

Højlund, Steven

Doctoral Thesis

Evaluation Use in Evaluation Systems: The Case of the European Commission

PhD Series, No. 4.2015

Provided in Cooperation with:

Copenhagen Business School (CBS)

Suggested Citation: Højlund, Steven (2015) : Evaluation Use in Evaluation Systems: The Case of the European Commission, PhD Series, No. 4.2015, ISBN 9788793155916, Copenhagen Business School (CBS), Frederiksberg,
<https://hdl.handle.net/10398/9117>

This Version is available at:

<https://hdl.handle.net/10419/208922>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/>

COPENHAGEN BUSINESS SCHOOL
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

WWW.CBS.DK

ISSN 0906-6934

Print ISBN: 978-87-93155-90-9
Online ISBN: 978-87-93155-91-6

EVALUATION USE IN EVALUATION SYSTEMS - THE CASE OF THE EUROPEAN COMMISSION

PhD Series 04-2015

Steven Højlund

EVALUATION USE IN EVALUATION SYSTEMS - THE CASE OF THE EUROPEAN COMMISSION

PhD School in Organisation and Management Studies

PhD Series 04-2015



COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Evaluation Use in Evaluation Systems:
The Case of the European Commission

Steven Højlund

Susana Borrás

Ph.D. School in Organisation, Management and Strategy

Copenhagen Business School

Steven Højlund
Evaluation Use in Evaluation Systems
– *The Case of the European Commission*

1st edition 2015
PhD Series 04-2015

© The Author

ISSN 0906-6934

Print ISBN: 978-87-93155-90-9
Online ISBN: 978-87-93155-91-6

The Doctoral School of Organisation and Management Studies (OMS) is an interdisciplinary research environment at Copenhagen Business School for PhD students working on theoretical and empirical themes related to the organisation and management of private, public and voluntary organizations.

All rights reserved.

No parts of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the publisher.

Foreword

This PhD thesis was made possible by the industrial PhD programme offered by the Danish Ministry of Science, Innovation and Higher Education as well as a cooperation between COWI A/S and Copenhagen Business School (CBS). In addition to the PhD programme, COWI, the COWI Foundation and CBS also provided funding.

A number of people at CBS and COWI supported me throughout the three year PhD programme. First, warm thanks go to my primary supervisor, Professor Susana Borrás, who reminded me what an academic contribution is and how to address a theoretical gap in the literature. I thank my co-supervisor, Jesper Strandgaard, who gave me valuable advice on organisational theory and enabled a very productive and life-changing stay at Stanford University. Also, Professor Emeritus Finn Hansson deserves warm thanks for his valuable comments on draft papers.

At COWI A/S, I would like to thank former Head of Department Arne Kvist Rønne, who made the project possible and believed in it from the beginning. Also warm thanks to Niels Eilchow Olesen, who acted as company supervisor in COWI A/S and gave valuable feedback on drafts. Moreover, Claus Rebién and many other good colleagues contributed valuable comments, good conversation and curiosity that made me feel that the project was worthwhile and interesting to others beyond a narrow group of academics and myself.

Finally, I wish to thank the European Commission for producing the documents I requested and for participating in numerous interviews. A special thanks goes to Anne-Louise Friedrichsen and Michael Sørensen in the European Commission, who both opened many doors. Without their help, I would not have accessed some invaluable information.

Abstract (English)

What effect do evaluation systems have on the use of evaluation? This is the research question guiding this PhD thesis. By answering this research question as well as three sub-questions, the thesis addresses three important gaps in the evaluation literature: the first gap is that evaluation theory does a poor job explaining non-use and justificatory uses of evaluation. The second gap is that evaluation theory does not account for the systemic context of evaluation in its more general explanations of evaluation use. Finally, the literature does not account empirically for the micro-level of evaluation use in evaluation systems.

The thesis draws inspiration from organisational theory and in particular organisational institutionalism. Organisational institutionalism explains organisational action and change to be driven by legitimacy-seeking organisational behaviour. Organisations seek to legitimise themselves in order to survive in their environment. This theory is applied to the concept 'evaluation system'. Hence, the assumption underlying this thesis is that organisations within an evaluation system are using evaluations to appear accountable rather than improve policies.

