

Ex-post legislative evaluations in the European Union

Questioning the use of evaluation as instruments for accountability

Paper prepared for the ECPR 5th Biannual Conference on Regulatory Governance, Barcelona,
25-27 June 2014.

Pieter Zwaan, Stijn van Voorst, and Ellen Mastenbroek¹

Please do not quote!

Abstract

Evaluations may perform a key role in a political system, as they provide an information basis for accountability to operate. This is equally the case in the European Union. Yet, several factors work against the use of evaluations. Members of European Parliament work under great time pressure and have to prioritize. Executives may not have an incentive to produce high-quality, objective evaluations, given the undesirable consequences this may have. This article therefore addresses the question to what extent and under what conditions the European Parliament uses ex-post legislative evaluations for accountability purposes. We present an analysis of a new dataset comprising 160 evaluations, studying how many of them were followed up through parliamentary questions. Our main finding is that about 25% of the evaluations is followed up by parliamentary questions. Variance in follow-up is not yet fully explained, but is likely to be influenced primarily by political consideration instead of usability.

¹ The three authors work at the department of Public Administration at the Radboud University Nijmegen, in the capacity of assistant professor (Pieter Zwaan), junior researcher (Stijn van Voorst), and associate professor (Ellen Mastenbroek). Queries can be directed to p.zwaan@fm.ru.nl.

1. Introduction

In response to the concerns about the democratic deficit of the EU and the lack of accountability of its institutions that were raised in the debate on the EU's legitimacy in the 1990s, increasing attention has been given to the accountability of the European Commission. When the Santer Commission had to resign in 1999 because of alleged fraud, this increasing attention was followed by action: the succeeding Prodi Commission implemented serious reforms to increase accountability of the Commission (Wille, 2010: 194).

In order to increase the Commission's *political* accountability, the Commission has been increasingly linked to the European Parliament (EP). Various institutional reforms resulted in a range of powers and instruments at the disposal of the EP to hold the Commission to account (see below). The increasing powers were followed amongst others by stronger demands on the Commission to send reports and studies (about EU policies) to the EP (Curtin, 2009: 256-57). Especially promising in this respect are ex-post evaluations of EU legislation and policies. Most legislation adopted nowadays includes a requirement for a mid-term or ex-post evaluation, focusing either on implementation process or the actual impact of EU legislation. In theory, these evaluation reports provide rich information on the fulfillment of policy goals and responsibilities of involved actors (Corbett et. al., 2011: 318-19).

So far, we know very little about the usage of ex-post evaluations in the EU: "even though evaluation at the EU level was introduced earlier than in many member states" as Hoerner and Stephenson (2012: 699) note, "academic analysis of evaluation remain rather scarce". This is true in particular for the relationship between evaluations and accountability. While evaluations have been identified at the EU level as a tool to enable accountability (see also below), the little existing research on EU evaluation indicates that evaluations are used quite

differently: studies by the Commission's DG Budget for example, suggest that evaluations are mainly used by the officials directly involved in the implementation of the interventions that are evaluated..." (EPEC, 2005 in Stern 2009). Other studies suggest that the Commission uses evaluations mainly to support its own decision making and strategies: this information can help the Commission consecutive policy cycles "to improve agenda setting and policy drafting..." (Bauer, 2006: 723).

Literature on EU evaluations thus suggests that ex-post evaluations, whereas theoretically useful for accountability practices, in reality play a limited role in this respect. Therefore, this article assesses the question *to what extent and under which conditions the European Parliament uses ex-post evaluations to hold the Commission accountable*. To answer this question we analyze the extent to which Members of the EP (MEPs) ask questions either about or based on ex-post evaluation reports produced or commissioned by the European Commission. We do so for a period of roughly three parliamentary terms from 1999-2014. By doing so, we not only seek to add to the literature on EU evaluations; we also seek to develop the quantitative knowledge base for understanding EU accountability, which is currently rather weakly developed in the literature (see Brandsma, 2013).

2. EP-Commission accountability relationship

The concept of accountability was once used mainly in the financial world but has made its entrance in political discourse in recent years. Nowadays, the concept is considered a key aspect of good governance that nobody can be against, even though it is unclear what it means exactly (Bovens, 2007: 448; Curtin, 2009: 246; Ebrahim, 2003: 193). Bovens (2010: 947-948) makes a distinction between authors who view accountability as a virtue (a trait of individual actors) and

authors who view it as a mechanism or procedure that structures the relation between two or more actors. The second view is followed by most authors who deal with the European Union. For example, both Curtin (2009: 248, 253; Curtin et. al., 2010:938) and Vedung (1997: 02) define accountability as a social relation in which an agent is held to account for his actions to a principle. This view is also adopted in this article.

Accountability, from this point of view, is an important mechanism, which allows a principle to deal with the so-called problem of ‘moral hazard’, i.e. that there is no guarantee that the agent acts in line with the principal’s interests or does so in an effective way (Blom-Hanssen 2005: 631). Through accountability the actions of the agent can be controlled. According to Curtin (2009: 257) a proper accountability process requires the following steps: 1) there must be sufficient information about the fulfillment of responsibilities by the agent; 2) there must be a debate in which the agent held accountable is given the chance to explain his actions; and 3) there must be the possibility for the principle to sanction or reward the agent in order to steer its behavior.

The possibilities for the EP to hold the Commission accountable during each of the different steps have increased over the last decade. The importance of holding the Commission accountable is also reflected in the increasing involvement of individual MEPs in holding the Commission accountable: “over the past decade, more questions are being posed, more debates held, and there are more committees occupying the time of many more MEPs, receiving the written answers or oral policy statements of the Commission.” (Wille 2010: 60). The next section discusses the different instruments available to the EP, starting with the latter step: the possibility for the EP to sanction or reward the Commission.

