Evaluation in the European Commission

For Accountability or Learning?

Steven Højlund*

The article accounts for the development of the European Commission's evaluation system. The article shows how internal and external developments shape an evaluation system aiming for both accountability and policy learning. In 58 interviews, several justifications for the evaluation system emerge, including four types of accountability as well as an evaluation system constructed to facilitate learning from past experience. In the system's commencement, financial and legal accountability overshadow the focus on policy learning that was perceived to be in contradiction with the two former. However, the article also demonstrates that accountability and policy learning are not necessarily opposed to each other.

I. Introduction

This article contains a historical account of the European Commission's evaluation system and how its implementation was related to concerns over accountability and policy learning. The article applies a historical research design to answer how and why evaluation practices were introduced, as well as looking at their consequences on policy learning. The historic approach allows us to see the internal and external developments that shape the evaluation system, as well as how the Commission balances the focus of the formal implementation between concerns for accountability and policy learning. Evaluation as a phenomenon makes the perfect case because evaluation can be both learning-oriented as well as ac-

countability-oriented.[1] To fully grasp the intricacies of the dichotomy between policy learning and accountability, the article breaks up accountability into four types that emerged during the data analysis. Over time, the emphasis on these four types of accountability changes with consequences for policy learning in the European Commission.

The European Commission's evaluation system was chosen as the empirical case. On one hand, the Commission is expected to learn from past experience to improve its policies and drive European integration. On the other hand, pressures to hold the Commission accountable to the European Parliament (EP), Member States (MSs) and ultimately to European citizens are strong. This double pressure on the Commission makes it a good case for the study of how the setup of the evaluation system in public administrations balances considerations for accountability and policy learning. Moreover, research on the Commission rarely focuses on its internal practices.[2] Also, very little attention has been given to the Commission's evaluation system, which has been developing gradually over the last 30 years.[3] It is in this area that the article makes its contribution. The Commission has had a significant impact on evaluation practices and the setting up of evaluation systems in public administrations across Europe through conditionality.[4] This warrants a focus on the justification and implementation of its own evaluation practices.[5]

The article concludes that policy learning and accountability are not necessarily opposed in an eval-

---

*   Steven Højlund holds a PhD from Copenhagen Business School – Department of Business and Politics.

1   Scriven M, 'Beyond Formative and Summative Evaluation' in McLauglin M and Philips DC (eds), Evaluation and Education: At Quarter Century (3 edn, University of Chicago Press 1991)

2   Gornitzka Å and Sverdrup U, 'Access of experts: information and EU decision-making' (2011) 34 West European Politics 1452. Smith AS, 'How the European Commission's Policies Are Made: Problematization, Instrumentation and Legitimation' [2013] Journal of European Integration

3   Stern E, Evaluation policy in the European Union and its institutions, vol 123 (Trochim WMK, Mark MM and Cooksy LJ eds, 2009)

4   Furubo J-E, Rist RC and Sandahl R, International Atlas of Evaluation (Furubo J-E, Rist RC and Sandahl R eds, Transaction Publishers. 2002)

5   Hoerner J and Stephenson P, 'Theoretical Perspectives on Approaches to Policy Evaluation in the EU: The Case of Cohesion Policy' (2012) 90 Public Administration 699

uation system. By decomposing accountability the article finds evidence for a change in the focus on accountability in time, which in turn changes the effects that the evaluation system has on policy learning. Hence, the focus and potential for policy learning from evaluation in the Commission's evaluation system was reduced after the major Commission reforms in the 1990s and 2000s due to a focus on financial and legal accountability. However, evaluations were more likely to induce policy learning before the evaluation system was in place and in the late 2000s. This points in the interesting direction that accountability sometimes can induce policy learning rather than only be a burden to public administration.

II. Studying the Commission's Evaluation System

Over the past three decades, the Commission's evaluation practices have evolved into an 'evaluation system'. An evaluation system can be understood as permanent and systematic evaluation practices taking place and institutionalized in several interdependent organizational entities with the purpose of informing decision-making and securing oversight. According to Leeuz and Furubo's four elements[6] characterize an evaluation system: 1) Participants share a common understanding of the objectives of evaluation and the means by which the objectives are attained; i.e. evaluation is taken for granted and considered 'normal' practice; 2) The evaluation system is institutionalized formally in at least one organizational structure, in which it is separated from the operational structure. Hence, the system has at least one formal institutionalized organizational element (e.g. 'an evaluation unit') that typically is in charge of planning, tendering, implementing, quality-checking and follow-up; 3) Evaluation systems are permanent in the sense that their setup has no time-limitation; 4) Evaluations are undertaken continuously and systematically and in relation to previous and future evaluations as well as to the cycle of activities of the organization (e.g. budget or policy cycle).

The EU institutions constitute a very good example of an evaluation system. The system's core consists of the Commission, the EP and the Council. It is the Commission that undertakes most evaluations

in the system, but the EP and MSs do also carry out or commission evaluations usually according to Commission evaluation standards. The Commission has a legal obligation to evaluate programmes and policies as stipulated in the Commission's management policies, as well as regulations of the programmes and legislative frameworks. For this reason, the Commission has institutionalized evaluation practices over the last 30 years in each Directorate General (DG) through a common evaluation policy, guidelines and standards. In the DGs, designated evaluation units supervise and guide evaluation activity with support from the Secretariat-General (SG). The evaluation units are subject to internal audits as they are described in the Internal Control Standards of the Commission services and evaluation is an integral part of the activity-based management and budgeting system of the Commission and is thus formally related to the decision-making regarding EU budgetary allocations.

