the Netherlands have found it easier to support Barroso and Verheugen first, Juncker and Timmermans later, in their effort to carry out the REFIT operations – as opposed to supporting more (draconian) de-regulatory initiatives like the British ‘cut red tape’ report (Business TaskForce 2013).

The EU-level regulatory reform coalition has not passed unnoticed. In response to this coalition, environmental/social organizations and research institutes have mobilized against the regulatory reforms of the Commission and REFIT, in particular. For example, the European Trade Union Institute has published over the last few years briefing papers arguing that REFIT as ex-post evaluation was essentially an attack on the European Social Model, together with ex-ante appraisals, which is criticized as incapable of providing an accurate evaluation of health and safety legislation (van den Abeele, 2014; 2015; Schömann, 2015). In 2015 these concerns have materialized into an umbrella organization called ‘better regulation watchdog’ supported by 66 green, social and consumerist groups across Europe (<http://www.betterregwatch.eu/>).

Finally, although ambiguity is conceptually a minus for an evaluation system, politically it can be a plus. It is hard to find a single analytical concept or method behind REFIT, apart from the general notion that legislation has to be ‘fit for purpose’. REFIT is a reference to a certain (de)regulatory discourse, perhaps a general objective, not a policy tool – certainly it is far from the analytical and methodological status of evaluation methods in cohesion policy.

Exactly because of its conceptual ambiguity (not a proper evaluation tool since some REFIT exercises are just appraisals of some regulatory costs, not a method, elusive on how its results should be used politically), REFIT is a springboard for a variety of actions: consulting stakeholders, preparing hit lists of regulations to be targeted, reducing administrative burdens and compliance costs for business, slowing down regulatory action in

some sectors, and withdrawal of proposals. Another mutation of REFIT took place in May 2015, when the Commission launched a platform to collect comments on irritating or burdensome regulations, following the example of similar platforms in the UK (red tape challenge, <https://cutting-red-tape.cabinetoffice.gov.uk/>) and France (faire simple, <http://www.faire-simple.gouv.fr/>). The platform is called ‘lighten the road – have your say’ and shows how better regulation has also become a major communication and signaling tool – far removed from the discipline of proper, analytical regulatory evaluation. In the same month, REFIT was ‘beefed-up’ with a two-tier platform for dialogue with stakeholders and member states. Essentially the platform consists of two standing groups, one for governments and one for stakeholders. (Incidentally, the European Economic and Social Committee was included in the latter group, causing outrage in the EESC quarters who believe they belong to an institution rather than being yet another stakeholder.)

Discussion and Conclusions: Evaluation for Whom?

One way to explore the association between evaluation as solution and different problems is to first consider the empirical information on usages, and then return to the question that motivates this chapter. Usages are mainly inside the Commission, but evaluations can also be used by stakeholders, the EP, the Member States and auditing institutions.

At the outset, there are no data on how many evaluations (ex ante and ex post) are simply tick-the-box exercises and, ultimately, are not used. Little is known about why organizations do not evaluate certain activities but make efforts to evaluate others (for an exciting approach to this question see Pattyn, 2014 on Flemish organizations). For sure, it’s easier to write about learning than publishing papers on ‘non-effects’ or lack of usage of evaluations. However, it is also fair to say that most research (and even more so evaluation

studies by consultancy firms) revolve around questions of value for money spent or whether the IAs meet the standards set in the guidelines (Turnpenney et al, 2009 on IA; Hoerner and Stephenson 2012 on cohesion funds). Future research should attend more systematically to sophisticated questions on who gets what from evaluation activities.

Empirical research over the last 20 years (see the literature summarized above) suggest the ex post project evaluations have been used for the purpose of guiding program officers (perhaps more so in the early days) and financial accountability. Borràs and Højlund (2016) have examined in depth two mid-term and one interim evaluation. Using interviews, they report learning effects on program units and external evaluators rather than external and organizational stakeholders. Their study digs deep on an important dimension of evaluation as solution to the problem of learning, but it draws on a limited sample of three. There is also evidence of evaluation as solution to the problem of financial accountability: when evaluation is legally binding, it is often carried out as a necessary step in the execution of the initiative (Højlund, 2015).

Turning to ex post legislative evaluations, research has just started, with van Golen and van Voorst (2016) suggesting limited use of legislative evaluations to inform the policy cycle – something not surprising, given what we know about the use of evaluations in other political systems like the USA. Mastebroek et al. (2016) argue that ex post legislative evaluations are a matter of legislative obligation rather than the expression of the Commission’s initiative to learn and-or close the policy cycle. Their scorecard (see table 3 in their article, 2016) indicates that there is a justification of methods only in 15 per cent of their sample. Also, 39 per cent show no involvement of stakeholders consultation – but again, this is in line with what we know from other jurisdictions where the vast majority of notice and comments generate little public input, whilst few consultations attract a disproportionate attention because they are politicized by pressure groups’ campaigns.