Sanctioning and rewarding

The Commission is selected by the heads of government and state of the EU member states. Its appointment, however, is subject to a vote by the EP. The political system of the EU in this respect has some elements of a parliamentary system, but also of a separation of powers that is more typical of a presidential system. This separation of powers between the Commission and Parliament affects the possibilities for the EP to steer the Commission through rewards or sanctions. For example, this means that there is no such thing as a vote of no confidence for individual Commissioners as a sanctioning mechanism. The EP has the power to censure the Commission as a whole, yet the supermajority requirement makes this instrument very hard to use. The option for the EP to steer the Commission is therefore mainly by threatening to hold back budgets. A less hurtful measure to do so is by drawing up resolutions or parliamentary questions in which it calls upon the Commission to take action.

Debate

Before judging the agent, the principal must consider the agent's point of view. Stufflebeam and Shinkfield (2007: 163) state that accountability requires considering information in its relevant context; this means there must be a debate in which the actor that is held accountable is given the chance to explain his actions. In the EU, this occurs for example when Parliament asks Commissioners to appear before its committees (Curtin, 2010: 258-61). EP committees not only have legislative tasks, but scrutiny tasks as well. Commissioners can be asked to attend meeting of EP committees to answer questions (Corbett et. al., 2011: 319-20). Commissioners can also be requested to come and make a statement in the plenary, after which there is a debate (Judge and Earnshaw, 2008: 222-224). Parliamentary questions (besides giving information, see below)

provide an alternative forum for the agent to explain its actions. In particular so-called “questions for oral answer with debate” and “question time” that is organized at each plenary session of the EP allow for discussion with and considering information of the Commission.

Information

Openness and transparency, are considered the hallmark of accountable government (Curtin et. al., 2010: 936). Accountability requires a baseline of information to demonstrate the fulfillment of responsibilities (Stufflebeam and Shinkfield, 2007: 163): it is on that basis that debates take place, or rewards or punishments are awarded through which the agent’s behavior is steered.

In general, legislatures have an information deficit as they have less expertise and resources than the executive (Brandsma, 2013: 4). Partly they can compensate their information deficit by acquiring more general political information about the preferences of the executive. This kind of information allows the legislature to judge the trustworthiness of the other information it receives from the executive. In the case of the EU, the information asymmetry between the legislature and executive is even bigger because the executive is not directly linked to the legislature through the ties of political parties as is the case in national political systems. As a result, information exchange between the legislature and executive takes place in a much more formalized way, for example through parliamentary questions or through the attendance of EP committee meetings (Proksch and Slapin: 2010).

However, the EP increasingly has successfully to gain political information by strengthening its general rights for information (Brandsma, 2013: 5). For this reason, the Commission is required to send myriad reports and studies to the European Parliament, amongst which evaluations (Curtin, 2009: 256-57). In the inter-Institutional Framework Agreement (IFFA) of

2010, which describes the responsibilities that the Commission has towards the European Parliament², a number of general principles on the exchange of information is provided. It includes for example, the general agreement that the EP is treated equally to the Council with regards to the information it gets (paragraph 9).

The IFFA also contains a specific paragraph on impact assessments (paragraph 42), stating that “the Commission shall ensure that its impact assessments are conducted under its responsibility by means of a transparent procedure which guarantees an independent assessment. Impact assessments shall be published in due time (...) and shall in principle be presented to the relevant parliamentary committee during the phase of the provision of information to national parliaments...”. There is no such paragraph for ex-post evaluations in the IIFA 2010, although the Commission has stated in its Policy Paper on Smart Governance (2013) that its services publish all ex-post evaluations, regardless of the author.

3. Evaluations and accountability

Evaluations can provide a useful tool for principals to gain information about the behaviour of their agents and the fulfillment of their responsibilities (Stufflebeam and Shinkfield, 2007: 163). Accordingly, enabling accountability is generally considered one of the three substantive aims of ex-post and midterm evaluations in the public sector- next to learning how to improve existing policy or generating basic knowledge of a policy’s intervention logic (Vedung, 1997: 101; Benjamin, 2008: 325).

The EP and the Commission also increasingly stress the accountability function of evaluations. The European Parliament has stressed the need of ex-post evaluations for

² This document is a good example of a negotiated agreement between European actors that have mostly served to improve the position of the European Parliament without having to change the European treaties (Brandsma, 2013: 6).

accountability purposes in its 2001 resolution (EP: A5-0399/2001): the Commission need to be more transparent and to prove its worth to the European public. The need for the EP to better scrutinize the executive is also stressed in the "European Parliament in 2025" paper by the EP's Secretary-General (2012: 26) and in a so-called "library briefing" for the European Parliament by Poptcheva (2013: 1), who stresses the need for ex-post evaluations to improve accountability.

The European Commission originally presented evaluations as a tool to increase EU policy effectiveness (European Commission, 2001: 10). Gradually, though, the Commission started stressing the potential of evaluation for enhancing accountability. The Commission's Evaluation Staff Guide (European Commission, 2004: 9) describes accountability as one of the several goals of evaluation, besides contributing to the design of interventions, assisting in the efficient allocation of resources and improving the quality of the intervention.³ In its 2007 Communication the Commission clearly stressed that evaluation "enhances the legitimacy of decisions and the accountability of decision-makers: (...) where evaluation results are communicated properly, they enhance transparency and democratic accountability." In the Commission Evaluation Staff Guide (European Commission, 2010), accountability - defined as "provid[ing] information on what has been achieved and at what cost" - however, was named as the primary goal of ex-post evaluations.⁴ Also in its recent Policy Paper on Smart Regulation, the Commission stressed the potential for ex-post evaluations for improving accountability (2013: 1).