The Commission's evaluation system has expanded over the last 30 years to systematically cover all expenditure programmes and most Commission activities including policies, agencies and strategies etc. An estimated 80% of all evaluations are contracted out[7] and therefore an evaluation industry has been created by the Commission's public procurement of these and similar services. Commission evaluations are usually carried out by teams of academic experts, consultants or research institutes. The Commission allocated an estimated 140 full time positions to managing evaluations in 2007 and spent 45 million EUR annually on evaluations (including external assistance and staff resources) (European Court of Auditors [CoA] estimate).[8]

III. Policy Learning and Accountability

This article investigates the concerns for policy learning as well as accountability during the setup of the Commission's evaluation system. A dichoto-

---

6   Leeuw FL and Furubo J-E, 'Evaluation Systems : What Are They and Why Study Them?' (2008) 14 Evaluation 157

7   Jacobsen Sven, "Evalueringsaktiviteter & -resultater på tværs af Europa-Kommissions tjenestegrene", 2007

8   Communication to the Commission from MS GryBauskaite in agreement with the President. Responding to Strategic Needs: Reinforcing the use of evaluation, SEC(2007)213, at p. 17.

mous and opposed relationship between the two is well-described in several research areas, such as evaluation studies;[9] audit studies[10] and studies on knowledge use and governance in the Commission.[11]

Commonly it is suggested that tools such as evaluation do not always generate learning even though they were implemented in order to do so and that control systems are implemented at a cost to general performance.[12] In fact, learning and accountability understood as outcomes are often found to be contradicting to one another.[13] There are several explanations for this contradiction. In the literature on evaluation the intended uses are commonly identified as the cause that can explain evaluation outcomes related to learning or accountability.[14] Learning-oriented evaluations (formative evaluation) are intended to offer advice aimed at improving the effectiveness of a program, whereas accountability-oriented evaluations (summative evaluations) are intended to evaluate the existence of the programme.[15] However, other explanations draw on organisation-

al theory to argue that an organisation can be compelled to evaluate for reasons of accountability because of its dependency on its organizational environment[16] instead of independently pursuing its preferred strategies of goal attainment and utility maximization.[17]

The evaluation literature has been more concerned with concepts of 'evaluation utilization' rather than learning.[18] But learning is implied in several concepts of evaluation utilisation such as 'conceptual use', 'process use' or 'instrumental use',[19] while accountability is related to 'legitimising use' or 'symbolic use'.[20] This article focuses on policy learning happening in the political organization or bureaucracy itself.[21] Policy learning is defined as 'a process of updating beliefs about policy based on lived or witnessed experiences, analysis or social interaction'.[22] Policies are understood as both legal and non-legal acts as well as expenditure programmes. Standard definitions of evaluation have a tendency to focus on the formative role of evaluation. As an example Vedung[23] defines evaluation as 'careful retrospective as-

9   Scriven M, 'Beyond Formative and Summative Evaluation' in McLauglin M and Philips DC (eds), *Evaluation and Education: At Quarter Century* (3 edn, University of Chicago Press 1991). Albæk E, 'Between Knowledge and Power: Utilization of Social Science in Public Policy Making' (1995) 28 Policy Sciences 79.

10   Power M, 'The Theory of Audit Explosion' in Ferlie E, Lynn LE and Pollitt C (eds), *Oxford Handbook of Public Management* (Oxford University Press 2005). Boven M, 'Public Accountability' in Ferlie E, Lynn LE and Pollitt C (eds), The *Oxford Handbook of Public Management* (Oxford University Press 2005). Dubnick M, 'Accountability and the promise of performance - In search of the Mechanisms' (2005) 28 Public Performance & Management Review 376.

11   Boswell C, 'The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy' (2008) 15 Journal of European Public Policy 471. Böhling K, 'Sidelined Member States: Commission learning from Experts in the Face of Comitology' [2013] Journal of European Integration. Mendez C and Bachtler J, 'Administrative reform and unintended consequences: an assessment of the EU Cohesion policy 'audit explosion'' 18 Journal of European Public Policy 746

12   Boven, Public Accountability, supra note 10.

13   Scriven M, 'The methodology of evaluation' in Stake RE (ed), *Curriculum evaluation* (Rand McNully 1967). Scriven, *Beyond Formative and Summative Evaluation*, supra note 8. Chen HT, 'A comprehensive typology for program evaluation' (1996) 17 Evaluation Practice 121.

14   Torres RT and Preskill H, 'Evaluation and Organizational Learning: Past, Present, and Future' (2001) 22 American Journal of Evaluation 387. Weiss CH, 'Have We Learned Anything New About the Use of Evaluation' (1998) 19 American Journal of Evaluation 21. Balthasar A, 'Institutional Design and Utilization of Evaluation - A Construction to a Theory of Evaluation Influence Based on Swiss Experience' (2009) 33 Evaluation Review 226. Chen, A comprehensive typology.

15   McDavid JC, Huse I and Hawthorn LR, Program Evaluation and Performance Measurement (Sage Publications 2013). Scriven, *The*

methodology of evaluation, supra note 13. Scriven, *Beyond Formative and Summative Evaluation*, supra note 8.

16   Dahler-Larsen P, *The Evaluation Society* (Stanford Press 2012). Højlund S, 'Evaluation use in the organisational context – Changing focus to improve theory' 20 Evaluation 26.

17   Meyer JW and Rowan B, 'Institutionalized organizations: Formal structure as Myth and Ceremony' 83 American Sociological Review 340.

18   Shulha LM and Cousins B, 'Evaluation Use: Theory, Research and Practice Since 1986' (1997) 18 Evaluation Practice 195. Preskill H and Boyle S, 'A multidisciplinary model of evaluation capacity building. ' (2008) 29 American Journal of Evaluation 443.