As for the ex ante dimension, IA is a solution to different problems. For Member States and pressure groups it is a vector of possible control on the policy formulation activity of the Commission, since it provides avenues for consultation and information on what kind of benefits and costs are considered. Inside the Berlaymont, IA is a vector of more strategic and operational management at the center – Sec Gen, key Commissioners like Timmermans in the Juncker Commission, and the Regulatory Scrutiny Board. Apart from these power-centralizing effects, IA has undoubtedly made the policy formulation process more complex but has also increased the quality of the decisions taken by the Commission (Radaelli and Meuwese, 2010). IA, however, does not necessarily determine the nature of the decision of whether to regulate or not, but it (may) improve on the quality of a decision.

Damonte et al. (2014) also link IA to accountability. Has IA made the Commission more accountable to the European Parliament? Zwaan et al. (2016) have gathered a sample of 220 ex-post legislative evaluations. They found that 22 per cent appear in questions raised by Members of the European Parliament (MEPs) and 16 per cent are used to steer the behavior of the Commission. The variance in questions asked by MEPs is best explained by political conflicts between the EP and the Commission. We do not have recent data on how the EP handles IA, although past research suggests limited interest of the MEPs (The Evaluation Partnership, 2007; European Court of Auditors 2010). And yet, Anne Meuwese (2009) provides an fascinating study of how even a few, emblematic cases of IA usage by MEPs can be instrumental in re-defining power between the Commission and the EP in lawmaking. The EP has increased its analytical capacity to deal with evaluation and support its members via its research service (www.eptthinktank.eu established in 2013). The EP research service has its own IA and evaluation experts and a ‘policy cycle’ unit – to mirror the better regulation mantra to evaluate ‘across the policy cycle’. An inter-institutional agreement (European Commission and European Parliament, 2016) on better regulation locks in the joint

commitment for systematic evaluation of the EU institutions, (), although at the time of this writing it is too early to know how the agreement will be implemented.

At the end of the first year of the ‘better regulation’ Schrefler and Huber (2015) f provided an analysis of all evaluations under way (expenditure as well as legislative). They found a staggering number of planned and ongoing evaluations: 525 – of which 35 per cent were prompted by a legal requirement (e.g. financial regulation, legal basis of the multiannual financial framework), REFIT contributed to six per cent of total ongoing evaluations, and three per cent originated in the Commission’s ‘evaluate first’ commitment in its work programme . The variety of information on these evaluations was at the time puzzling, making it difficult for the EP to keep track of this evaluation activity. In the same year, the regulatory oversight body of the Commission, the Regulatory Scrutiny Board, was given a mandate to appraise the quality of both IA and ex post legislative evaluations to improve quality standards evenly across the different stages of the policy process. However, whilst experience for regulatory oversight of IA has been built since late 2007, the oversight activity on ex post regulatory evaluation is a novelty for the Commission, and capacity will have to be built over the years.

What is the overall EU evaluation portfolio? With the goal of ‘closing the policy cycle’ the bar has been set quite high. Perhaps this is a syndrome of the Commission, that when challenged does not retreat towards simple, deliverable plans: it scales up in ambition instead (Cram, 2001). Be that as it may, can evaluation in its different mutations become the solution to the problem of closing the policy cycle? At the moment the constellations of actors as well as the problems and methods of the different types of evaluations are not the same. Ex ante evaluation is produced in house for certain purposes, whilst the ex post project evaluations are produced by consultancy firms. The classic ex post evaluation of cohesion projects provides little guidance on the wider problem of designing policy. In principle,

REFIT contains the tools to fill the gap, but it is not clear what the analytical core of REFIT is or may be. Regulatory oversight has developed mostly with IA, but in the future the Regulatory Scrutiny Board will have to check on the quality of ex post evaluations too. The IA of a single directive makes the Commission accountable to certain stakeholders, whilst the wider REFIT initiatives should in principle make the Commission accountable to the stakeholders operating in an entire policy sector. IA has witnessed a limited but steady sophistication of methods (Fritsch et al., 2013) whilst ex-post legislative evaluations are not robust in methods (Mastenbroek et al., 2016).