³ The guide states that accountability is the primary goal of ex-post evaluations (EC, p. 13), while for Impact Assessments the emphasis is on improving the intervention.

⁴ Annex C of the guide (p. 75) further describes what this information should be about: an ex-post evaluation should assess causality and make clear if the criteria of effectiveness, efficiency, relevance and sustainability have been met in a certain intervention. The Commission recommends external evaluations for the purpose of accountability, as they would be more independent (p. 14). If accountability is the aim, the judgement should be entirely with the evaluator and not with the steering group (p. 89).

This changing view on evaluations is reflected in the objects of evaluation studies conducted. Originally, the evaluation of EU policies started off largely on a sectorial basis in the 1980s, with a focus on the EU (re)distributive policies such as the EU Regional Policy (Structural Funds) (Stern 2009). Since the late 1990s a more general interest in evaluation has grown due to an increasing focus on budgetary stringency and effective program execution and accountability (EC 2007, Annex 1; cf. Bauer 2006). As a result, the Commission has made great investments in its evaluation system. Nowadays, also regulatory policies are being evaluated, although it should be noted that most attention of the Commission has been given to ex ante evaluations or Impact Assessments (IA). Even though the Commission has made repeated promises to step up their numbers, ex-post legislative evaluations are less common. Yet, earlier research (Mastenbroek and Van Voorst, 2013) has shown that 31% of the EU's major directives and regulations have been evaluated ex-post.

4. Theory and hypotheses

The literature provides a wide variety of explanations for variance in use of evaluation studies. These include practical, technical, financial, institutional, political as well as psychological reasons (McDavid and Huse, 2012: 12). In order to provide an explanation of the use of ex-post evaluations by the EP for accountability purposes we formulate a number of hypotheses, respectively inspired by a political and rationalist view of evaluations.

The political view

Our first set of hypotheses is based on *political explanations*. This view is based on the understanding of accountability as a relationship in which one party, the principal, has delegated a task to another party- the agent. Earlier literature has argued that the nature of this relationship influences how evaluation data is being used (Benjamin, 2008: 335). We therefore expect that the way MEPs engage in accountability and use evaluation data for this purpose depends heavily on its political relationship with the Commission.

(Political) risks play an important role in accountability relationships. By delegating the task to an agent the principal faces the risk that the agent may not deliver, i.e. the problem of moral hazard. For this reason, an agent must be answerable to its principal. Benjamin (2008: 336) argues that this risk contains a number of aspects that have an effect on the exact accountability relationship between the principle and the agent.

First of all, the accountability relationship is affected by the *significance of the principal's risk*: the question here is how important it is for the principle if the agent doesn't do what it wants it to do? The greater the significance of the principle's risk, Benjamin (ibid.) argues, the greater the need for holding the agent accountable. For the EP-Commission accountability relationship It can be expected that the significance of this risk for the EP is greater when they feel more responsible for a EU policy. This is the case for codecision measures. On that basis we expect the following:

H1: Evaluations of co-decision measures are more likely to be referred to by MEPs for accountability purposes than consultation and cooperation measures.

Secondly, the accountability relationship is affected by *the nature of the principal's risk*. According to Benjamin (2010) the question here is whether there is a possible conflict of interest between the principal and agent. That the EU has a strict separation of powers - with the Commission not depending on the support of a clear majority of parties in the EP (Jacqué, 2004; Kreppel, 2011) - is important in this respect.

It results in a situation in which the EP can act relatively independently from the Commission and can pursue its own interests. It has been observed that while the means of the EP to hold the Commission accountable are more limited than those of national parliaments, the EP does make greater use of it. When it comes to the relationship between the EP and the Commission we can expect that both have their own interests, which may conflict. This leads to our second hypothesis:

H2: The higher the conflict during decision-making, the higher the chances that an evaluation of the resulting legislative measure is referred to by MEPs for accountability purposes.

The usability view

The basic idea behind our second set of hypotheses is that the use of an evaluation is dependent on the *usability* of the report. Our second set of hypotheses thus concerns the characteristics of the evaluation itself. The literature has argued that specific evaluation characteristics improve usability. Below we turn to a number of characteristics, such as the type of evaluation, the presence of a concise executive summary, methodological quality, and independence of the report.

To begin with, different audiences and activities require different *type of evaluations* (McDavid and Huse, 2012). When it comes to using evaluations for accountability purposes, literature suggest that evaluations must be aimed at finding out if a policy has actually worked, rather than how it can be improved. Lehtonen (2005: 170-1) has argued that evaluations meant for accountability must be useful in a more direct way than evaluations aimed at learning: holding someone accountable needs to happen immediately. Stufflebeam and Shinkfield (2007: 161) make a similar point when stating that accountability evaluations need to focus on outcomes. Typically, this is the case for *product* evaluations, which focus on impacts, outcomes or effects, whereas *process* evaluations focus on the transposition, implementation, or enforcement of EU measures in the member states.

H3: Product evaluations are more likely to be referred to by MEPs for accountability purposes than process evaluations.

The EP's "accountability task" often competes with other activities. In addition, evaluations often compete with other sources of information (Weis, 1993 (1973): 94): there is simple to limited time and information that can be processed meaningfully by MEPs and their staff (Simon, 1972). We know that the EP often gets overloaded with the information by the Commission; as a result, it has a hard time deciding what is important and what is not (Linter & Vaccari, 2005: 23). We therefore expect that evaluations, in order to be referred to, must not only have a clear accountability focus, but also must be readable in order to be referred to by MEPs for accountability purposes. In particular, this means that evaluations need to have concise executive summaries. This lead to the following expectation:

H4: Evaluations with a concise executive summary are more likely to be referred to by MEPs for accountability purposes than those without a concise executive summary.

