Evaluation and Policy Learning.
The Learners' Perspective

Susana Borrás and Steven Højlund

Copenhagen Business School, Denmark

Abstract

This paper examines how evaluation induces policy learning, a question largely neglected by the scholarly literature on evaluation and policy learning. Following a learner's perspective, the paper attempts to ascertain who the learners are, and what and how learners actually learn from evaluations. In so doing, it focuses on the learners' views within the context of the evaluation framework (the set of administrative structures defining the evaluation goals and process). Taking the empirical case of three EU program evaluations, the paper examines the patterns of policy learning emanating from them. The findings are that only two types of actors involved in the evaluation are actually learning (program units and external evaluators), that learners learn different things (programme overview, small-scale programme adjustments, policy change, and evaluation methods), and that learners are generally in control of learning objectives and processes within a well-defined evaluation framework established by the European Commission.

1.   Introduction

Most democratic countries have significantly increased their emphasis on the evaluation of government interventions in the last two decades (Furubo et al., 2002). Evaluation generally has two objectives: it aims at enhancing accountability, as well as learning (van der Meer and Edelenbos, 2006, Scriven, 1991). Whereas the former emphasizes retrospective assessment in view of attaining political legitimacy of the government intervention in question (summative evaluation), the latter emphasizes improvement of the government intervention via learning (formative evaluation). Both dimensions are included in standard definitions of evaluation: "[Evaluation is the] careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of government interventions, which is intended to play a role in future, practical action situations" (Vedung, 1997). This paper looks at the second dimension of evaluation, examining how it induces policy learning[1]. 'Evaluation' is understood here as the generic name of different assessment exercises of governmental interventions (intervention programs or regulations). These are usually done in ex-ante, midterm, final or ex-post forms.

Generally there is the understanding that evaluation provides systematic knowledge about the evaluated government intervention in question, and that this knowledge constitutes an important basis for the social process of policy learning (Van der Knaap, 1995, Balthasar and Rieder, 2000, Radaelli, 2009). However, the scarce literature that looks into that has tended to address learning normatively, in aspirational terms, rather than analytically. This means that these studies tend to treat learning in generic terms, obscuring the learner as the agent of learning and disregarding different types of learning.

This paper aims at bridging the gap between the evaluation literature, on the one hand, and the policy learning literature, on the other. It does so by focusing on the empirical questions of "who is learning" and "what is being learnt and how" when government interventions are evaluated. Hence, it aims at developing an approach that puts the learner's role at the centre of evaluation, one that understands learners' learning as a socially embedded process. In particular, this paper focuses on the "evaluation framework", understood as the set of administrative structures, rules and practices that – to varying extent - define the objectives as well as the processes of the evaluation (Leeuw and Furubo, 2008). Introducing the evaluation framework into the equation of policy learning is the second contribution of this paper.

However, the evaluation framework is far from automatically determining the actual learning (what and how learners learn). Learners will always act according to their own interpretation of the evaluation framework, and will make sense of the evaluated intervention. In so doing, the learners might be learning along the lines expected by the evaluation framework, or might be

---

[1] This paper uses the term "policy learning". See section 2 for the definition and analytical levels of policy learning.

learning in different ways. This means that the way in which the learners learn is an empirical question after all.

Empirically, the paper takes the case of the European Commission and studies what is learned from three midterm evaluations of three distinct EU expenditure programmes. It also ascertains the patterns of policy learning emanating from them and who the learners are.

The study finds that two types of actors (program units and external evaluators) learn from evaluation. Learners learn different things related to program overview, small-scale program adjustment, policy change and evaluation methods. Learners in the Commission are in control of learning objectives, while the external evaluators mainly control the evaluation process. Learners in the Commission only have control over the process within a well-defined evaluation framework.

The paper proceeds as follows: the next section explores the literature on evaluation and policy learning and identifies some important gaps. Sections three and four describe the theoretical assumptions underlying the study, as well as the analytical approach and methodology. Sections five, six and seven analyse the data, and section eight concludes based on the paper's findings.

2.   Evaluation and policy learning: Multiple actors and multiple forms of learning

The literature on evaluation has been preoccupied with the use of evaluations and the conditions under which evaluations are being used (Shulha and Cousins, 1997). Naturally, for authors interested in studying the conditions for the utilization of evaluation results, the assumption is that organizational learning and the use of evaluation results are intrinsically linked (Preskill and Boyle, 2008, Radaelli and Meuwese, 2010). Several scholars have argued in favour of participatory evaluation forms that let the evaluated participate in the process of evaluation in order to increase evaluation use and learning (Patton, 1997). The current approach linking evaluation use, participation and learning is related to what some scholars have portrayed as the transition from a "traditional-objectivist" paradigm towards an "argumentative" turn in evaluation studies. The former assumes that the knowledge produced by evaluation is invariably a source of information used for changing policy programs, while the latter focuses on evaluation results and processes as a framework for dialogue within which learning takes place (Van der Knaap, 1995, Owens et al., 2004, Radaelli and Meuwese, 2010). These two approaches might seem to have quite different views on "evidence-based knowledge", not only in epistemological terms, but also in terms of how to design the process of evaluation as such (Sanderson, 2000, Hansen et al., 2012). Nevertheless, they share the acknowledgement that evaluation takes place in complex policy and organizational contexts formed by different levels of action and populated by multiple actors (Sager and Rissi, 2011).