The thesis investigates the European Union's evaluation system with a particular focus on the European Commission. This is done in four articles. The first article is a theoretical article introducing organisational institutionalism to the evaluation literature in order to explain non-use and justificatory uses of evaluations. The second article is a historical analysis of the development of the European Commissions evaluation practices. The third article is a case-based analysis of evaluation use in the European Commission. The fourth article is also an empirical article on policy learning from evaluations in three different Directorate-Generals in the European Commission. The methodology used in the empirical articles is qualitative content analysis and the data were more than a hundred Commission documents and 58 interviews with Commission staff.

The thesis concludes that formal structures are introduced in the Commission to increase oversight of the Commission by other organisations in the system and that evaluation is used to increase accountability in the Commission. Article 2 finds that evaluation is, in fact, primarily institutionalised in the Commission for accountability purposes. The evaluation system is thus set up with a main aim of

securing the legitimacy through accountability for the Commission. Nevertheless, articles 2, 3 and 4 all show that, despite this aim of the evaluation system, there is still room for evaluation use within the framework of the evaluation system's rules and standards. The three main effects of the evaluation system on evaluation use can be summarised as: 1) the 'sacrifice' of process use for findings use and accountability in decision points; 2) a very narrow scope for evaluation use, due to the formal institutionalisation of evaluation; 3) a de-politicisation of evaluation.

First, the possibility of evaluation use in the evaluation process is decreased because of the tightly managed and standardised evaluation process and the stress on evaluator independence that ultimately secures the legitimacy of the evaluation output and the Commission. Process use is sacrificed as a logical consequence of the fact that programme changes are usually attainable only in the design phase of a new programme (and not during its implementation), at which time the Commission needs credible, trustworthy and independent evaluations to increase its own legitimacy as well as that of the new proposal. Second, evaluation recommendations tend to suggest small procedural programme changes rather than large-scale programme changes that only the EP and Council could decide upon. Third, the two previous findings imply a de facto de-politicisation of programme evaluations in the EU evaluation system, where evaluation information conforms to the administrative context of programme management in the Commission instead of the political context of policy-makers.

A number of other findings from the four articles are indirectly linked to the research question. First, the articles all together show the importance of analysing phenomena such as evaluation and evaluation use in their systemic organisational context. When trying to explain evaluation use, the evaluation literature has focused on the evaluation much more than on the context of the evaluation. The main contribution of this thesis is to introduce to the evaluation literature empirically tested assumptions of organisational institutionalism, thereby illustrating that a theory of organisation is better at explaining evaluation uses than evaluation theory. The purpose of the evaluation system is to secure the Commission's accountability. Justificatory use is therefore the most important type of use for the Commission and this *raison d'être* can explain why process

uses are not made possible in the evaluation system and why findings uses are significantly limited to mainly small-scale programme changes.

An important finding of this thesis is that the concept 'evaluation system' needs more theoretical depth. If an 'evaluation system' is defined only in terms of its boundedness, units and institutionalisation then we fail to understand how accountability and organisational effectiveness affects evaluation practices and evaluation use. This thesis shows very clearly how organisational accountability in the system plays an important role in determining how evaluations are used.

Resume (Dansk)

Hvilken effekt har evalueringssystemer på anvendelsen af evalueringer? Det er forskningsspørgsmålet, som besvares i denne ph.d.-afhandling. Ved at besvare dette forskningsspørgsmål samt tre delspørgsmål, adresserer afhandlingen tre vigtige mangler i evalueringslitteraturen: For det første kan evalueringsteori ikke på tilfredsstillende vis forklare, hvorfor evalueringer nogle gange ikke anvendes eller, hvorfor evalueringer bruges til at legitimere organisationer. For det andet tager evalueringsteori ikke højde for den systemiske kontekst, når evalueringsanvendelse forklares. Endelig er litteraturen ikke udviklet empirisk i forhold til at forklare anvendelse af evalueringer i evalueringssystemer på mikro-niveauet.

Teoretisk trækker afhandlingen på organisationsteori og især organisatorisk institutionalisme. Organisatorisk institutionalisme forklarer organisatorisk handlen og forandring ved organisationers behov for at legitimere dem selv vis-à-vis det organisatoriske miljø. Denne teori bringes i anvendelse i relation til begrebet 'evalueringssystem', og det er således antagelsen i denne afhandling, at organisationer inden for et evalueringssystem bruger evalueringer til at legitimere sig selv ved at vise ansvarlighed (accountability) frem for at forbedre politikker.