When deciding whether or not to use an evaluation, actors use different tests. One of the tests that is used is the so-called “truth test” (Weis and Bucuvalas, 1981 cited in Cooksy and Caracelli, 2005: 31), which involves considering the extent to which the evaluation can be challenged based on the research quality. A critical aspect when it comes to the research quality is the research methodology. Questionable quality, is likely to reduce usability of a report.

H5: The higher an evaluation’s methodological quality, the higher the chances that it is referred to by MEPs for accountability purposes.

A related aspect affecting usability of an evaluation is the independence of the study. Several authors (Rutman, 1990; Vedung, 1997; Stufflebeam and Shinkfield, 2007) have argued that evaluations meant for accountability should be external evaluations, as external researchers are likely to be more independent than internal ones. The literature on the use of EU evaluations suggests that the independence of evaluations is considered as part of this test by MEPs before deciding to refer to it. Stern (2009: 70-2) stated that there is a widespread belief in the evaluation community that the Commission advocates objectivity in evaluations on paper, but doesn’t appreciate it in practice. In practice, evaluation is mostly a tool to support its own strategies. The very little use that has been made of IAs by the EP is telling in this respect. While IA can provide the EP with a powerful tool to gain information and improve its control over the Commission

(Meuwese, 2006), they are hardly being used during the policy making process. Poptcheva (2013: 4-5) found that out of 12.000 EP Committee documents in the 2004-09 parliamentary term, only one document made explicit reference to a Commission IA. A study by the Court of Auditors and an Evaluation of the Commission IA system led to similar conclusions. Amongst the reasons being mentioned of why the EP makes little use of IAs is the EP's distrusts in the objectivity of the information provided by the Commission. On that basis we formulate the following hypothesis for the referral to ex-post evaluations:

H6: External evaluations are more likely to be referred to by MEPs for accountability purposes than external evaluations.

5. Data and methods

The hypotheses above are tested a dataset on EU ex-post evaluations (Mastenbroek & Van Voorst, 2013). The dataset contains 160 ex-post legislative evaluations, which were gathered from multiple sources: DG websites, Commission evaluation documents, Commission work programmes, the Commission evaluation search engine, and systematic Google searches. For all these evaluations, we searched for corresponding EPQs, using the EP's website on parliamentary questions and declarations.⁵ For three parliamentary terms we searched for EPQs that referred to an evaluation using a Boolean search term that was adapted for every single evaluation⁶. The EPQs that resulted from this search were read so as to make sure that reference was indeed made to the evaluation included in our dataset.

⁵ <http://www.europarl.europa.eu/plenary/en/parliamentary-questions.html#sidesForm>

⁶ The search term was: "(evaluation OR study OR report OR review OR assessment) AND ([legislation number] OR [key-words from evaluation report])."

Dependent variable

There are three main procedures for asking questions to the Commission (Corbett et. Al., 2011, p. 316-18): 1) asking written questions, 2) asking a question during ‘Question time’ of 90 minutes in each plenary session; and 3) asking an oral question followed by a debate. EPQ’s have been recognized in literature as one of most visible instruments that *can* be used by the EP to ensure accountability (Proksch and Slapin, 2010: 60; Sozzi, 2011: 1; Wille, 2010: 60).

EPQs can of course perform different functions. In literature a variety of functions is referred to. Three main functions can be distinguished: requesting for information, gaining publicity and holding the executive to account for its actions. Literature on PQs in national democracies questions the extent to which PQs are used for purposes of accountability; according to this literature PQs are most often asked by MPs for attention and re-election. However, different scholars have suggested that this is different for the EP given the second order nature of EP elections and little media attention. As a result MEPs cannot use questions for personal publicity or in order to gain votes (Proksch and Slapin, 2010: 60; Sozzi, 2011: 1). Accountability and information are therefore a much more important function.

Earlier research yet, has also shown that accountability often mixes with other parliamentary motivations in particular scrutiny processes. Given the centrality of policy-seeking motives, MPs may go beyond demonstrating an executive’s shortcomings, and try to influence the executive, for instance by summoning it to change existing legislation. Given this close empirical link between retrospective and prospective parliamentary scrutiny (Wille 2010), we have adopted a broader understanding, in line with the existing literature on quantitative measurement of accountability that also includes these questions (Brandsma, 2013). The independent variable was measured as a dichotomous variable (referred to and not-referred to).

Independent variables

Our independent or explanatory variables have been operationalized as follows. Co-decision has been measured by assessing whether or not the measure evaluated was adopted through co-decision. The nature of the risk for the EP was measured by counting the number of amendments that have been proposed by the EP on a Commission's proposal. This variable thus is a continuous variable. To measure the function of the evaluation, two categories were created: product evaluations versus process evaluations. To assess the presence of a concise executive summary, we counted the number of pages of a report. If a report was longer than 10 pages, we checked whether an executive summary of maximum ten pages was available. To establish methodological quality we analyzed eight aspects of the evaluation reports: the presence of a clear operationalization, the presence of a clear research problem, an explanation of the methods used, the presence of questionnaires and list of respondents for possible reproduction, the use of more than one research method (triangulation), an explanation of the selection of countries, an explanation of the selection of cases and a response rate for surveys and interviews of more than 50% (see Mastebroek and Van Voorst, 2013). To analyze whether an evaluation was carried out internally by the Commission or carried out by external researchers, we consulted the title pages of the evaluation reports gathered. The different variables and their expected effect on the referral to evaluations in EPQs is summarized in Table 1 below.