19   See for example Balthasar A and Rieder S, 'Learning from Evaluations: Effects of the Evaluation of hte Swiss Energy 2000 Programme' 6 Evaluation 245. Weiss CH, 'Have We Learned Anything New About the Use of Evaluation' (1998) 19 American Journal of Evaluation 21. Patton MQ, *Utilization-focused evaluation: The new century text* (3 edn, Sage 1997). Alkin MC and Stecher B, 'Evaluation in Context: Information Use in Elementary School Decision Making' (1983) 9 Studies in Educational Evaluation 23. Finne H, Levine M and Nilssen T, 'Trailing research: A model for useful program evaluation' (1995) 1 Evaluation 11.

20   Alkin M and Taut S, 'Unbundling Evaluation Use' (2003) 29 Studies in Educational Evaluation 1

21   Schofield J, 'A model of learned implementation' (2004) 82 Public Administration 283. Bennett CJ and Howlett M, 'The lessons of learning: Reconciling theories of policy learning and policy change' (1992) 25 Policy Sciences 275.

22   Radaelli CM and Dunlop CA, 'Learning in the European Union: theoretical lenses and meta-theory' (2013) 20 Journal of European Public Policy 923.

23   Vedung E, *Public Policy and Program Evaluation* (NJ: Transaction Publishers 1997), at p. 3.

sessment 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'.

Public accountability is the hallmark of democratic governance. Similar to the definition of policy learning, accountability is broadly defined in this article 'as a social relationship in which an actor feels an obligation to explain and to justify his or her conduct to some significant other'.[24] As implied by the definition, accountability in a setting of public governance can be many things and ultimately relies on the norms, values and cognitive frameworks of actors. The literature distinguishes between several types of accountability of which traces of hierarchical, democratic, legal and financial accountability were found in the interview data used in this article. The four types of accountability are described as follows:

1. **Hierarchical** accountability is the internal accountability to hierarchy that most public administrations build on. Therefore, hierarchical accountability is sometimes also called 'managerial accountability'.

2. **Legal** accountability refers to the legal obligations formal organization builds upon with regard to implementing formal structure (e.g., evaluation units) as well as practices (e.g., obligation to evaluate). The legal framework of the Commission relates to the Treaty, the Financial Regulation as well as other community legislation.[25]

3. **Financial** accountability is related to budgetary execution and financial controls that an organization can be subject to. One example is audits.[26]

4. **Democratic** accountability is the external accountability to democratically elected bodies that undertake oversight and the public: for example the EP's oversights of the Commission.[27]

An evaluation can be related to several types of accountability at the same time. As we shall see in the analysis below interviewees refer to several reasons why they evaluate, including several types of accountability.

IV. Methodology and Data

This article is based on a historical case study of the Commission's evaluation system.[28] This article's analysis is based on data generated from 58 recorded interviews, 2 group interviews and 1 conference on evaluation in the EP along with numerous informal talks with experts and Commission desk officers, as well as personal observations made during evaluations conducted for the Commission. Interviewees were sampled purposefully and according to availability. Interviewees included Commission employees working in evaluation units and policy units, as well as external evaluators, evaluation trainers and consultants working with the Commission in the setting up of the evaluation system. Several of the interviewees were senior staff with key roles in the early implementation of the evaluation system, and thus with a good overview of the history of evaluation in the Commission. Interview data were validated with historical document data (e.g. internal evaluation policy papers, guidelines, minutes of meetings in the evaluation network etc.). Data were analysed according to the principles of qualitative content analysis and by using the NVIVO software package[29] and coded according to the interviewees' perceptions of accountability and policy learning.[30]

V. Analysis

The analysis is presented chronologically and broken down into four periods. Each period emerged from

24 Boven, *Public Accountability, supra* note 10.

25 Harlow C, *Accountability in the European Union* (Oxford University Press 2002).

26 Laffan B, 'Auditing and accountability in the European Union' (2003) 10 Journal of European Public Policy 762.

27 Wille A, 'Political–Bureaucratic Accountability in the EU Commission: Modernising the Executive' (2010) 33 West European Politics 1093.

28 For similar approach to studies of the Commission see for example Wonka A, 'Decision-making dynamics in the European Commission: partisan, national or sectoral?' (2008) 15 Journal of European Public Policy 1145. Schout A, 'Organizational learning in the EU's multilevel-governance system' (2009) 16 Journal of European Public Policy 1124. Smith, *How the European Commission's Policies Are Made, supra* note 2. Böhling, *Sidelined Member States, supra* note 11. Burns C, 'How and When Did We Get Here? An Historical Institutionalist Analysis of EU Biotechnology Policy' (2012) 34 Journal of European Integration 341.

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

30 Schreier M, Qualitative Content Analysis in Practice (SAGE Publications Ltd. 2012). Mayring P, 'Qualitative Content Analysis' in Flick U, Von Kardorff E and Steinke I (eds), *A Companion to Qualitative Research* (SAGE 2004). Kohlbacher F, 'The Use of Qualitative Content Analysis in Case Study Research' (2006) 7 Forum: Qualitative Social Research.

Table 1 - Periods of implementing evaluation in the Commission

	1980–94	1995–99	2000–06	2007–13
Key events	- Increased budget - NPM paradigm starts in Europe	- SEM2000 - Evaluation linked to policy-cycle - Evaluation units - MEANS guidelines	- Kinnock reform - Better Regulation - Impact Assessment - Evaluation linked to control standards - SG takes over evaluation from DG Budget	- Evidence-based policy making - EP and CoA policy evaluation capacity - DGs develop regulatory evaluation guidelines - Fitness check

the data and was marked by one or more events related to the evaluation system's focus on policy learning and/or accountability. Table 1 sums up the periods.