This acknowledgement of social and institutional embededness and actors' interactions in the learning process provides the background for this paper, which builds from the few evaluation studies that look into learning. More concretely, this approach is based on two general assumptions, which our paper follows: First, the understanding, in analytical terms, that the multiple actors, who take part in the evaluation process, have in fact multiple and different views (van der Meer and Edelenbos, 2006). Multiple actors have multiple views, and this reflects on the learning process. For van der Meer and Edelenbos (2006) the multiplicity of actors has two implications: "First, not all actors will perceive a certain evaluation as equally relevant and usable. This need not be a problem in itself, but it may complicate cooperation and mutual adaptation. Second, certain insights from evaluation may not lead to learning effects at all if they do not - to some extent - fit within existing institutional patterns and distribution of competencies." (van der Meer and Edelenbos, 2006). The second general assumption is that there are different types of learning (Balthasar and Rieder, 2000). These authors (inspired by Van der Knaap, 1995) identify two types of learning that result from evaluation activities: direct learning, caused by the feed-back information from findings and recommendations in the evaluation, and indirect learning caused by stimuli and strong political arguments indirectly associated with the evaluation (Balthasar and Rieder, 2000). Their findings suggest that learning from evaluation activities will tend to be direct learning at the operational level.

Turning now to the literature on policy learning, it is worth noting that recent scholarly developments have focused on the above two key assumptions too (multiple actors and different types of learning). Following the general views on learning by James March (March, 2010), and more specifically on policy learning (Bennett and Howlett, 1992, Dunlop and Radaelli, 2013), learning is defined in this paper as the adaptation of beliefs and views that result from the learners' sense-making of past experience. Therefore, the policy learning that emanates from a specific evaluation exercise is the adaptation of that specific policy area, programme or other government intervention which results from making-sense of past experience systematically gathered through that evaluation. Our definition of learning follows specific strands in the policy learning literature that views learning in terms of reflexive and epistemic processes, rather than the outcome of bargaining or hierarchical control (van der Meer and Edelenbos, 2006, Dunlop and Radaelli, 2013).

Learning studies gained prominence in the 1980s and 1990s (Grin and Loeber, 2007), particularly with several normative approaches, all of them suggesting that learning is important. However, just like in the evaluation literature mentioned above, the empirical studies in the field of policy learning are scarce (Grin and Loeber, 2007, Freeman, 2006, Dobbin et al., 2007). According to Dunlop and Radaelli (2013), there is a general feeling of disappointment with empirical achievements in the field. One of the most influential notions in the field is the Bennett and Howlett (1992) distinction of three levels of policy learning according to what is being learned and who are the learners. This paper focuses on the first level, which is the 'government learning'

level, where the learners are state officials, learning about specific process-related issues in policy programs. The other two levels are 'lesson-drawing' and 'social learning'. In the former, the learners are policy networks learning about policy instruments. The learners on the 'social learning' level are the broad policy communities that learn about policy-related ideas. This multi-level approach has been influential in subsequent studies. However, somehow these studies have been more concerned with identifying multiple levels of learners, to the disregard of the multiple actors on each of the levels. In other words, a partial lack of attention to the fact that, on each of these levels, there are many different learners with different views and interests.

The different types of policy learning are another fundamental issue in the literature. One of the most influential typologies in the literature is the one by Mocker and Spear (1982), recently refined (see also Dunlop and Radaelli, 2013). This typology is based on the understanding that the variation of types of learning depends on the control that learners retain over the learning objectives and the learning process. Using this typology to analyse the evaluation framework, this paper expands our understanding of 'governmental learning' by combining a study on policy evaluation with theory on policy learning. Thereby this paper contributes to the few studies on evaluation and learning by focusing on the evaluation framework that, together with actors' sense making of past experience, defines policy learning at the governmental level.

3.   Who learns what and how from evaluations?

We are interested in studying how evaluation contributes to policy learning, and we are going to do it from a learner's perspective that takes into consideration the multi-actor nature of learning, as well as its social embeddedness.

Hence, one of the first issues to address in our analytical framework is: who are the potential learners from an evaluation? In the context of evaluation, there are numerous actors involved in different aspects of the evaluation, each of them assuming a different role. Generally speaking we can distinguish four types of actors (Eser and Nussmueller, 2006). The "Program Unit" is the administrative unit responsible for the evaluated governmental intervention. In a sense, the program unit can be seen as the "evaluated". Another type of actor is the "Organisational Stakeholders", who typically represent the organisational hierarchy, here including evaluation coordination units typically higher up in the hierarchy. The organisational stakeholders supervise and control the evaluation exercise. They are internal to the organisation, and act as a linchpin between the program unit and the broader set of social and political actors outside the organisation. This broader set of social and political actors is what we define here as the "External Stakeholders", who do not necessarily know about the intervention in detail or the program unit, but who might be interested in the outcomes of the evaluation for political reasons. External stakeholders are external to the evaluating organisation that the program unit and the

organisational stakeholders belong to. Finally, the "External Evaluators" are consultants and experts contracted to carry out the evaluation. When studying the policy learning that emanates from an evaluation, all four types of actors in figure 1 may potentially learn. Identifying who actually does learn is an empirical question.