Afhandlingen undersøger Den Europæiske Unions evalueringssystem med særlig fokus på Europa-Kommissionen. Dette sker i fire artikler. Den første artikel er en teoretisk artikel, som introducerer forklaringer på manglende anvendelse og legitimerende typer anvendelse ved hjælp af antagelser fra organisatorisk institutionalisme. Den anden artikel er en historisk analyse af udviklingen af Europa-Kommissionens evalueringspraksis. Den tredje artikel er en case-baseret analyse af evalueringsanvendelse i Europa-Kommissionen. Den fjerde artikel er også en empirisk artikel om politik læring fra evalueringer i tre forskellige generaldirektorater i Kommissionen. Den metode, der anvendes i de empiriske artikler, er kvalitativ indholdsanalyse og data udgør mere end hundrede Kommissionsdokumenter og 58 interviews med Kommissionsmedarbejdere.

Afhandlingen konkluderer, at de formelle evalueringspraksisser, der er indført i Kommissionen til at øge tilsynet, anvendes primært til at øge Kommissionens ansvarlighed (accountability). Artikel 2 fastslår, at evaluering primært er

institutionaliseret i Kommissionen for at øge organisationens ansvarlighed udadtil. Evalueringssystemet er således oprettet med det formål at sikre legitimitet gennem ansvarlighed (accountability) for Kommissionen. Ikke desto mindre, viser artikel 2, 3 og 4, at evalueringer stadig anvendes inden for rammerne af meget snævre rammer i evalueringssystemet. De tre vigtigste effekter af evalueringssystemet i forhold til evalueringsanvendelse kan opsummeres som: 1) en 'ofring' af procesanvendelse for at forbedre anvendelse af evalueringsresultater og legitimerende anvendelsestyper; 2) evalueringer anvendes inden for et meget snævert anvendelsesområde på grund af den formelle institutionalisering af evaluering; 3) der sker en afpolitisering af evalueringer i evalueringssystemer.

Vi kan uddybe disse tre overordnede konklusioner. For det første er muligheden for evalueringsanvendelse i evalueringsprocessen reduceret på grund af den stramt styrede og standardiserede evalueringsproces og fokus på evaluators uafhængighed, som i sidste ende sikrer evalueringens- og Kommissionens legitimitet. Procesanvendelse ofres som en logisk konsekvens af, at programændringer normalt kun er opnåelige i designfasen af et nyt program (og ikke under gennemførelsen). I designfasen har Kommissionen behov for troværdige, pålidelige og uafhængige evalueringer for at øge sin egen legitimitet såvel som legitimiteten af det nye forslag. For det andet har evalueringer i Kommissionen tendens til at anbefale små proceduremæssige programændringer snarere end store programændringer. For det tredje indebærer de to ovenstående resultater at der sker en de facto afpolitisering af programevalueringer i EU's evalueringssystem, da informationer fra evalueringer tilpasses til den administrative kontekst i Kommissionen i stedet for den politiske kontekst i Europa-Parlamentet.

En række andre resultater fra de fire artikler er indirekte knyttet til forskningsspørgsmålet. For det første viser artiklerne alle sammen, hvor vigtigt det er at analysere fænomener som evaluering og evalueringsanvendelse i deres systemiske og organisatoriske kontekst. Denne afhandlings væsentligste bidrag er at introducere empirisk testede antagelser fra organisatorisk teori til evalueringslitteraturen og dermed illustrere, at en teori om organisation er bedre til at forklare evalueringsanvendelse end standard evalueringsteori.

Et andet vigtigt resultat af denne afhandling er desuden at tilføre begrebet 'evalueringsystem' mere teoretisk dybde. Hvis et 'evalueringsystem' kun defineres i form af sin afgrænsning, aktører og institutionalisering, så bidrage det stadig ikke meget til at forstå, hvordan ansvarlighed (accountability) og organisatoriske effektivitet påvirker evalueringspraksis og evalueringsanvendelse. Denne afhandling viser meget tydeligt, hvordan organisatorisk ansvarlighed (accountability) i systemet spiller en vigtig rolle i forhold til at forklare, hvordan og hvorfor evalueringer anvendes.