Table 1 Overview of independent variables

Variable	Indicator	Measurement	Expected effect on number of EPQ's
Political view			
Co-decision	Not relevant	0 = consultation procedure 1 = codecision procedure	+
Nature of risk	Number of amendments	Continuous variable	+
Usability view			
Product evaluations	Not relevant	0 = process evaluation. 1 = product evaluation.	+
Concise executive summary	Number of pages of executive summary of max. 10 (for reports over 10 pages long)	0 = evaluation is longer than ten pages and does not have a summary of less than ten pages. 1 = evaluation is shorter than ten pages or contains a summary of less than ten pages.	+
Methodological quality	Composite measure of quality	Continuous variable, ranging from 1-8	+
External evaluations	Not relevant	0 = internal evaluation 1 = external evaluation	+

Method of analysis

Since our dependent variable is dichotomous (ex-post evaluations referred to and not-referred to in EPQ), and our independent variables are either categorical or continuous, we chose to use a binary logistic regression analysis (Hosmer and Lemeshow, 1989) to test our hypotheses.

External validity

Finally, some words about external validity are in place. This study is particularly aimed at understanding the functioning of accountability in the EU, which seems warranted given its very accountability deficit. Yet, it is thought to have broader relevance. That is, even though the EP has less formal powers than some national parliaments, it seems, as stated above, more willing to

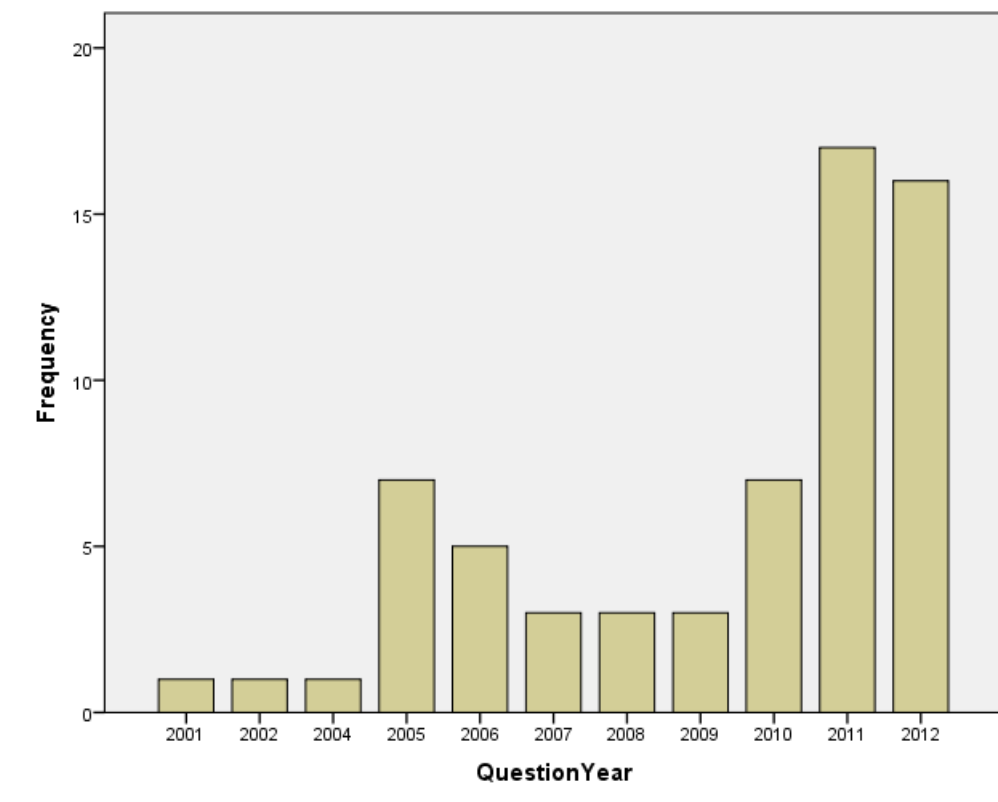
use them (Scully, 2010, p. 167). As the EU has a much stricter separation of powers than most national parliamentary systems, the EP behaves relatively independently in holding the Commission accountable. Therefore, it seems a very likely case for evaluations to be followed up for accountability purposes.

6. Analysis

We first provide some descriptive results of our analysis of EPQs. 26 of the 160 ex-post evaluations included in the dataset were referred to in EPQs. More specifically, 15 evaluations were referred to once, 5 were referred to twice, 2 were referred to thrice and one was referred to four times. Three cases appeared much more than usual: the first report from the Commission on the application of the Tobacco Products Directive from 2005 (seven times), the animal welfare legislation evaluation from 2010 (ten times) and the evaluation of the data retention directive from 2011 (seventeen times).

In total, 69 questions were found referring to ex-post legislative evaluations. Figure 1 below shows the number of EPQ's referring to evaluations per year. While the peaks at 2005, 2010 and 2011 can be explained by the three specific evaluations mentioned above, there still seems to be an upward trend, with only one question found from the first couple of years and a particularly large number of questions found from 2011 on.

Figure 1 Distribution of EPQs referring to evaluations over time



Of the 71 questions, 54 questions were "written questions". 15 questions were "oral questions during question time". Two questions were asked as an "oral question with debate".

Table 2 below shows the number of EPQ's referring the evaluations asked by every political group in the European Parliament. As would be expected, the larger groups in Parliament have asked more questions, although the table also shows that the S&D faction refers to evaluations in its questions relatively more than the other major groups in the EP.