Figure 1: The multiple actors involved in an evaluation.

The second essential analytical issue is related to what and how learners learn from an evaluation. This study determines learning by the objectives of the evaluation and the evaluation process. This is what we define here as the "evaluation framework"; namely, the administrative structures that define the evaluation objectives as well as how the evaluation is carried out in an evaluation process. These administrative structures include formal and informal rules and practices related to evaluation, such as guidelines, evaluation policies and standards, methodology and best-practices. Naturally, administrative structures relate to the organization in which evaluation practices are embedded, and in which several of the previously mentioned stakeholders play different and important roles (van der Meer and Edelenbos, 2006). The evaluation framework is always a crucial component of any evaluation exercise because it defines the frame for the execution and utilisation of the evaluation results and the subsequent learning (Højlund, 2014, Leeuw and Furubo, 2008). For this reason, studying what and how learning takes place must inescapably refer to the features of the evaluation framework in question, and the way in which learners act within it.

So far, the focus on evaluation frameworks (or 'evaluation systems' in the evaluation literature) has not adequately captured the tension between the actors and the structure (Højlund, Forthcoming, Leeuw and Furubo, 2008, Rist and Stame, 2006). To capture this tension between the structuring evaluation framework and the learners' actorness and sense making, we seek

inspiration in the two dimensions of learners' control that Dunlop and Radaelli mentioned (Dunlop and Radaelli, 2013). Hence, the learning mode is defined by the degree of learner's control of the objectives and process of the evaluation. The evaluation framework is the administrative structure that potentially defines the degree of control that learners have regarding their learning situation. Hence, the evaluation framework influences who learns, what is learned and how learning occurs.

However, by putting too strong an emphasis on the structuring features of the evaluation framework, one runs the risk of overlooking the actual learning of the learners. As we saw earlier in this paper, the evaluation literature emphasises the importance of actors and their different views, not just across different types of actors, but within the same type of actors. For this reason, we are inspired by van der Meer's remark that learning from evaluation "is not based on the characteristics of the evaluation [framework] as such, but on the outcome of processes of collective sense-making in networks". This collective sense-making has to do with the way in which the learners interpret and develop an understanding of the past experiences and developments of the evaluated program.

Our overall point of departure is the understanding that the way in which an evaluation is defined (the evaluation framework defining the goals and the process of the evaluation) might contribute to specific forms of policy learning. However, it does not entirely or completely determine learning, as it is intrinsically related to the learners' own sense-making of their experience with the program under evaluation. For this reason, we need to open the black box and study the actual processes of learning among and across specific types of actors within the frame provided by the evaluation framework, in order to understand who learns, what is learned and how learning actually takes place.

4.   Designing our Empirical Study

Our empirical study is based on three retrospective program evaluations carried out by the European Commission at approximately the same time. They constitute a collective of qualitative case studies, which are suitable for answering complex 'how' research questions (Stake 2000; 437).

We have chosen three cases of EU program evaluations because the European Commission has played an instrumental role in proliferating evaluation practices and institutionalising programme evaluation practices in Europe over the last three decades (Furubo et al., 2002, Toulemonde, 2000). We have also chosen the three cases because empirical studies are still relatively scarce in relation to learning from evaluation in the Commission and in the European Union (Boswell, 2008, Zito and Schout, 2009, Radaelli and Dunlop, 2013).

The case evaluations are midterm evaluations commissioned by the European Commission, and produced by external evaluators using a relatively similar and comparable methodology common to all European Commission midterm evaluations. The evaluated programmes are of comparable budget sizes for the programme cycle 2007-2013 and work in similar ways through project support.

The first case is the midterm evaluation of the Programme for the Environment and Climate Action (LIFE). The second case is the midterm evaluation of the environmental research program within the Framework Program 7 for Research and Development that was conducted in 2010. The third case is the interim Evaluation (2009) of Competitiveness and Innovation Framework Programme (CIP) - Intelligent Energy - Europe (IEE).

It is important to define who the learners in our study are. The learners in any Commission-led EU program evaluations, like our 3 cases, are potentially the four types identified above: the program unit, the external stakeholders (other Directorates-General (DGs), the Secretariat-General and others), the organisational stakeholders, which are gathered formally in the evaluation Steering Committee normally consisting of the evaluation unit, heads or deputy heads of unit of the program units and other staff. Finally, the external evaluators are external experts or consultants hired mainly through Commission framework contracts.

Prior to the actual data collection, we conducted 35 explorative semi-structured interviews with evaluation experts in the Commission, in order to gain important background knowledge of the evaluation framework and identify the relevant actors to interview for the actual data collection. The actual data for this article was generated from 25 structured in-depth interviews with different types of actors in the three case evaluations presented earlier. The interviews were structured around 20 questions that all interviewees were presented with. Appendix 1 reproduces the questionnaire sent to the respondents. The questions were organized around three themes: the respondents' views on the learning that emanated from the evaluation exercise, their views on the objectives of the evaluation, and their views on the process of the evaluation. Interviewees were sampled purposefully and according to availability. Appendix 2 provides an anonymized list of the respondents according to their program and group of actors they belong to. Our sample included Commission officers in the programming unit, in the Steering Committee, and external experts and consultants. Interview data were validated with document data comprising the three case evaluations and other relevant documents such as the impact assessments and ex ante evaluations of the new programme cycle 2014-2020, as well as several non-public Commission documents relevant to our understanding of the Commission evaluation system.