1. 1980–94: Decentralized Sector-based Evaluation

The first period that emerged from the data was the period between the early 1980s and 1994. Evaluation in the early 1980s, and possibly earlier, happened unsystematically in a few DGs based on sectoral best practices.[31] Most prominent were the evaluation structures developed in DG Aidco (Development aid) and DG Rtd (Research), which both had large expenditure programmes operating in sectors with evaluation norms and practices already established. A former Head of Evaluation Unit (HoEU) explained the relationship like this: '[the DGs] were sharing experiences and practices, but at the end of the day each DG was doing its own thing.'

As evaluations were unsystematically implemented and managed in the DGs without a general Commission policy or standard for evaluation, evaluation practices could be tailored more easily to focus on policy learning. One interviewee notes about this period: 'the main issue was the effectiveness of the policy [in "the golden age [of evaluation]". I think the following decade was more dominated by audits and control.' In this early period prior to the setting up of the evaluation system, the Commission had not experienced any major reforms and was functioning primarily according to ideas of ex post appraisals of effectiveness and performance. Accountability was primarily hierarchical, emphasizing procedural rules and hierarchy in order to deliver effective policy formulation according to the Commission mandate.[32] In DG Aidco and DG Rtd, evaluation was introduced to adapt to best practices in the sector but also to satisfy demands from MSs for increased controls with the expenditure programmes.

However, in 1993 the budget for expenditure programmes increased and the demand for financial accountability increased too.[33] In DG Regio, systematic programme evaluation started in the 1990s as a consequence of a more than doubling of the budget for Structural Funds in the second Delors package.[34] According to one HoEU there were two purposes for the introduction of systematic evaluation in DG Regio; 'DG Regio had to justify [the programme] in 1993 because of the doubling of the funds. The other reason was that there was a wave of public sector reforms in the EU.' Severl interviewees refer to the New Public Management (NPM) reforms in Europe that stressed a focus on control of public spending, ac-

31 Summa H and Toulemonde J, 'Evaluation in the European Union: addressing complexity and ambiguity' in Furubo J-E, Sandahl R and Rist R (eds), *International Atlas of Evaluation* (Transaction Publishers 2002), at p. 409.

32 Christiansen T, 'Tensions of European governance: politicized bureaucracy and multiple accountability in the European Commission' (1997) 4 Journal of European Public Policy 73. Ellinas A and Suleiman E, 'Reforming the Commission: between modernization and bureaucratization' (2008) 15 Journal of European Public Policy 708.

33 Summa, *Evaluation in the European Union*, *supra* note 31, at p. 412. Laffan B, 'From policy entrepreneur to policy manager: the challenge facing the European Commission' (1997) 4 Journal of European Public Policy 422.

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

countability, and effectiveness and efficiency of public interventions.[35] Together with the budget increases the NPM trend spawned a greater demand for scrutiny and control in the Commission. One HoEU explained the development in the following way: 'it was an audit culture from the very beginning'. At the end of this period, evaluation of expenditure programmes became an instrument primarily to produce financial accountability rather than just internal hierarchical accountability.

In sum, the period until 1994 saw no systematic and centralized evaluation policy in the Commission. DGs could evaluate more or less as they saw fit and in turn evaluation was primarily a tool for policy learning as well as hierarchical accountability.

2. 1995–1999: Centralized control and Accountability

The period from 1995 to 1999 was a turbulent period for the Commission. Before the Santer Commission resigned, it started the most comprehensive reform programme of the Commission since its commencement. The reform was called SEM2000 and was inspired by ideas of NPM, and in 1995 the SEM2000 Stage II Communication called for the systematic, timely and rigorous evaluation of expenditure programmes.[36] Evaluation practices were drawn up in the Communication from 1996 entitled 'Concrete Steps towards Best Practice across the Commission' that formally set up the evaluation system in the Commission. The Communication justified evaluation in the following way:

> The increasing size of the general budget and the concern of European Union taxpayers to secure value for money from their contributions, together with developments in Member States and elsewhere towards results-oriented public sector management have created the necessity to review and to update the Commission's approach to managing the budget.[37]

It is evident from reading this justification that financial accountability is the main concern for the Commission at this point. In relation to accountability, evaluation had previously been related to hierarchical accountability. But with the implementation of the evaluation system and its association with the budget, financial accountability became more impor-

tant. The Financial Regulation (FR) was amended in 1995 and read: 'All operations must be subject to regular review, in particular within the budgetary procedure, so that their justification may be verified.'[38] Evaluation thus became legally tied to the sound financial management of the Commission and in this period evaluation was therefore mainly linked to financial accountability.

The overall responsibility for the system's implementation was delegated to DG Budget and DG Financial Control. These two DGs – headed by their Nordic Commissioners, Erkki Liikanen from Finland (DG Budget) and Anita Gradin from Sweden (Financial control) – also undertook the drafting of the Communication establishing the evaluation system.[39] The Communication called for each DG to establish an evaluation capacity with responsibility for evaluation and drawing up of an annual evaluation plan for all programmes. DG Budget, DG Financial Control and SG were responsible for common evaluation guidelines, training and compiling evaluation experience. Retrospective evaluations (midterm and ex post) were to be implemented through a 'learning by doing' process and should be applied every sixth year.[40]

To support policy learning the evaluation system was decentralized, placing evaluation functions in the DGs to improve the timing and knowledge development for decision-making as close to the sectoral expertise and decision-making as possible[41]. Evaluation functions were implemented across all DGs even though some DGs did not have spending programmes. One HoU in DG Budget argued that;

> the misunderstanding when introducing evaluation [in the SEM2000 program] was that... evalu-

---

35 Pollitt C and Bouckaert G, *Public Management Reform: A Comparative Analysis* (2nd edn, 2004). Ellinas, *Reforming the Commission, Supra* note 32.