Table 2 Distribution of EPQs over the political groups

Political group	Number of EPQs
S&D	19
EPP	15
ALDE	10
Greens/EFA	9
EFD	5
ECR	2
GUE/NGL	2
Not associated	5

S&D = Socialists and Democrats, EPP = European People’s party, ALDE = Alliance of Liberals and Democrats, Greens/EFA = Greens/European Free Alliance and its predecessor Verts/ALE, EFD = Europe of Freedom and Democracy and its predecessor UND, European Free ECR = European Conservatives and Reformists, GUE/NGL = Conferderal Group of the European United Left/Nordic Green Alliance.

Concerning the content of the questions asked, we identified four broad types of questions. The first and largest group we identified includes questions about *evaluation follow-up*: what has the Commission done or is the Commission doing based on the results of a report? The following question by a large group of MEP’s from the ALDE group on the Data Retention Directive illustrates this type of question:

“In the light of its evaluation report on the Data Retention Directive (Directive 2006/24/EC), the Commission declared on 18 April 2011 that it will propose a

revision of the current data retention framework. In its Action Plan implementing the Stockholm Programme the Commission announced a proposal for revision for 2012.

When will the Commission present its proposal?”

The second group of questions is about *evaluation content*: why are certain topics included, excluded or treated in a specific way in the evaluation report? The following question about the Data Retention Directive, asked by green MEP Carl Schlyter, is an example of such a question:

“In the context of the current evaluation and impact assessment of the future options with regard to the Data Retention Directive (2006/24/EC), is it true that the Commission is not assessing the option of an EU-wide ban on blanket communications data retention in favor of a system of expedited preservation and targeted collection of traffic data, as agreed on in the Council of Europe's cybercrime convention?”

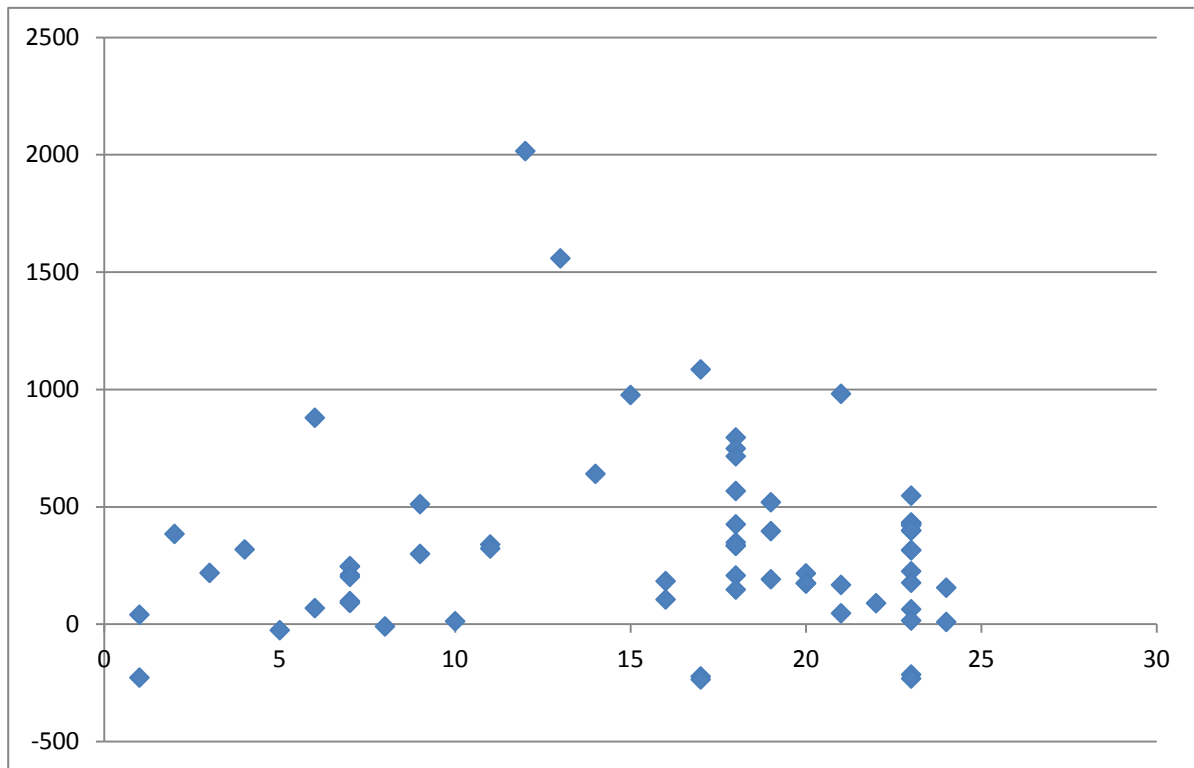
The other type of questions are more related to the evaluation process. The third group of questions is about *evaluation timing*: is the Commission planning an evaluation about a piece of legislation and/or when would this evaluation take place? The fourth group is finally is about *evaluation stakeholders*: why were certain actors included or excluded during the evaluation process?

Out of these four categories, *evaluation follow-up* was the largest category by far: 49 questions referred to the Commission taking action based on a report. Only for 4 out of 26 evaluations that were referred to in EPQs this category was absent. This indicates a forward-looking rather than a backward-looking attitude of the EP when it comes to making use of ex

post evaluations. Only 14 questions were actually about the evaluation content, 6 studies were about stakeholder involvement and just 2 questions were about evaluation timing.