To analyse the data we used the principles of qualitative content analysis (Schreier, 2012, Kohlbacher, 2006). The actual coding and analysis of interview data were carried out using the NVIVO software package (Bazeley, 2013). The data was classified (coded) along the lines of the

existing conceptual framework from Dunlop and Radaelli (Dunlop and Radaelli, 2013) and following the interview guide. In practice, categories (sub-codes) were added qualitatively under each classification when there was a need for them, all the while observing the relevance to the overall research question. We use this methodology in Table 2 in the analysis to illustrate the patterns in the data that conclusions on learning are based on.

5.   Learning and learners from Commission evaluations

This section analyses what learners learn from evaluations. On this matter, the interviewees identified the learners in the Commission to be primarily the Commission's program units or evaluation units. In principle, there could be learners in the Steering Committee and among external stakeholders, such as decision-makers in the Commission as well as in the European Parliament and the Council. But these external stakeholders were not identified as learners by the interviewees. In general, interviewees distinguish sharply between program management and policy-making when it comes to learning. One program manager put it this way: "I think that the evaluation of a program has a more important outcome for the program implementation than for policy making." This finding is supported by previous research (Højlund, Forthcoming). The respondents did not mention the Steering Committee in relation to learning. One evaluator observed: "I suspect that most of our findings were already on the agenda of the European Commission." This point was mentioned by half of the interviewees. One evaluation unit desk officer said: "I cannot really remember any fundamental things [in evaluations], where I could just say; 'well, this is really something new or something that I never looked at before'." This indicates that learning is gradual and cumulative rather than path-breaking, due to the program units' control over learning goals in the evaluation framework (see next section).

Further, interviewees also associated learning positively with the external evaluators. Evaluators have a steep learning curve when they start the evaluation compared to the Commission staff, who already know the program inside-out. Even though external evaluators are selected on the basis of their expertise and usually through consecutive and competitive rounds of appraisals, they still find EU programs relatively complex. At the beginning of evaluations, external evaluators acquire detailed knowledge about the program and its management and implementation. On the other hand, program units are naturally the most knowledgeable with regard to the programs they manage, and therefore the learning about the program is gradual. In other words, the acquired knowledge usually builds on or confirms hunches and assumptions already made in the program units over the course of time.

Almost all interviewees mentioned learning in relation to evaluation, indicating that learning takes place. Our coded interviewee answers clustered around four types of learning: 1) an 'overview' of the program; 2) small-scale program change; 3) learning in relation to program change; and 4)

learning conceptually about evaluation methodology. Table 2 summarizes the data clustered around these four types of learning in the interviewees' responses. The table contains the number of interviewees who mentioned each of the learning types as well as the total number of times each learning type was mentioned. The types are exemplified with quotes, and a column in the far right of the table is used for the most pronounced learner of each learning type.

Table 2   Overview of Indications of learning and no learning

What is being learned	# Interviewees	# References	Examples of coded quotes	Typical learner
Learning as 'overview'	15	28	- *"It was actually a meta evaluation of the program…and that is also what you learn from, because you see where all the holes are." (Desk Officer – Evaluation Unit)* - *"Yes, we did learn a lot from the evaluation. I think one of the reasons is that it is very hard for us to find the time to do this kind of analysis." (Head of Program Unit)* - *"It is an overview that you would not normally get. But a midterm or final evaluation can provide this overview." (Former Deputy Head of Program Unit)*	Program unit External evaluators
Learning as small-scale program adjustments	11	31	- *"I think that one of the issues that was raised was the question of measurable indicators, which was very important, and at the end of the day it led to a new set of indicators, which were introduced in the program based on the findings of the midterm evaluation." (Programme Desk Officer)* - *"The things we can change, we change. We do not wait. If we think that they can be useful, we do it immediately outright." (Head of Program Unit)*	Program unit
Learning as evaluation's positive impact on policy change	14	35	- *So, it was already foreseen from the design stage as a very important input to another step in the impact assessment for Horizon 2020. (Desk Officer - Evaluation unit)* - *"[the evaluation] is there [and] I presume that the Commission also prioritizes accordingly for the next years' work programs." (Sub-contractor)* - *"So the main things we wanted to get from the evaluation were: how can we improve things for the next Regulation? What things should we modify in the way that the program is designed, so that we can have a better program in the next seven years?" (Deputy Head of program Unit)*	Program unit
Learning about 'methodology, and evaluation	8	11	- *"[The evaluators] derived the kind of topology of impact, which - …from my perspective - was very interesting for the decision making process within the Directorate." (Program Desk Officer)* - *"I also learned something about evaluation and how it is conducted and things like that." (Desk Officer – Evaluation Unit)*	External evaluators
No learning	12	17	- *"The [evaluation] we did made no real 'change the world' findings. I believe that all the changes that were done in [the new program] …were mainly driven top down based on the grand challenges that could be traced out." (External evaluator)* - *"The feeding back into the political decision making process still could be done better. … if I was looking for a way of improving things, I would be trying to do something about feeding [the evaluation] into the higher level decision making somehow, but I don't know exactly how to do that I have to confess." (Head of Program Unit)*	All actors

About half of the interviewees related learning to an 'overview' of the program in response to a question that they did not have before the evaluation. Both program units and evaluators described learning as an overview of general information about the program being evaluated. The overview included general program information such as statistics and compilations of, for example, program-related reports and their content. In addition, evaluators learned extensively about program details that they did not have prior knowledge of.