This leads us to conclude that at least at this moment in time the pieces of the puzzle do not fall into place. But we also know that ambiguity is a resource for political organizations. The less-than-Cartesian rationality of the current system of different ex ante and ex post appraisals may generate local equilibria that can be exploited within the Commission and externally. Ambiguity is organizationally acceptable if the problem is to generate local power equilibria. Evaluation, in fact, is also a frame of reference and locus of activity where the Member States, the Sec Gen, the DGs of the Commission, the EP test and constantly re-define the question of who has control over EU policy.

References

- Allio, L. (2008). *The emergence of better regulation in the European Union*. PhD, King's College. University of London.
- Barca, F. and P. McCann (2011). *Outcome indicators and targets. Towards a new system of monitoring and evaluation in EU cohesion policy*. Brussels.
- Borrás, S. and S. Højlund (2014). Evaluation and policy learning: The learners' perspective. *European Journal of Political Research* 54(1): 99-120.
- Bozzini, E. and J. Hunt (2015). Bringing evaluation into the policy cycle: CAP cross compliance and the defining and re-defining of objectives and indicators. *European Journal of Risk and Regulation* 6(1): 57-67.
- Business Taskforce (2013). *Cut EU Red Tape. Report from the Business Taskforce*. London, H M Government.
- Coglianesi, C. (2012). *Evaluating the impact of regulation and regulatory policy*. Paris, Expert paper no.1 for the OECD programme Measuring Regulatory Performance.
- Cram, L. (2001). Governance 'to go': Domestic actors, institutions and the boundaries of the possible. *Journal of Common Market Studies* 39(4): 595-618.
- European Commission (1999). *MEANS Collections – Evaluation of socio-economic programmes*. Brussels: Office for Official Publications of the European Union.
- European Commission (2000). *Focus on results. Strengthening evaluation of Commission's activities*. SEC(2000) 1051. Communication to the Commission from Mrs. Schreyer in agreement with Mr. Kinnoek and the President.
- European Commission (2001a). *Ex Ante Evaluation: A practical guide for preparing proposals for expenditure programmes*. D. Budget. Brussels: 10 December.
- European Commission (2001b). *European Governance: A White Paper*. Brussels, Com (2001) 428 final.
- European Commission (2013). *Commission follow-up to the 'Top Ten' consultations for SMEs on EU Regulation*. Retrieved from http://ec.europa.eu/europe2020/pdf/top10_en.pdf.
- European Commission (2015a). *Better regulation for better results – An EU agenda*. C. final. Strasbourg, 19 May.
- European Commission (2015b). *Better Regulation Toolbox*. Brussels.
- European Commission (2015c). *2016 Work programme of the Commission*. Retrieved from http://ec.europa.eu/atwork/key-documents/index_en.htm.
- European Commission (2015d). Commission confirms withdrawal of 73 pending proposals announced in 2015 Work Programme [Press release]. Retrieved from http://europa.eu/rapid/press-release_IP-15-4567_en.htm.
- European Commission and European Parliament (2016). *Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European*

- Commission on Better Law-Making*. Retrieved from http://ec.europa.eu/smart-regulation/better_regulation/documents/iaa_blm_final_en.pdf
- European Court of Auditors (2010). *Impact Assessments in EU institutions: Do They Support Decision-Making?* Luxembourg.
- Falkner G., O. Treib, M. Hartlapp, and S. Leiber (2005). *Complying with Europe. EU Harmonisation and Soft Law in the Member States*. Cambridge: Cambridge University Press.
- Fritsch, O., C.M. Radaelli, L. Schrefler and A. Renda (2013). Comparing the Content of Regulatory Impact Assessments in the UK and the EU. *Public Money & Management* 33(6): 445-452.
- Gravey, V. (2016). *Does the European Union have a reverse gear? Environmental policy dismantling, 1992-2014*. PhD, University of East Anglia. Retrieved from <https://ueaeprints.uea.ac.uk/59419/1/ThesisVGravey.pdf>
- Hoerner, J. and P. Stephenson (2012). Theoretical perspectives on approaches to policy evaluation in the EU: The case of cohesion policy. *Public Administration* 90(3): 699-715.
- Højlund, S. (2015). Evaluation in the European Commission: For accountability or learning? *European Journal of Risk and Regulation* 6(1): 35-46.
- Kassim, H. (2008). 'Mission impossible', but mission accomplished: the Kinnock reforms and the European Commission. *Journal of European Public Policy* 15(5): 648-668.
- König, T. and B. Luetberg (2009). Troubles with transposition? Explaining trends in member-state notification and delayed transposition of EU directives. *British Journal of Political Science*, 39: 163-194.
- Mastenbroek, E., S. van Voorst and A. C. M. Meuwese (2016 in press). Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy*.
- Meuwese, A. C. M. (2009). *Impact Assessment in EU Lawmaking*. The Hague, Kluwer Law International
- Molle, W. (2008). *European Cohesion Policy*. London, Routledge.
- OECD (2014). *Framework for Regulatory Policy Evaluation*. Paris, OECD Publications.
- Pattyn, V. (2014). Why organizations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation* 20(3): 348-367.
- Pattyn, V., S. van Voorst, E. Mastenbroek and C. A. Dunlop (2017). Policy Evaluation in Europe. *The Palgrave Handbook of Public Administration and Public Management in Europe*. E. Ongaro and S. van Thiel. Palgrave.
- Radaelli, C. M. (1999). Steering the Community regulatory system: the challenges ahead. *Public Administration* 77(4): 855-871.
- Radaelli, C. M. (2007). Whither better regulation for the Lisbon agenda? *Journal of European Public Policy* 14(2): 190-207.