<u>1</u>	<u>INTRODUCTION</u>	10
1.1	BACKGROUND	10
1.2	TOWARDS A RESEARCH QUESTION	12
1.3	LITERATURE REVIEW	17
1.4	CONTRIBUTIONS AND COHERENCE	21
1.5	DATA	28
1.6	INDUSTRIAL PhD THESIS	29
<u>2</u>	<u>THEORETICAL FRAMEWORK</u>	31
2.1	COMMON ASSUMPTIONS IN THE EVALUATION LITERATURE	31
2.2	ORGANISATIONAL INSTITUTIONALISM	33
2.3	KEY THEORETICAL ASSUMPTIONS	36
2.4	RESEARCH QUESTION	38
<u>3</u>	<u>CONCEPTUAL FRAMEWORK</u>	41
3.1	EVALUATION	41
3.2	EVALUATION SYSTEMS	45
3.3	EVALUATION USE	51
<u>4</u>	<u>OBJECT OF STUDY</u>	65
4.1	CHOICE OF CASE	65
4.2	THE EU EVALUATION SYSTEM	66
<u>5</u>	<u>RESEARCH DESIGN AND METHODOLOGY</u>	68
5.1	CASE STUDIES	68
5.2	ARTICLES	72
5.3	METHODOLOGY	77
<u>6</u>	<u>CONCLUSION</u>	83

6.1	ANSWERING THE SUB-QUESTIONS	83
6.2	ANSWERING THE RESEARCH QUESTION	88
6.3	PERSPECTIVES ON FINDINGS	91
7	REFERENCES	94
8	APPENDIX – ARTICLES	108
8.1	ARTICLE 1 – EVALUATION USE IN THE ORGANIZATIONAL CONTEXT – CHANGING FOCUS TO IMPROVE THEORY	109
8.2	ARTICLE 2 – EVALUATION IN THE EUROPEAN COMMISSION - FOR ACCOUNTABILITY OR LEARNING?	132
8.3	ARTICLE 3 – EVALUATION USE IN EVALUATION SYSTEMS – THE CASE OF THE EUROPEAN COMMISSION	159
8.4	ARTICLE 4 – EVALUATION AND POLICY LEARNING – THE LEARNERS’ PERSPECTIVE	198

1 INTRODUCTION

I start by explaining the background to this PhD thesis and the reasons I spent three years studying evaluation and the European Commission – two things that for most people would provoke boredom at best and aversion more likely. Then I clarify how a puzzle developed into a research question, how far the literature on evaluation has answered the research question and which parts of the story are still untold. Finally, this introduction summarises the findings of this thesis and their significance.

Chapters 2, 3 and 4 describe the theoretical and conceptual framework of the thesis. Chapter 4 describes the object of study and Chapter 5 expands on the research design and methodology. Chapter 6 concludes this thesis by answering the research question and putting the findings into perspective.

1.1 BACKGROUND

About four years ago in Copenhagen, three consultants in the Danish consultancy COWI, were pondering the meaning of their work. They worked primarily with evaluations of European Union policies and spending programmes. But they did not know if the European Commission used the evaluations they produced. One day they decided to find out. And so this industrial PhD project came into being, as collaboration between COWI and Copenhagen Business School with support from the Danish Ministry of Higher Education and Science. The project was undertaken to shed light on evaluation use in the Commission.

As it turned out, the consultants' puzzle over evaluation use was shared by the evaluation community in relation to all types of evaluating organisations. In fact, the phenomenon of evaluation use had been studied empirically since the 1970s with widely varying results. Some studies described how evaluations so rarely change policies that a general evaluation 'utilization crisis' was pronounced by Michael Q Patton (Patton, 1997; Pawson and Tilley, 1997: 2-3). Other studies revealed extensive use of evaluations (Leviton and Hughes, 1981; Shulha and Cousins, 1997; Johnson et al., 2009a). So what to believe?

Two aspects made this question difficult:

First, the literature is mainly empirical and under-theorised. It comprises many case studies that are hard to generalise from. Attempts to build theory from the many observations also stay conceptual. Without a theory of evaluation use, we have to resort to the theory behind evaluation itself. The main theory assumes that evaluations are carried out so that public policy can be improved. According to this theory, evaluations are used because public organisations have implemented and institutionalised them (see for example Mark and Henry, 2004: 38; Cousins and Leithwood, 1986; Pawson and Tilley, 1997). But that does not explain why evaluations are not always used.

Second, no evaluation theory adequately explains the non-use of evaluations. The literature is merely empirical, and highlights only that evaluation use sometimes does not take place. There are few attempts to build a theory of non-use of evaluation. In other words, we can understand non-use of evaluation only in terms of the general evaluation theory described above. Probably unaware, Patton perceives non-use of evaluation to be so unusual that he proclaims a 'utilization crisis'. Maybe, if utilisation is in crisis when it does not materialise, we expect utilisation to happen too often. In other words, non-use of evaluation is considered an empirical curiosity or abnormality. Despite 40 years of research, no generalised assumptions can yet explain and predict the phenomenon properly.