Interviewees also related learning to small-scale program changes either expressed as such or through examples of new (or old) ideas in the evaluations. Small-scale instrumental changes may be implemented by the Commission without a change of the program Regulation or other legal frameworks, which can only be changed through a political process involving the European Parliament and the Council. Small-scale adjustments might, for example, consist of new indicators to monitor the program, less project visits or more communication across DGs or within a DG. Interviewees from program units related most examples to this type of learning compared to the other learner groups. One program manager explained: "if there were any changes [resulting from the evaluation], they had to be minor."

It was mainly interviewees from program units who related new information from evaluations to policy change. Predominantly, the reference made was to the connection is between the midterm evaluations and the ex ante and IAs for the programming period 2014-2020. One Evaluation Unit desk officer noted: "[The evaluation] was foreseen from the design side as a very important input to another step in the impact assessment for Horizon 2020." Interviewees gave numerous examples of recommendations in the midterm evaluations that were fed into these two documents, and there were numerous references in the IAs and ex ante evaluations to the three midterm evaluations as well. Nevertheless, about a third of the interviewees indicated that there was no policy change as a result of the evaluation.

Finally, respondents associated learning with new information about evaluation methodology. In the main, evaluators related to this type of learning, whereas it was less prominent among officers in program units. This is not surprising, as programs are evaluated every two to three years and program units acquire evaluation knowledge continuously, on top of the support they get from the evaluation unit. Particularly in the FP7 evaluation, most of the external evaluators had little experience with evaluation methodology because they were sector-experts.

Indications of no learning are less predominant than indications of learning. These interviewees referred to learning not taking place, or in particular to policy change not taking place as a consequence of learning from the evaluation. The main reasons for no learning are that the evaluations do not contain any new information and that the group of users is small. Apart from these main reasons, Commission staff also complained about having too few resources for learning, a remark that corresponds to findings in other studies (Borrás, 2011). They mentioned

as well that the evaluation framework was not conducive to learning, that midterm evaluations often lacked project data to produce findings on results and impacts, and finally that programs were already optimized. In the case of the LIFE program, the program was twenty years old and had undergone quite a few changes over time. Hence, several interviewees thought that everything possible within the legal framework had already been tried.

6.   Learners' control of objectives

The learners' control of the evaluation objectives in the Commission is determined by the evaluation framework and the time, experience and resources the learners have to draft the Terms of Reference (ToR). Evaluation objectives are stated in the evaluations' ToR, which are prepared up to a year before the procurement procedure is launched according to the evaluation plan of the Commission. The responsibility for conducting evaluations of a program rests with the program units, and they also draft the ToR in which learning objectives are specified. The ToR contains evaluation questions and contractual obligations as well as special information on the program. In the three cases, drafting the ToR was done by one or two desk officers in close cooperation with the evaluation unit and/or unit responsible for impact assessment. The role of the DG evaluation unit is to make sure that evaluation activities conform to the evaluation guidelines and standards of the Commission, and to help the program unit conduct the evaluation.

In the Commission, the framework for evaluation is mainly the Terms of Reference and the concrete evaluation guidelines and evaluation policy, which consist of several communications related to evaluation. Evaluations in the Commission are supposed to follow a set of standards that are found in the 'common' Commission's evaluation guidelines (Commission, 2004). The guidelines include suggestions for evaluation criteria and questions. Evaluation is stipulated in the Commission's Control Standards as well, and is thus subject to compliance audit by the Internal Audit Services and by the auditors in each DG. One Head of Evaluation Unit described the rigor of the evaluation framework in the following way: "The evaluation standards are pretty clear in what an evaluation has to look like, how it is to be conducted, how it is has to be structured and how it is to be communicated. It is a very straight jacket of very clear rules."

After the program unit has drafted the ToR with the guidance of the evaluation unit, several other stakeholders are involved in the finalization of the ToR. Usually, the Secretariat General comments on the draft ToR, and the horizontal unit responsible for impact assessment (IA) might also be involved to increase the fit between the midterm evaluation and the IA of the subsequent program. This was the case in DG RTD and DG Env. In the former, the evaluation was commissioned particularly to feed into a parallel ex ante evaluation. In the latter, the midterm as well as the ex ante and IA-study were implemented by the same desk officer in the

Commission and with the same consultants. The ToR might also be reviewed by the hierarchy and other DGs, though the latter usually play a minor role in determining the content. In principle, all the organizational stakeholders are also learners, who are invited into the process to enrich it with ideas and create synergies with other programs etc. However, in practice, they also have a controlling function that limits the control of the learners in the program units. Hence, other organizational and external stakeholders might have an interest in certain evaluation questions related, for example, to other programs or rule-following.