36 The Commission's Communication 1996 'Concrete steps towards best practice across the Commission'. Communication to the Commission by Mr Liikanen and Mrs Gradin, in agreement with the President, SEC(1996)659.

37 Commission, *Concrete steps, supra* note 36, at p. 1.

38 Council Regulation (EC, Euratom, ECSC) No 2335/95 of 18 September 1995.

39 Commission, Concrete steps, supra note 36. Laffan B, *The Finances of the European Union* (Macmillan 1997). Laffan. *From policy entrepreneur to policy manager*, supra note 33.

40 Commission, *Concrete steps, supra* note 36.

41 Williams K, Laat Bd and Stern E, *The Use of Evaluation in the Commission Services* (2002).

ation was limited to spending programs. And that was actually why many of the DGs into which evaluation was meant to be introduced...did not quite understand what they were supposed to be doing. Because they did not have any spending programs.

Consequently, the interest in evaluation across DGs was negligible and the implementation sluggish. Also, in DG Budget staff questioned the rationale of evaluation. According to a former HoU in DG Budget: 'DG Budget was [initially] rather suspicious about this activity on evaluation. ...they did not really see the point [and] you could sense that nobody really took it very seriously.' In addition, systematic evaluation was only linked to expenditure programmes and not legislation. Some spill-over in terms of learning was envisaged. But, as one programme manager notes: '...the interest is only with the [programme implementing units]. The [policy units] have no interest in the evaluation and they do not use evaluations'. Therefore the setup of evaluation functions in each DG often decoupled evaluation from policy units and the attention of the hierarchy.

In practice, as the majority of evaluations are outsourced to consultants, the evaluation function and the decentralized evaluation system came to operate de facto as evaluation management (managing and not conducting evaluations) ensuring 'that, where external evaluators have to be employed to undertake studies, they are given clear and feasible terms of reference and consistently supervised so as to ensure that the Commission (and its partners) retain control over the content and timing of studies (evalua-

tions)'.[42] The externalization of evaluation to consultants was stressed ultimately to secure objective and sound evaluations contributing to democratic accountability. However, as evaluation responsibility and oversight were removed from hierarchy and compartmentalized, their potential impact on policy learning was reduced. An evaluation unit desk officer adds on learning: 'often we are not surprised about the results [of evaluations]'. A former staff member in DG Budget adds to that: '[evaluation] is very divorced from any serious intellectual academic thinking. ...It becomes a ritual.' Therefore, the evaluation system created an 'evaluation bureaucracy' that managed mandatory evaluations according to the formalized rules of the system, with the aim of learning as the secondary objective.[43]

In 1995, DG Regio commissioned the MEANS guidelines that were the first comprehensive evaluation guidelines in the Commission. They laid the foundation for an utilization-focused and non-paradigmatic evaluation methodology in the Commission respecting the different practices in each DG. The MEANS guidelines connected evaluation with the Commission's 7-year policy cycle.[44] According to one key observer, the main consequences of this were that there were too many evaluations conducted during the programme cycle resulting in bad timing in relation to decision-making, which ultimately had a negative effect on policy learning.[45][46] A programme manager noted about the frequency of evaluation over the 7-year programme cycle: 'We do not even have time to really do a follow-up [of the evaluation] because we are just really doing the next evaluation and [then] we have to start drafting the new Regulation.' Tying evaluation to the policy cycle was done more for accountability reasons than learning reasons, as one key observer puts it: 'MEANS was just bothering [people] with evaluation. [The Commission] was so silly [implementing] MEANS [and] SEM2000 to provide for accountability [and not for learning].' The MEANS became standard in the Commission as most subsequent DG-specific guidelines and all internal evaluation training in the Commission were developed and implemented based on the MEANS guidelines.

In 1999, the Commission issued the Communication 'Spending more wisely: Implementation of the Commission's evaluation policy' that took stock of the implementation of the evaluation system.[47] It concluded that more needed to be done in order to

---

42 Commission, *Concrete steps, supra* note 36.

43 Summa, *Evaluation in the European Union, supra* note 31, at p. 420.

44 Eser, *Mid-term Evaluations, supra* note 34.

45 See also Højlund S, 'Evaluation use in evaluation systems - the case of the European Commission' (2014) 20 Evaluation 428.

46 Ex ante evaluation is conducted about two years ahead of programme start. Midterm evaluation is conducted three years into programme implementation, final evaluation is conducted at programme termination, and ex post evaluation is conducted two years after programme termination. Effectively, the evaluation cycle is therefore not seven years but eleven years and overlapping each other. Including ex ante/IA, there are effectively 2–3 years between evaluations of a programme.

47 The Commission's Communication 1999, *"Spending more wisely: Implementation of the Commission's evaluation policy"*, SEC(1999)69/4.

institutionalize evaluation in the Commission. Ex ante appraisals were not used systematically, DG Budget should systematize the use of evaluations in budgetary decision-making and DG Financial Control should consider how performance information from evaluation could be used to assess the extent to which EU spending programmes achieved their objectives.[48] In 1998, DG Audit produced an annual report concluding that evaluation practices were well developed in the areas of Structural Funds, Rtd and external relations. However, at this point in time, most DGs had neither evaluation process standards nor any strategy for using evaluation results or evaluating evaluation quality.[49] Thus the evaluation system was designed to secure financial accountability rather than hierarchical and democratic accountability. In relation to democratic accountability, a key external observer of the Commission found that the 'actual use of EU evaluation for accountability in the EP [does] not [take place] at all.' Thus evaluation was not yet related to democratic accountability in practical terms.