It constitutes a theoretical problem for evaluation theory that it cannot explain its non-use; it is also a problem that there is no credible theoretical framework that can. In practical terms, evaluation theory's predictions of evaluations being used to improve public policies and interventions is paradoxical when, in fact, this is not always the case. It would be pertinent to ask: 'Why evaluate with the objective of improving policy if the evaluation is not used afterwards?'

The focus of this thesis is the European Union's evaluation system and, in particular, the European Commission, where most of the EU evaluation activity takes place. Over the last 30 years, the Commission institutionalised evaluation practices in response to internal and external pressures (see Article 2). Previously, very little attention was given to the Commission's evaluation practices. This is unfortunate, since the Commission has had a significant impact on evaluation

practices and the setting up of evaluation systems in public administrations across Europe through EU programme conditionality (Toulemonde, 2000; Toulemonde et al., 2005; Olejniczak, 2013). Therefore, this thesis focuses on the European Commission evaluation system's effects on evaluation use with a particular stress on the Commission's role.

1.2 TOWARDS A RESEARCH QUESTION

Before introducing the research question, it is important to address the gaps in the literature as well as the underlying hypotheses that are the foundation for proposing the research question. In section 2, the theoretical assumptions and the research question are described in more detail.

1.2.1 GAPS IN THE LITERATURE

This thesis addresses three gaps in the literature on evaluation. The first gap is a theoretical gap, and concerns the problem that evaluation theory cannot explain non-use of evaluation and justificatory uses of evaluation.

Theories of organisation have, for many years and with great success, explained why organisations do not act upon the knowledge they collect through means such as evaluations. Organisational institutionalism, in particular, has demonstrated how organisations have adopted evaluation to seek legitimacy in the organisational field. This thesis aims to fill this gap by proposing a theoretical framework that can explain why organisations do not use or learn from evaluations, and why organisations often use evaluations to justify themselves or the intervention in question. The thesis addresses the gap in the literature by applying organisational institutionalism to explain non-use of evaluation. Assumptions are drawn from this theory to formulate a key hypothesis and the overall research question (see Section 2.4).

The second and related gap in the literature exists because the literature has not sufficiently incorporated ideas from organisational theory to explain evaluation use more generally in order to include the organisational context of evaluation. The evaluation literature has tried to focus on the organisational context by focusing on concepts such as 'evaluation system'. However, this conceptualisation has never developed into a genuine theory of evaluation systems and the ways in which they relate to phenomena such as evaluation use.

The third gap relates to the lack of empirical evidence concerning evaluation use in evaluation systems – in particular, the effect that the evaluation system has on evaluation use and learning at the micro-level and in relation to the users and uses of evaluation. This thesis uses the concept ‘evaluation system’ to explain the context of the evaluation (see Section 3.2 for an introduction to evaluation systems). A particular empirical focus that considers the special dynamics of an evaluation system is necessary to understand evaluation use in highly institutionalised contexts.

To sum up, three gaps are addressed in this thesis.

Box 1-1 Gaps addressed by the thesis

First gap: Evaluation theory cannot explain non-use and justificatory uses of evaluation.

Second gap: Evaluation theory does not account for the systemic context of evaluation in its more general explanations of evaluation use.

Third gap: The literature does not account empirically for the micro-level of evaluation use in evaluation systems.

Table 1-1 below contains an overview of the thesis’ four articles and how they address the three gaps.

Table 1-1 Overview of the way gaps are addressed in the articles

Article number	How gaps are addressed
Article 1	Gaps 1 and 2 are addressed by proposing a theory of the organisation and its context that will account for various types of evaluation use. The proposed theory adapted from organisational institutionalism assumes that evaluation use is dependent on the organisation's internal propensity to evaluate, and the external pressures on the organisation.
Article 2	Gaps 2 and 3 are addressed as article 2 analyses the implementation of the EU evaluation system. The article breaks down accountability into four different types and illustrates how learning from evaluation and accountability are not necessarily mutually exclusive. Therefore, article 2 explains evaluation outcomes with the organisational context (secondary gap) while also explaining under which circumstances non-use of evaluation outcomes is likely.
Article 3	Article 3 addresses all three gaps by demonstrating empirically how important the organisational context (in the form of the evaluation system) is in order to determine types of evaluation use.
Article 4	Article 4 addresses gaps 1 and 3 by demonstrating empirically how important the organisational context (in the form of the evaluation system) is in order to determine different types of learning.