Radaelli, C.M. (2017). European Union policies and the Europeanization of domestic policy-making. In H. Heinelt and S. Munch (Eds.) *Handbook of European Policy: Formulation, Development and Evaluation*.

Radaelli, C. M. and A. C. M. Meuwese (2010). Hard Questions, Hard Solutions: Proceduralisation through Impact Assessment in the EU. *West European Politics* 33(1): 136-153.

Schömann, I. (2015). *EU REFIT machinery 'cutting red tape' at the cost of the acquis communautaire*. Brussels, ETUI Policy Brief, 2015. Retrieved from <https://www.etui.org/Publications2/Policy-Briefs/European-Economic-Employment-and-Social-Policy/EU-REFIT-machinery-cutting-red-tape-at-the-cost-of-the-acquis-communautaire>.

Schrefler, L. and S. Huber (2015). *Evaluation in the European Commission: Rolling Check-List and State of Play*. Brussels, European Parliament, EPRS, Policy Cycle Unit.

Schrefler, L., G. Luchetta and F. Simonelli (2015). A new tool in the box? The cumulated cost assessment. *European Journal of Risk and Regulation* 6(1): 68-78.

Shadish, W. R., T. D. Cook and L. C. Leviton (1990). *Foundations of Program Evaluation: Theories of Practice*. London, Sage.

Smismans, S. (2015). Policy evaluation in the EU: The challenges of linking ex ante and ex post appraisal. *European Journal of Risk and Regulation* 6(1): 6-26.

Stephenson, P. (2015). Reconciling audit and evaluation? The shift to performance and effectiveness at the European Court of Auditors. *European Journal of Risk and Regulation* 6(1): 79-89.

The Evaluation Partnership (2007). *Evaluation of the Commission's Impact Assessment System. Annexes to the Final Report*. Richmond, The Evaluation Partnership.

Turnpenny, J., C. M. Radaelli, A. Jordan and K. Jacob (2009). The policy and politics of policy appraisal: Emerging trends and new directions. *Journal of European Public Policy* 16(4): 640-653.

Van den Abeele, E. (2014). The EU's REFIT strategy: a new bureaucracy in the service of competitiveness? European Trade Union Institute. Retrieved from <https://www.etui.org/Publications2/Working-Papers/The-EU-s-REFIT-strategy-a-new-bureaucracy-in-the-service-of-competitiveness>.

Van den Abeele, E. (2015). *'Better regulation': a bureaucratic simplification with a political agenda*. Brussels, European Trade Union Institute. Retrieved from <http://www.etui.org/Publications2/Working-Papers/Better-regulation-a-bureaucratic-simplification-with-a-political-agenda>.

van Golen, T. and S. van Voorst (2016). Towards a regulatory cycle? The use of evaluative information in impact assessments and ex-post evaluations in the European Union, *European Journal of Risk and Regulation* 7(2): 388-403.

Versluis, E., M van Keulen and P. Stephenson. (2011). *Analyzing the European Union Policy Process*. Palgrave Macmillan.

Table 33.1 - Types of evaluations

	Expenditure/Programs	Regulatory/Legislative
Ex ante	Ex-ante analysis of expenditure	Impact Assessment (IA)
Ex post	Evaluation of European structural and investment funds (*). Evaluation of other programs e.g Research, Innovation, Media and Culture	Ex post legislative/regulatory evaluation, REFIT (including Fitness Checks and cumulative cost assessments)

Source: Authors, 2016

(*) This category includes cohesion policy. Cohesion policy is an umbrella term. It gathers all the projects funded by the European Regional Development Fund, the European Social Fund, and the Cohesion Fund. See http://ec.europa.eu/regional_policy/en/faq/#1 Molle (2008) and Hoerner and Stephenson (2012).