As a consequence of the focus on financial accountability, policy learning from retrospective evaluation in the evaluation system's early years did not receive much attention. However, the evaluation system was not implemented overnight and the following period would see a reinforcing of the legal structures to support a faster implementation of the system as a whole with a continued emphasis on financial accountability.

3. 2000–2006: Kinnock Reform and Evidence-Based Policy-Making

In 1999, the Santer Commission resigned due to alleged corruption.[50] To re-establish trust, the subsequent Commission headed by Romano Prodi issued the Kinnock reform, named after the newly appointed British Vice President of the Commission. The Kinnock reform moved the Commission's management system from a centralized system to a performance-oriented system with activity-based management as one of the key features. It was the most comprehensive reform undertaken to date in the Commission.[51] Focus on financial accountability increased and financial management, control and audit came to play an important role in Commission thinking.[52]

To implement the Kinnock reform in relation to learning from evaluation, the Commission issued the Communication 'Focus on Results: Strengthening Evaluation of Commission Activities'.[53] It stressed in particular the role of the evaluation capacity to ensure better policy-making, programme design and planning functions in relation to resource allocation as well as the consolidation of existing evaluation principles. As the title suggests, the focus was now on all activities of the Commission and not only expenditure programmes. According to Toulemonde et al.[54], the evaluation system was consolidated in relation to expenditure programmes at this time.

More importantly for this period was that the evaluation system was also linked to the Commission's Internal Control Standards, which are binding to the Commission and can be audited.[55] Thus, DG Financial Control was given the competence to conduct 'regular review of the organization and the systems DGs have put into place to achieve proper evaluation and feedback.' Though the entire evaluation system has not yet been audited by the Internal Audit Service, examples exist of evaluation functions in the DGs being audited by the audit capability in the DG. However, the main point to make here is that legal accountability became more important in order to reinforce the evaluation and ultimately strengthen financial accountability. Due to a previous sluggish implementation of the evaluation system in some DGs

---

48 See Laffan, *The Finances of the European Union, supra* note 39.

49 Toulemonde J, Summa H and Usher N, 'Assessing EU Evaluations' in Schwartz R and Mayne J (eds), *Quality matters: seeking confidence in evaluation, auditing and performance reporting* (Transaction Publishers 2005), at p. 82.

50 Spence D, 'Plus ca change, plus c'est la meme chose? Attempting to reform the European Commission' (2000) 7 Journal of European Public Policy 1.

51 Commission, Reforming the Commission: A White Paper (2000). Kassim H, ''Mission impossible', but mission accomplished: the Kinnock reforms and the European Commission' (2008) 15 Journal of European Public Policy 648. Schön-Quinlivan E, 'Implementing organizational change - the case of the Kinnock reforms' (2008) 15 Journal of European Public Policy 726.

52 Bauer M, 'Introduction: Organizational change, management reform and EU policy-making' (2008) 15 Journal of European Public Policy. Mendez, *Administrative reform and unintended consequences, supra* note 11.

53 Communication to the Commission from Mrs. Schreyer in agreement with Mr. Kinnock and the President, *"Focus on results: Strengthening evaluation of Commission activities"*, SEC(2000)1051.

54 Toulemonde, *Assessing EU Evaluations, supra* note 49.

55 Commission, Responding to Strategic Needs, Supra note 8.Commission, Revision of the Internal Control Standards and Underlying Framework - Strengthening Contol Effectiveness (European Commission 2007).

the implementation was reinforced by making evaluation subject to audits and thus adding an extra level of oversight.

In order to streamline and secure the full implementation of the evaluation system, a working group in the Commission's inter-service evaluation network had begun working on common standards for evaluation. However, due to the heterogeneity of evaluation methods and standards across DGs, the common standards became watered down to common evaluation guidelines, which should represent the best practice in the Commission but which did not exclude the possibility of each DG developing their own guidelines. All guidelines were inspired by the MEANS and other already existing guidelines in the DGs.[56] Also, inspiration was drawn from the standards and guidelines in the OECD and academia. The final work was called '*Good Practice Guidelines for the Management of the Commission's Evaluation Function*', and was adopted in 1999 with a revision in 2000. A final edition came in 2004 and is still in use today.

In relation to policy learning the results were limited. The Kinnock reform represented the completion of the evaluation system's implementation along the lines of NPM principles and thus completed the negative impact this integration had on policy learning due to what was perceived by many staff members in the Commission as burdensome controls.[57] According to a HoEU, evaluation 'became more systematic [with the Kinnock reform], but [also] a kind of bureaucratic thing. I mean if you think how evaluation results fed into the budget cycle of the Commission, [the] impact is very limited.' This is confirmed by several contributions on EU evaluation.[58] One programme manager put it this way:

well you know, this midterm evaluation, it is not something that is really very present in our minds. I mean we have to do it, because it is an obligation....in terms of lessons learned ..., I would say we did not learn something really fundamental.