1.2.2 ASSUMPTIONS AND HYPOTHESIS

This thesis draws extensively on organisational theory and in particularly organisational institutionalism (Meyer and Rowan, 1977; DiMaggio and Powell, 1991). Organisational institutionalism proposes assumptions about organisational behaviour that are supported by empirical findings (see Section 2.3 for a more detailed explanation of assumptions). The general assumption is that organisations

seek to legitimise themselves in order to survive in the organisational field. They do so by adopting institutions or norms of behaviour. One such institution is evaluation including norms and values, which are adopted to support claims of accountability in the evaluation system.

Therefore, and in accordance with organisational institutionalism, the general **hypothesis is that public organisations implement evaluation to generate accountability.**

Below, I will explain the content of the hypothesis. First, evaluation is linked mostly with public organisations. This follows from the definition of evaluation used in this thesis (see Section 3.1). Therefore, this thesis analyses accountability in relation to public organisations (see Section 2). In this thesis I focus on accountability rather than legitimacy, because accountability is more specific in relation to evaluation's function in public administration. Because this thesis focuses on the evaluation system and the distribution of power and control among organisations in the system, accountability is a more appropriate concept than legitimacy (see Section 3.2 for further elaboration).

1.2.3 RESEARCH QUESTION

In this subsection, I present the research question of the thesis. Section 2.4 explains the research question in greater detail and explains its importance. Following from the discussion of the gaps in the literature and the theoretical assumptions, the research question becomes:

What effect do evaluation systems have on the use of evaluation?

The research question contains three important elements (see Section 2.4 for further explanation): 'evaluation system', 'evaluation use' and 'effect'.

The 'evaluation system' concept is the study object of the thesis, together with evaluation use. It signifies the organisational context that organisational institutionalism assumes can explain organisational behaviour and, therefore, organisations' use of evaluations. In this thesis, 'evaluation system' is used as a concept to capture the organisational interdependency in a system of

organisations, while relating this interdependency to evaluation practices and outcomes. This is captured in my definition of evaluation system: ‘Evaluation system’ is understood as permanent and systematic evaluation practices taking place and institutionalised in several interdependent organisational entities with the purpose of generating accountability and informing decision-making’ (Højlund, 2014a) (see Section 3.2). This definition builds on existing definitions and in particular on Leeuw and Furubo (2008).

The concept ‘evaluation use’ is the object of study of the thesis together with evaluation systems. Evaluation use refers to the use of evaluations, including the learning that takes place during and subsequent to evaluation processes. In Section 3.3, evaluation use is described in more detail, including the many types of evaluation use. The concept ‘evaluation use’ is to be understood broadly to include all types of uses. It is a key finding of this article that non-use as well as justificatory uses of evaluations are also to be included as important and central use types.

The concept ‘effect’ refers to the relationship between evaluation system and evaluation use. I am interested in understanding this relationship; in particular, whether the evaluation system has an effect on evaluation use.

To answer the research question, I propose three auxiliary sub-questions that address more specifically the gaps in the literature:

- 1) *How can non-use and justificatory uses of evaluation be explained?*
(addressed in articles 1, 3 and 4)
- 2) *How can evaluation use be explained in its systemic organisational context?*
(addressed in articles 1, 2, 3 and 4)
- 3) *How are evaluations used in evaluation systems?* (addressed in articles 2, 3 and 4)

The next section contains a review of the evaluation literature in relation to the concepts of evaluation system and evaluation use.

1.3 LITERATURE REVIEW

In this thesis, three gaps in the evaluation literature are addressed. All of them relate to evaluation theory and its potential for improvement. Therefore, this literature review starts by focusing on existing evaluation theory and its assumptions regarding evaluation use. Then it reviews the attempts the evaluation literature has made to integrate organisational theory and, in particular, organisational institutionalism in order to answer questions related to evaluation use. The review also looks at how the literature has developed in order to gauge the theoretical depth of this emerging line of thinking in the evaluation literature. The concept of 'evaluation system' represents another attempt to bring contextual factors into evaluation theory. But does the present literature on evaluation systems also speak about evaluation use? Finally, the review assesses the empirical contributions of the EU evaluation system.