Our data suggest that the learners' control of the evaluation objectives is high within the limits of the Commission's evaluation framework. This was so in all three cases. In the case of IEE, the responsibility for the program is with DG Energy, but the implementation is conducted by EACI, an executive agency. DG Energy commissioned the evaluation and wrote the ToR, but that took place in close cooperation with EACI. One Deputy Head of Program Unit described the liberties of drafting in the following way: "When you are the one writing [the ToR], you can more or less write what you want within the framework and infuse it with the ideas you like. Depending on the energy the other [stakeholders] will commit to changing something, you get to have a rather large influence when you draft."

At the same time, the program units' influence on the evaluation objectives is bound by the evaluation framework consisting of evaluation criteria and evaluation questions. One desk officer in a program unit put it this way: "Evaluation questions are never radical questions. They are questions that very implicitly study the basics. …you are not supposed to go for a real investigation of the whole thing." Given the elaborate evaluation framework described above, the degrees of freedom associated with the formulation of the ToR are reduced in scope. One Head of Program Unit noted that writing the ToR and evaluation questions was "not a big task [for the DG], because…it is always about the effectiveness, the efficiency…-- I mean, [it] is a standard set of questions." However, in practice, the appliance of evaluation methodology varies across DGs, and evaluations depend on the experience of the staff in charge and the resources allocated to the evaluation. This was reflected to varying extent by the interviewees in the three cases. In the case of the FP7, the evaluation guidelines were not adhered to with the same rigor as in the two other evaluations.

Overall, the control of evaluation objectives rests mainly with the program units within the scope allowed by the evaluation framework. The interviewees from program units perceived the margins of maneuver to be substantial in terms of formulating the questions and controlling the content of the evaluation, to be substantial. In the case of FP7 evaluation, which served the same purpose as a midterm evaluation, the evaluation framework did not play a significant role. Differences in evaluation practices exist across DGs according to tradition for evaluation and local practices. In DG RTD, it is common to contract academic sector experts, whereas

other DGs usually contract evaluation consultants or country experts in the case of DG DEVCO. The evaluation framework is therefore not necessarily perceived to be a limiting factor in the construction of evaluation objectives, but just as much an aid for the program unit to ask meaningful questions within the limits of what is possible in an evaluation. Therefore, program units retain a high level of control by means of the evaluations that they are also responsible for within the evaluation framework.

On the other hand, evaluators, who also learned from the evaluations studied for this article, were almost completely without control of the objectives of the evaluation. Thus, the independence of external evaluators does not apply to the definition of objectives, issues and questions that the evaluation addresses. In all cases, evaluators had some influence on the methodology through their proposals and the initial kick off and inception phases, where the methodology was agreed with the Commission. However, the evaluation questions were fixed and only changed insofar as data was not available to completely answer the questions satisfactorily. Overall, external evaluators retain almost no control of evaluation objectives.

7.   Learners' control of the evaluation process

The evaluation process is explained by the Commission's evaluation guidelines and standards, and is thus a part of the evaluation framework. The latter stipulates how externalization normally takes place through framework contracts, where consortia compete for evaluation contracts. According to the Commission, around 80% of evaluations in the Commission are externalized (Commission and Jacobsen, 2007). The competitors submit proposals on how they intend to carry out the assignment, and answer the evaluation questions specified in the ToR. Once the best bid is selected, the actual implementation starts with a kick-off meeting and an inception phase, where the methodology and timing are specified and where the Commission can ask questions about the proposal and the further intentions of the evaluation team. The external evaluators and the Commission Steering Committee typically hold between three to six meetings in order for the Commission to oversee the evaluation process. Throughout the process, the designated project manager from the program unit oversees, helps, guides and negotiates with the external evaluators on any matter necessary for the finalization of the evaluation. Over the course of the evaluation, the external evaluators write an inception report, an interim and a draft final report, before the evaluation is finalized. Normally, each report is presented to the Steering Committee that provides guidance and approves the progress of the evaluation.

The three evaluations in this study were implemented by external consultants (LIFE and IEE) or an external expert group (FP7). The evaluation process was managed by desk officers from the

program unit in all three cases, and both the Steering Committee and program unit had some control over guiding the process. However, in practice, the external evaluators are independent and mainly in contact with a representative from the program unit. One external evaluator explains the relationship: "We went along doing the evaluation as we felt, and the major influence on how that evaluation would shape from the Commission side [was NN from the program unit]. The HoU …had a lot of interesting things to say and was very helpful". This is a consequence of the externalized nature of program evaluations in the Commission, where the external evaluators work independently to answer the evaluation questions. The Commission respected this independence in the three cases and thus relinquished its control over the process. The interviewees from program units felt that their needs and comments were taken on-board, and at the same time the external evaluators did not feel pressured by the Commission to take findings in any particular direction. Interviewees from both parties described the evaluation processes in all three cases as good with a mutual understanding between the external evaluators and the Commission. One Head of program unit made the following remark: "To me it was an objective, impartial, professional approach that was adopted by the consultants. I do not think we influenced particularly the way that they did it." One external evaluator described the Commission's involvement in the whole evaluation exercise in the following way: "The commission made quite clear what they needed, but they were open in terms of how they were to get it." This quote illustrates the Commission's control and focus on the objectives and renouncement of control of the evaluation process. In all three cases, the external evaluators also perceived their role to be independent, and the involvement of the Commission in the evaluation process to be positive and constructive.