In an inter-service survey from 2002, the majority of evaluation unit staff considered legal requirement to be the main reason why evaluations were undertaken.[59] A later stocktaking exercise in the Commission concluded that 'all evidence shows that overall allocation of EU resources is certainly not determined on efficiency grounds, but much more by political decision-making. The study shows especially that such decisions are generally very little informed by evaluation'.[60] Another consultant report took stock of the implementation of ABM in the Commission and found that there is 'limited information in the Annual Activity Reports on the "lessons learnt" from the information on the performance of the policies. ...The design of the SPP/ABM cycle and the evaluation framework does not yet realise the potential benefits of evaluation results for the effective management of budgetary resources'.[61] One example of an internal audit of a DG evaluation unit confirms that feedback to from evaluations is not processed optimally.

Evaluation was eventually moved to SG in 2009 in order to bring retrospective evaluation closer to policy formulation and thus enhance policy learning. One HoEU explains:

The traditional DG Budget approach had been very much focused on accountability and not so much on policy design. But with the move of the Evaluation Unit from DG Budget to the SG, it was acknowledged that we tried to have different purposes for an evaluation. So, [now] we really do it for accountability purposes but also for learning purposes [and] for policy design purposes.

DG Budget could enforce evaluation of spending programmes (with basis in the FR and the Control Standards), but could not impose regulatory evaluations. SG has the clout to impose regulatory evaluation, which it has been working on ever since in the update of the evaluation guidelines.

An important event for the evaluation system was the parallel introduction of Impact Assessment (IA) in 2002. SG headed the development of IA. IA was introduced as a supplement to ex ante appraisals and

---

56   Toulemonde, *Assessing EU Evaluations*, supra note 49, at p. 78.

57   Ellinas, *Reforming the Commission, Supra* note 32. Mendez, *Administrative reform and unintended consequences*, supra note 11.

58   Toulemonde J, 'Evaluation culture(s) in Europe: differences and convergence between national practices' (2000) 69 Vierteljahrshefte zur Wirtschaftsforschung 350. Batterbury SCE, 'Principles and purposes of European Union policy evaluation' (2006) 40 Regional Studies 179. Eser, Mid-term Evaluations, supra note 34.

59   Williams, *The Use of Evaluation in the Commission Services*, supra note 41.

60   See executive summary in Laat Bd, *Study on the Use of Evaluation Results in the Commission* (2005).

61   Ecorys and COWI, *Study on the state of implementation of Activity Based Management in the European Commission* (2008).

as an important part of the Better Regulation agenda of 2000.[62] In the Mandelkern Group report from 2001, IA was envisaged to play a key role in securing evidence behind policy proposals from the Commission. Quickly it gained popularity in the Commission, because of its usefulness in policy- and decision-making processes. In relation to learning, the FR states that IAs are supposed to be based on earlier evaluation findings. However, in practice this integration was not so easy.[63] One HoEU argues that 'evaluations are relatively disconnected from IAs.' Another HoEU supports this claim: 'so we have two very strict and formalized systems next to one another and up until today (2013), this has not fully been sorted out.' Several interviewees suggest the same and that SG was even more uncertain about the usefulness of retrospective evaluation than DG Budget had been previously.

As a consequence of IA's relevance for decision-making, IA crowded out evaluation in budgetary terms. One HoEU describes that 'the pure evaluation activity in the Commission is actually gone down because of the growth of IA. With all the resources going into IA it was very hard to convince people to carry out [regulatory] evaluations on top of this.' Furthermore, regulatory evaluation is often not a legal obligation written into the legislation, it is often perceived to be unnecessary costly and risky to perform such a large-scale evaluation. Therefore, regulatory evaluation was still working ad hoc in this period as well as in parallel with both the IA-system and the mandatory expenditure evaluations.

Overall, this period is characterized by more focus on enforcing the legality of the evaluation system and thus reinforcing the legal structures that would secure compliance from the DGs in order to finally implement the evaluation system and secure accountability. This was the case when evaluation was introduced in the Commission's control standards. Now the Commission was not only legally obliged to evaluate (legal base in the FR) but could also be audited accordingly by internal audit services in relation to its implementation of the appropriate organisational structures as well as evaluation practices. Hence, legal accountability was the focal point in this period in order to ultimately secure financial accountability, which was the main purpose of the evaluation system in the first place. As a consequence of the focus on financial and legal accountability, policy learning from retrospective evaluation in the evaluation system's early years did not receive much attention. Furthermore, IA reduced the potential for retrospective policy learning by de facto decoupling retrospective evaluation and prospective policy appraisals (ex ante and IA).

4. 2007–2014: Regulatory Evaluation and Fiscal Constraint

In the period from 2007 the Better Regulation agenda was succeeded by the Smart Regulation agenda. The period was characterized by more focus on policy learning from the introduction of systematic regulatory evaluation in the Commission du to an increased emphasis on evidence-based policy making. The Communication on Smart Regulation reads:

> [The Commission] has begun evaluating legislation...[and]... this approach must be extended so that evaluations of legislation become an integral part of smart regulation. Evaluating the effectiveness and efficiency of EU legislation will improve the quality of policy-making.[64]

Moreover, the CoA suggested that 'legislation and other non-spending policies are not yet systematically evaluated. [And that there] is the need to obtain more relevant results to support political decision-making within or across policy areas.'[65] Particularly in relation to policy learning, the SG issued the communication 'Reinforcing the Use of Evaluation',[66] which had as its objectives improvement of the quality of evaluation; linking evaluation to the concept of 'evidence-based policy-making', as well as integrating evaluation further in the 'Strategic Policy and Programming cycle' and ABM-system of the Commission.

---

62 Kassim, Mission impossible, supra note 51. OECD, *Regulatory policies in OECD Counctires: From Interventionism to Regulatory Governance* (2002). Francesco FD, Radaelli CM and Troeger VE, 'Implementing regulatory innovations in Europe: the case of impact assessment' (2011) 19 Journal of European Public Policy 491.