The concept 'evaluation' first appeared in the 1950s, linked with a common belief that society could be engineered through large public spending programmes and infrastructure projects (Vedung, 2010). Targeted interventions would improve life for everyone and evaluation would ensure that decision-makers learned from earlier mistakes or successes (Porter, 1995). Underlying this thinking were rationalist and economic assumptions about human behaviour containing the underlying positive and evolutionary assumption of progress and betterment (Henry, 2004; Dahler-Larsen, 2012). According to Henry and Mark, the ultimate objective of evaluation is social betterment, because evaluation helps policy-makers make better policies, and in turn those policies improve people's lives. In other words, part of this positivist paradigm was a belief in the ability of policy-makers and administrators to learn lessons from previous interventions and thus to constantly improve the quality, efficiency and effectiveness of public spending and interventions (Vedung, 1997). The inherent logic of mainstream evaluation theory is therefore realist and rational, and perfectly aligned with the economic theory and theories of rational choice from which it arose (Albæk, 1995; Van der Knaap, 1995; Schwandt, 1997; Sanderson, 2000).

Theoretically, the rational approach to government intervention was reflected in David Easton's system theory of policy-making (Easton, 1965). In Easton's model, evaluation relates to the feedback that policy-makers receive as input

toward improved policies. But as this rational view on policy-making and public interventions proliferated, particularly in political science and economics, sociologists studying evaluative and scientific knowledge were more sceptical about whether this learning and evaluation use actually took place (Lazarsfeld et al., 1967). Though Easton's policy model assumes feedback, little evidence supported this assumption. In fact, subsequent research into the use of scientific knowledge and evaluation often illustrated how existing knowledge was not used to improve policies. This phenomenon was referred to by some scholars as a 'utilization crisis' (Patton, 1997).

This concern over non-use of evaluation findings prompted the emergence of a large body of literature related to evaluation use. It was derived from literatures on the use of scientific results in policy making (Lazarsfeld et al., 1967; Porter, 1995; Vedung, 2010; Weiss, 1998; Weiss and Bucuvalas, 1980). In fact, evaluation use is probably the most researched theme in the literature on evaluation (Christie, 2007: 8; Henry and Mark, 2003: 294). At present, the substantial 'evaluation use' literature exists independently of the literature on the use of scientific results and knowledge, mainly because the field of evaluation is a relatively specific practitioner field.

In the wake of the disenchantment over the scarce evidence of evaluation use, the literature sought reasons for it (Leviton and Hughes, 1981; Cousins and Leithwood, 1986; Johnson et al., 2009a). After the evaluation literature had more or less converged on a typology of four key uses (see Section 3.2), research was then dedicated to explanatory variables and to answering the question that interests all evaluators: What makes my evaluation useful? This resulted in a large number of case studies that inferred relationships between contextual variables or conditions and several different use types. This research, which undoubtedly comprises more than one thousand case studies, has been summarised in reviews: (Burry et al., 1985; Beyer and Trice, 1982; King and Thompson, 1983; Thompson and King, 1981; Shulha and Cousins, 1997; Leviton and Hughes, 1981; Cousins and Leithwood, 1986; Cousins et al., 2004; Johnson et al., 2009a). As the reviews make clear, the literature focused on factors related to the attributes of the evaluation (e.g., methodology, quality, relevance of findings, etc) or the immediate contextual factors pertaining to the organisation in which the

evaluation was implemented (e.g., political climate, timing of the evaluation relative to decision-making, etc). These categories were empirically informed from the late 1970s and onwards (see for example Leviton and Hughes, 1981).

While research has clearly and comprehensively explained evaluation use, the explanations of non-use or justificatory use (sometimes called misuse) are not as satisfactory (Højlund, 2014b). There is a large body of empirical literature illustrating – but not adequately explaining – the phenomenon of non-use. This is mainly because the literature focuses on the evaluation and not very often on its context (Van der Knaap, 1995). This thesis addresses this gap by seeking an explanation of non-use and explanatory uses in organisational sociology.

The majority of literature explaining evaluation use focuses on identifying uses as well as factors and conditions relevant to evaluation use. Though some examples exist, it is rare to find any work that attempts a systematic reflection on the organisational context and the implications for evaluation use. Van der Knaap (1995) and Sanderson (2000) are two important voices calling for a contextual analysis of evaluation research. Earlier, Levin (1987) found that contextual factors were highly important in explaining evaluation use, and Shulha and Cousins (1997) later produced an important review, concluding that contextual factors and organisational contexts were becoming more prominent in the literature. Thereby, the contextual characteristics of the evaluation, such as the decision-making or policy setting, became a focal point in the 1990s. However, the research community realised that it was not enough to describe different types of use and to catalogue their contributing factors (Shulha and Cousins, 1997: 197).