Even though the external evaluators are in control during the evaluation process, several factors suggest that internal forces are also at play. First, the three cases show that the Commission retains some control through a well-guided and managed evaluation process, in order to steer the overall direction of the evaluation, secure timeliness and mitigate problems in the process. This is found to be the case in a recent research on the evaluation system in the Commission (Højlund, Forthcoming). The number of times an evaluation report is scrutinized by the Steering Committee throughout the evaluation process does reduce the control of the external evaluators. Second, the external evaluator (the consultants more than the experts) depends on the Commission for future evaluation contracts, as the Commission is an important procurer of evaluation services. However, the data do not suggest salient support for this assumption of dependency and its influence on the control of the evaluation process. This is due to the fact that the Commission has an interest in keeping the evaluation independent, in order to secure its legitimacy and thus retain its usability in later policy-making (Højlund, Forthcoming).

Overall, the external evaluators are in control of the evaluation process within the evaluation framework and the subtle control of the Steering Committee and the program unit's project manager of the evaluation. In turn, the program unit in its entirety has little control over the evaluation process.

8.   Conclusions

This paper follows a learner's perspective to ascertain who learns, and what and how learners actually learn from evaluations. Aiming to bridge the gap between the evaluation and policy learning literatures, this paper focuses on the multiple learners' sense-making of their experience with the program evaluated within the frame of the evaluation framework. Taking this learners' perspective means we start by identifying who the learners are before addressing the matter of what and how learning takes place. On this basis, our perspective makes a double contribution. First, it recognizes different types of actors who take part in an evaluation, and identifies empirically those among them who are learners. Second, it underlines the importance of understanding that evaluations are public administration activities designed within an "evaluation framework".

In order to understand 'who learns', our paper has identified four overall types of actors who are commonly involved in evaluations, and who are potentially the learners from evaluation. Those are: program units, external stakeholders, organisational stakeholders and external evaluators. Our empirical findings show that the learners are primarily program units and external evaluators. Organisational stakeholders and external stakeholders are not reported as learning from evaluations to any noteworthy degree.

As to the question of what is learnt from evaluations, our findings show that there are different responses. Some respondents indicate that learning is an overview of the program that gives a different perspective than its day-to-day implementation. Other respondents indicate that learning is a series of small-scale adjustments in the program as a direct result of the knowledge attained from the evaluation. Other respondents emphasize that learning has to do with policy change that is beyond the specific program. And last but not least, respondents also point to the learning of evaluation methodology as such.

Most respondents indicate that learning is not taking place in a radical and disruptive way. Instead most point out that gradual and incremental learning is the norm. Somehow it is as if the outcomes of learning are not surprising, but expected. This brings us to our next finding, which has to do with 'how' learning takes place.

Following previous studies and our focus on the evaluation framework, we have focused on the extent to which the learners were control the evaluation objectives and processes. Our findings show quite clearly that, in the cases studied, the program unit retains control over evaluation objectives, whereas external evaluators retain control of the process. The European Commission has a standardized way of conducting evaluations, so that evaluation objectives are stated in the evaluations' Terms of Reference (ToR), prepared according to relatively fixed procedures. The program units draft the ToR, and they retain control over the specific learning objectives of the evaluation. However, organizational stakeholders monitor this drafting process, which somewhat limits the control exercised by the program unit. The process of the evaluation, for its part, is mainly controlled by the external evaluators. Again, the Commission standards, which define the specific evaluation framework in our three empirical cases, tend to assign control over the process to external evaluators, a process which is well defined beforehand by the evaluation framework combining the specific terms of reference and extended praxis in this area. However, control over the evaluation process is also limited, as it is formally supervised by the organizational stakeholders.

Even in a highly standardized evaluation framework like the one in the European Commission, variations occur according to the nature of the program and the multiple views of the actors who participate in the evaluation. Different actors learn differently from evaluations, and in some contexts, like the ones examined here, the learners tend to be concentrated in two of the four types of actors engaged in evaluation activities. Our data shows that external evaluators are more inclined to learn about evaluation methodologies, and that program units learn about small-scale program improvements and potential policy changes.

Nonetheless, these findings give rise to several research questions. First, it is unclear whether these findings will be found in other empirical cases of evaluation such as prospective ex-ante or impact assessment or retrospective regulatory evaluation. Another open research question is whether learning at the level of organisational and external stakeholders would result in more path-breaking and disruptive learning. These are questions for future research, and the backbone of our plea for further scholarly attention to the link between evaluation and policy learning.

REFERENCES

Balthasar, A. & Rieder, S. (2000) Learning from Evaluations: Effects of the Evaluation of hte Swiss Energy 2000 Programme. *Evaluation,* 6**,** 245-260.

Bazeley, P. (2013) *Qualitative Data Analysis with NVIVO,* London, Sage Publications Ltd.