The analysis of the moment of asking questions on the basis of ex post evaluations shows that out of the 69 questions, 17 questions were asked within ½ year after publication of the evaluation. 45 questions referring to ex post evaluation were being asked within 1 year. Almost all questions, 61, were asked within 2 years. The figure below shows the number of questions that we asked on the basis of a particular ex post evaluation (horizontal axis) and the time (days) that was taken to ask a question.

Figure 2 Timing of EPQs following publication of the evaluation report



Turning to the explanatory analysis, we first estimated a model only containing the two political variables. As the Chi-square statistic of 4,754 indicates, the model containing only the political variables is not significant. This also becomes clear from looking at the estimated effects for the variables of co-decision and the number of amendments. Whereas the sign of the effect is positive for both variables, as effect, none of the effects is significant at the 0.05 level.

Unfortunately, the model's fit does not improve after adding the four usability variables, as evidenced by the Chi-square value, which is not significant at the 0.05 level. Also, it is striking that some predicted effects actually point into another direction than expected: product evaluations and higher quality evaluations are less likely to be followed up by MEPs. The expectations for the variables concise executive summary and external evaluation are as expected, even though none of them are significant.

Table 3 Results of the logistic regression

	Political model			Full model		
	B	Sig	Exp(B)	B	Sig	Exp(B)
Constant	-2,41	,00	1,76	-1,79	0,03	0,17
Codecision	0,57	,24	1,01	0,57	0,26	1,77
Amendments	0,01	,08	0,09	0,01	0,10	1,01
Product evaluation				-0,53	0,29	0,59
Methodological quality				-0,19	0,24	0,83
Concise executive summary				0,31	0,54	1,37
External evaluator				1,00	0,85	1,10

Yet, it should be noted that among the parliamentary questions, two cases actually consisted of critical questions *about the quality* of an evaluation. If we remove these two atypical cases from the analysis, the outcomes change somewhat. Even though the political model still is not significant, the effect of amendments changes, getting close to significance. The other parameter estimates are fairly robust to this change, the direction of the effects does not change.

In order to gauge the fit of the resulting full model, we identified those cases with a studentized residual greater than 2. In all these cases, the predicted chances of an evaluation being followed up were slight, whereas in reality a parliamentary question was asked about the evaluation. To obtain more insights into these cases, we list these outliers in Table 4. It becomes clear from these table that most of these outliers are highly politicized topics, such as family reunification, airport noise, drug policy. Other topics seem to include cases with strong lobbying: animal welfare and cross-compliance in the field of agriculture. This analysis sustains the careful suspicion arising from our statistical results that evaluation follow-up is primarily a matter of political factors, rather than rather technical usability aspects. Further analysis is required to further enquire into this suspicion.

Table 4 Overview of outliers

Case ID	Title of evaluation report	Studentized residual
13	Report from the Commission to the European Parliament and the Council on the application of Directive 94/80/EC on the right to vote and to stand as a candidate in municipal elections by citizens of the Union residing in a Member State of which they are not nationals.	2,39418
33	Animal Welfare Evaluation	2,36040
43	Directive 2003/86/EC on the right to family reunification. Synthesis report.	2,29305
50	Study on the implementation of the regulation 1435/2003 on the statute for the European cooperative society (SCE)	2,20970
107	Noise Operation Restrictions at EU Airports (Report on the application of Directive 2002/30/EC)	2,14301
117	Study on the implementation of Directive 94/62/EC on packaging and Packaging Waste and Options to strengthen Prevention and Re-use of Packaging	2,13395
148	Evaluation of EC North South Cooperation on drugs	2,09624
158	Evaluation of the application of cross compliance as foreseen under Regulation 1782/2003	2,02994

7. Conclusion

To analyze the value of ex-post evaluations for accountability, this article has studied the effect of ex-post evaluation characteristics and political context factors on the extent to which MEPs refer to evaluations in EPQs. Although literature and EU institutions have stressed the potential for ex-post evaluations in holding the Commission to account, little is known about the extent to which this happens and the factors that affect this.

Drawing on a measurement of EPQs that explicitly refer to ex-post evaluation, we developed a new measurement for accountability activities and the importance of ex-post evaluations in this process. Our empirical analysis shows that out of 160s evaluations 26 evaluations were being referred to. About 16% of ex-post evaluations hence, is being followed

up by MEPs and used for other purposes that merely assisting the Commission in its agenda setting.

Yet, our analysis seems to suggest that the EP as well is using evaluation mainly in a strategic manner to set the agenda of future policies. As much as the Commission, the EP has policy-seeking motives, and aims to go beyond demonstrating the Commission's shortcomings. The high number of EPQs that can be described as *evaluation follow-up questions* and the low number of *questions about the findings of the evaluation* itself is an illustration of this. The fact that MEPs increasingly refer to ex-post evaluations in their questions in this respect can be interpreted as much as an increasing commitment to hold the Commission to account, as an effort to control the political agenda of the EU.

The political use of evaluation follow-up in the EP is also evidenced by the results of the logistic regression, even though we must interpret these results with great caution. Even though none of the variables are statistically significant, the variable of 'number of amendments' has the best cards for predicting evaluation follow-up. In addition, inspection of the outliers showed that the largest residuals concerned fairly political topics. This impression needs to be further pursued through additional theory-formation and data-gathering.