63 Radaelli CM and Meuwese ACM, 'Hard Questions, Hard Solutions: Proceduralism through Impact Assessment in the EU' (2010) 20 West European Politics 923.

64 Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the regions Commission, *Smart Regulation in the European Union*, COM(2010)543.

65 Commission, *Responding to Strategic Needs,* Supra note 8.

66 Commission, *Responding to Strategic Needs, Supra* note 8.

In parallel to this development towards more regulatory evaluation in the Commission, the EP started building its own evaluation capacity and conducting policy evaluations and assessments of Commission IAs. Also the CoA increased its capacity to perform performance audits on EU programmes. The evaluation landscape with regard to regulatory evaluation thus became more diversified in this period forcing the Commission to increase its efforts to produce sound assessments on the effects of legislation. Therefore and inspired by on-going work in DG Markt and DG Infso, the SG started work on new guidelines and new methods to support the evaluation of legislation. The new common guidelines are still in public consultation (in the moment of writing) but will most likely include a new instrument called Fitness Check, which is a type of regulatory evaluation that assess the combination of Community interventions in a particular policy area (e.g. all legal interventions related to waste, including waste treatment etc.).

In 2007, the Commission[67] concluded that evaluation is now an established management tool with regard to expenditure programmes and that learning and evaluation's impact on decision-making needs to be reinforced. This conclusion was based on three reports on evaluation use in the Commission.[68] The communication emphasizes evaluation use to increase learning at the evaluation unit level in the DGs: 'The evaluation function must promote the use of evaluation in decision-making and organizational learning by ensuring that policy implications and lessons learnt from (and across) evaluations are synthesized and disseminated.' This effort is still ongoing and in the latest proposal to common evaluation guidelines, the Commission mainly stresses learning as three out of four objectives with evaluation are learning-oriented ('organisational learning', 'relevant advice to decision-making' and '[input to] efficient resource allocation'). 'Transparency and accountability' is the only objective related to account-

ability.[69] One desk officer from SG stated: 'I think the policy [evaluations] from their origin are going to be more used. Because if you plan a policy evaluation that already means that there was some plan behind the evaluation in order to use the findings for political reasons.'

In relation to accountability, no measures were taken in this period to dis-integrate programme evaluation from the policy cycle or ABM. Therefore, it is unlikely that implementation of programme evaluation and the stress on financial and legal accountability will be undone any time soon . However, as systematic regulatory evaluation is implemented with the new guidelines, democratic accountability is reinforced. As we learned previously, programme evaluation does not receive much attention in the EP, the Council nor in the media. However, regulatory evaluation is likely to get more attention as legislation generally has a more tangible impact on society than expenditure programmes, where only few benefits directly from the support. Also, regulatory evaluations are rare and they therefore generally produce new knowledge that stakeholders and politicians can use strategically for decision-making.

Taken together, expenditure programmes are still managed according to a stress on financial and legal accountability in the Commission. However, the push for regulatory evaluation is linked much more closely with democratic accountability vis-à-vis the EU, the Council and MSs. In this period, a new focus on policy learning has emerged together with regulatory evaluation. However, due to the novelty of regulatory evaluation, further research is needed to elucidate whether regulatory evaluations in fact induce policy learning.

VI. Conclusion

The article accounts for the development of the European Commission's evaluation system. The article shows how internal and external developments shape an evaluation system aiming for both accountability and policy learning. In the system's commencement, financial and legal accountability overshadow the focus on policy learning that was perceived to be in contradiction with the two former. However, the article also demonstrates that accountability and policy learning are not necessarily opposed to each other.

---

67 Commission, *Responding to Strategic Needs, Supra* note 8.

68 Laat, *Study on the Use of Evaluation Results in the Commission, supra* note 60. Williams, *The Use of Evaluation in the Commission Services, supra* note 41. European Court of Auditors, *Presidential Letter* (2005).

69 European Commission, Public Consultation on Commission Guidelines for Evaluation (European Commission 2013).

Table 2 - Overview of implementation history of evaluation in the Commission

		1980–94	1995–99	2000–06	2007–13
Accountability	Focus	Hierarchical	Financial	Legal	Democratic
	Place	Internal - Decentralized	Internal - Centralized	Internal – Centralized	External
Learning	Focus	Organizational- and policy learning.	Incremental programme adjustments.	Incremental programme adjustments.	Policy learning and evidence-based policy making.
	Place	Ad hoc evaluation creates learning on DG-level.	Learning limited to operational- and evaluation units.	Learning limited to operational- and evaluation units.	The Commission, the EP and the Council.
Context		DGs implemented evaluation systems independently of DG Budget, DG Financial Control and SG and according to needs and best practice in their sector field.	Successive NPM reforms institutionalize programme evaluations legally and administratively to increase financial control and the overall legitimacy of the Commission.		Regulatory evaluation tools are developed and focus on learning is increased.
Key events		- Increased budget - NPM paradigm starts in Europe	- SEM2000 - Evaluation linked to policy-cycle - Evaluation units in each DG - MEANS guidelines	- Kinnock reform - Better Regulation - Impact Assessment - Evaluation linked to control standards - SG takes over evaluation from DG Budget	- Evidence-based policy making - EP and CoA policy evaluation capacity - DGs develop regulatory evaluation guidelines - Fitness check

Table 2 is a summary of accountability and policy learning in the four periods that emerged from the data.

The historic approach allows us to see the internal and external developments that shape the evaluation system, as well as the consequences the evaluation system has on accountability and policy learning. However, evidence is still weak and more specific research on the consequences of regulatory evaluation in the Commission is needed to fully grasp the impact on policy learning of these new types of evaluation.