Bennett, C. J. & Howlett, M. (1992) The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences,* 25**,** 275-294.

Borrás, S. (2011) Policy learning and organizational capacities in innovation policies. *Science and Public Policy,* 38**,** 725-734.

Boswell, C. (2008) The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy,* 15**,** 471-488.

Commission, E. (2004) Evaluating EU Activities - A practical guide for the Commission services. IN Budget, D. (Ed.) Brussels, European Commission.

Commission, E. & Jacobsen, S. (2007) Evalueringsaktiviteter & -resultater på tværs af Europa-Kommissions tjenestegrene. *Dansk Evalueringsselskabs konference.* Kolding, Denmark.

Dobbin, F., Simmons, B. & Garrett, G. (2007) The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning. *Annual Review of Sociology,* 33**,** 449-72.

Dunlop, C. A. & Radaelli, C. M. (2013) Systematising Policy Learning: From Monolith to Dimensions. *Political Studies,* 61**,** 599-619.

Eser, T. W. & Nussmueller, E. (2006) Mid-term Evaluations of Community Initiatives under European Union Structural Funds: A Process between Accounting and Common Learning. *Regional Studies,* 40**,** 249-258.

Freeman, R. (2006) Learning in public policy. IN Moran, M., Rein, M. & Goodin, R. E. (Eds.) *Oxford Handbook of Public Policy.* Oxford, Oxford University Press.

Furubo, J.-E., Rist, R. C. & Sandahl, R. (2002) *International Atlas of Evaluation,* London, Transaction Publishers.

Grin, J. & Loeber, A. (2007) Theories of policy learning: agency, structure, and change. IN Fischer, F., Miller, G. J. & Sidney, M. S. (Eds.) *Handbook of Public Policy Analysis: Theory, Politics, and Methods.* Boca Raton, CRC Press.

Hansen, M., Alkin, M. C. & Wallace, T. L. (2012) Depicting the logic of three evaluation theories. *Evaluation and Program Planning*.

Højlund, S. (2014) Evaluation Use in the Organisational Context – Changing Focus to Improve Theory. *Evaluation,* Special Issue.

Højlund, S. (Forthcoming) Evaluation Systems and their Implications on Evaluation Use – The Case of the European Union. *Evaluation,* Forthcoming.

Kohlbacher, F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research,* 7.

Leeuw, F. L. & Furubo, J.-E. (2008) Evaluation Systems : What Are They and Why Study Them? *Evaluation,* 14**,** 157-169.

March, J. (2010) *The ambiguities of experience* Ithaca, N.Y., Cornell University Press.

Mocker, D. W. & Spear, G. E. (1982) *Lifelong Learning: Formal, Nonformal, Informal, Informal, and Self-Directed. ,* Columbus OH, ERIC.

Owens, S., Rayner, T. & Bina, O. (2004) New agendas for appraisal: reflections on theory, practice, and research. *Enrivonment and Planning,* 36**,** 1943-1959.

Patton, M. Q. (1997) *Utilization-focused evaluation: The new century text,* Thousand Oaks, Sage.

Preskill, H. & Boyle, S. (2008) A multidisciplinary model of evaluation capacity building. . *American Journal of Evaluation,* 29**,** 443.

Radaelli, C. M. (2009) Measuring policy learning: regulatory impact assessment in Europe. *Journal of European Public Policy,* 16**,** 1145-1164.

Radaelli, C. M. & Dunlop, C. A. (2013) Learning in the European Union: theoretical lenses and meta-theory. *Journal of European Public Policy,* 20**,** 923-940.

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

Rist, R. & Stame, N. (2006) *From Studies to Streams: Managing Evaluative Systems,* London, Transaction Publishers.

Sager, F. & Rissi, C. (2011) The limited scope of policy appraisal in the context of referendum democracy - the case of regulatory impact assessment in Switzerland. *Evaluation,* 17**,** 151-163.

Sanderson, I. (2000) Evaluation in Complex Policy Systems. *Evaluation* 6**,** 433-454.

Schreier, M. (2012) *Qualitative Content Analysis in Practice,* Thousand Oaks, SAGE Publications Ltd.

Scriven, M. (1991) Beyond Formative and Summative Evaluation. IN McLauglin, M. & Philips, D. C. (Eds.) *Evaluation and Education: At Quarter Century.* 3 ed. Chicago, University of Chicago Press.

Shulha, L. M. & Cousins, B. (1997) Evaluation Use: Theory, Research and Practice Since 1986. *Evaluation Practice,* 18**,** 195-208.

Toulemonde, J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *Vierteljahrshefte zur Wirtschaftsforschung,* 69**,** 350-357.

Van der Knaap, P. (1995) Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation? *Evaluation,* 1**,** 189–216.

van der Meer, F.-B. & Edelenbos, J. (2006) Evaluation in Multi-Actor Policy Process. Accountability, Learning and Co-operation. *Evaluation,* 12**,** 201-218.

Vedung, E. (1997) *Public Policy and Program Evaluation,* New Brunswick, NJ: Transaction Publishers.

Zito, A. & Schout, A. (2009) Learning theory reconsidered: EU integration theories and learning. *Journal of European Public Policy,* 16**,** 1103-23.