8. References

- Bauer, M.W. (2010) Co-managing programme implementation: conceptualizing the European Commission's role in policy execution. *Journal of European Public Policy* Vol. 13, Iss. 5, 2006
- Benjamin, L. M. (2008). Evaluator's role in accountability relationships. *Evaluation*, 14(3), p. 323-343.
- Brandsma, G. (2013a). Bending the rules: Arrangements for sharing technical and political information between the EU institutions. In: *Ripoll Servent, Ariadna and Amy Busby (eds.): 'Agency and influence inside the EU institutions', European Integration online Papers (EIoP), 17(1)*, Article 8, <http://eiop.or.at/eiop/texte/2013-008a.htm>, pp. 1-22.
- Brandsma, G.J. (2013b) 'Quantitative Research into Accountability', in M. Bovens, R. Goodin and T. Schillemans (eds.), *Oxford Handbook of Public Accountability* (Oxford: Oxford University Press).
- Bovens, M. (2007). Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal*, 13, 447-68.
- Bovens, M. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics*, 33, 946-67.
- Blom-Hansen, Jens. "Principals, agents, and the implementation of EU cohesion policy." *Journal of European Public Policy* 12.4 (2005): 624-648.
- Cooksy and Caracelli, 2005: Quality, Context, and Use Issues in Achieving the Goals of Meta evaluation. *American Journal of Evaluation*, 26(1), 31-42.
- Corbett, R. G., Jacobs, F. B. & Schackleton, M. (2011). *The European Parliament (8th Ed.)*. London: Harper.

- Curtin, D. M. (2009). *Executive power of the European Union: Law, practices and the living constitution*. Oxford: University Press.
- Ebrahim, A. (2002). Making Sense of Accountability: Conceptual Perspectives for Northern and Southern Nonprofits. *Nonprofit management & leadership*, 14(2), p. 191-214.
- European Commission and European Parliament (2010) inter-Institutional Framework Agreement
- European Parliament / Cabinet of the Secretary-General Strategic Planning Team. (2012). *The European Parliament in 2025. Preparing for complexity*. Brussels: the European Parliament.
- European Parliament (2001) European Parliament resolution on the Commission White Paper on European governance (COM(2001). Retrieved 2-3-2014, at <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2002:153E:0314:0323:EN:PDF>
- European Commission (2001). European governance: A white paper. COM(2001)428.
- European Commission (2004). *Evaluating EU activities: A practical guide for the Commission services*. Retrieved 4-4-2012, at http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf
- European Commission (2007) Commission Communication "Responding to Strategic Needs: Reinforcing the use of evaluation" - SEC(2007)213
- European Commission (2010) Commission Evaluation Staff Guide
- European Commission (2013) Policy Paper on Smart Regulation
- European Commission. (2013). Regulatory Fitness and Performance (REFIT): Results and next steps [COM(2013)685 final]. Brussels: European Commission.

- European Court of Auditors. (2010). *Impact assessments in the EU institutions: Do they support decision-making? [Special report no. 3]*. Luxembourg: European Court of Auditors.
- Hoerner, Julian, and Paul Stephenson. "Theoretical perspectives on approaches to policy evaluation in the EU: The case of Cohesion policy." *Public administration* 90.3 (2012): 699-715.
- Judge, David, and David Earnshaw. "The European Parliament." (2008).
- Jacqu e, J. P. (2004). The principle of institutional balance. *Common Market Law Review*, 41(2), 383-391.
- Kreppel, A. (2011). Looking "up", "down" and "sideways": understanding EU institutions in context. *West European Politics*, 34(1), 167-179.
- McDavid, James C., and Irene Huse. "Legislator Uses of Public Performance Reports Findings From a Five-Year Study." *American Journal of Evaluation* 33.1 (2012): 7-25.
- Lehtonen, M. (2005). OECD Environmental Performance Review Programme. *Evaluation*, Vol 11(2), p. 169–188.
- Linter, P. & Vaccari, B. (2005). The European Parliament's Right of Scrutiny over Commission Implementing Acts: a Real Parliamentary Control? *EIPASCOPE*, 1, p. 15-25.
- Mastebroek, Ellen, and Van Voorst, S. (2013). Towards better problem-solving in the European Union? Analyzing ex-post evaluation of EU legislation. Paper prepared for presentation at the NIG Annual Conference, November 28-29, 2013, Enschede, The Netherlands.
- Meuwese, A. (2006). *Informing the EU legislator through impact assessments*. Draft Paper prepared for the ECPR/CRI conference 'Frontiers of Regulation: Assessing Scholarly Debates and Policy Challenges'.

- Poptcheva, E. M. (2013). Library Briefing. Policy and legislative evaluation in the EU. Brussels: European Parliament.
 - Proksch, S.O. and Slapin, J.B. (2010), Parliamentary Questions and Oversight in the European Union, in: European Journal of Political Research, vol. 50, pp. 53- 79.
 - Rutman, L. (1990). Serving management and accountability: a conflict for program evaluation. *Evaluation Practice*, 11(1), p. 39-43.
 - Simon, Herbert A. "Theories of bounded rationality." *Decision and organization*1 (1972): 161-176.
 - Sozzi, F (2011) Questioning the Executive at European Level. <https://www.uio.no/english/research/interfaculty-research-areas/democracy/news-and-events/events/seminars/2011/papers-roma-2011/Rome-Sozzi.pdf>
- Stern, E. (2009). Evaluation policy in the European Union and its institutions. In: W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice: New directions for evaluation*, p. 67-85.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007) The nature of program evaluation theory. In: Stufflebeam DL and Shinkfield AJ (eds.) *Evaluation. Theory, Models and Applications*. San Francisco, CA: Jossey-Bass, p. 57–79.
 - Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick: transaction.
 - Wille, A. (2010). Political–Bureaucratic Accountability in the EU Commission: Modernising the Executive. *West European Politics*, 33(5), 1093–1116.
 - Weiss, Carol H. "Where politics and evaluation research meet." *American Journal of Evaluation* 14.1 (1993): 93-106.