Researchers started to theorise about the context of evaluation use more broadly. This resulted in many theory-building attempts. One of the most significant concepts is contained in the concept of the ‘evaluation system’ (Rist and Stame, 2006; Leeuw and Furubo, 2008). The literature on evaluation has been increasingly interested in the question of evaluation systems and their effects on evaluation use. But only recently has it started investigating the effects of substantial formal and informal institutionalisation of evaluation practices on evaluation use. The discussion was initiated with publication of *From Studies to Streams*, edited by Ray C Rist and Nicoletta Stame (Rist and Stame, 2006) and

was continued in a small number of other studies (Williams and Imam, 2007; Leeuw and Furubo, 2008).

However, the literature on evaluation systems both lacks a sound theoretical framework and empirical evidence related to evaluation systems' effects on evaluation use. Evaluation systems are generally assumed to have a negative impact on the use of information and knowledge in policy-making (Power, 1997; Leeuw and Furubo, 2008; Dahler-Larsen, 2012; Pollitt et al., 1999; Furubo, 2006). Previous studies suggest that evaluation use tends to be made relevant primarily for administrators and not for policy-makers, and that use in administrations will be linked to procedural assurance and legitimisation of the organisation rather than policy-making (see also Furubo, 2006; Langley, 1998). However, empirical research of evaluation systems at the micro-level is still needed to increase our understanding of their role with regard to the use of evaluations.

Despite the focus on evaluative context, the evaluation literature has only to a limited extent looked to mainstream organisational theory (such as organisational institutionalism) for explanatory frameworks by which to explain evaluation use. Shulha and Cousins (1997) and Van der Knaap (1995) point to the emergence of 'context' in relation to explaining evaluation use; however, no theoretical framework has since been developed. Peter Dahler-Larsen has written extensively on evaluation from an institutional and constructivist perspective (Dahler-Larsen and Krogstrup, 1998). He analyses organisations' adoption and subsequent ritualisation of evaluation as an institution. However, Peter Dahler-Larsen does not explicitly address the use of evaluation results. When reviewing the evaluation literature, it is clear that since Cousins' call for more contextual explanations for evaluation use, there has been very little written on this subject. A bibliographical search in the leading evaluation journals found only a few articles that analyse evaluation and its institutional role (Barnes et al., 2003; Hanberger and Schild, 2004; Varone et al., 2005; Jacob, 2005). Only two articles focus on the influence of institutionalised evaluation practices on performance and evaluation use (Sager and Rissi, 2011; Eckerd and Moulton, 2011). Only the article by Eckerd and Moulton (2011) specifically explains evaluation use by institutional logics in the organisational context.

Thus, the literature falls short on three main issues. First, it fails to adequately explain why evaluations are sometimes not used at all, or why evaluations are often used to justify an organisation or its interventions. Second, the literature fails to explain how evaluation use can be explained by the systemic organisational context. Finally, there is very little empirical research on evaluation use in an evaluation system.

Academic interest in evaluation systems has increased as evaluation systems in national and international public administrations have increased. As part of a general trend that began in the 1990s, the European Union has also developed an evaluation system, with the European Commission as the main organisation. Despite this evaluation system's increasing influence on the evaluation practices of EU Member States and 'third countries', surprisingly little academic attention has been paid to the EU's evaluation system. Only a few academic studies exist (Toulemonde, 2000; Furubo et al., 2002; Toulemonde et al., 2005; Hoerner and Stephenson, 2012; Mendez and Bachtler, 2011; Eser and Nussmueller, 2006), and the main empirical works are Commission-sponsored consultant reports (Williams et al., 2002; Laat, 2005). On EU evaluation more generally, a small body of literature exists, in particular on evaluation of the Structural Funds or other EU programmes (Eser and Nussmueller, 2006; Toulemonde et al., 2005; Eureval-C3E, 2006; Stern, 2009; Summa and Toulemonde, 2002; Olejniczak, 2013; Ferry and Olejniczak, 2008). This thesis also addresses this shortage of research on the EU evaluation system. The section below explains the contributions of the thesis as well as the coherence of the four articles.

1.4 CONTRIBUTIONS AND COHERENCE

This section elaborates on the contributions of this thesis as well as on the coherence of the four articles that constitute the main work of the thesis (see the Appendix).

1.4.1 COHERENCE OF ARTICLES

This thesis consists of four articles, each of which makes a unique contribution to our knowledge about evaluation systems and evaluation use. Table 1-2 provides an overview of the articles and their findings.

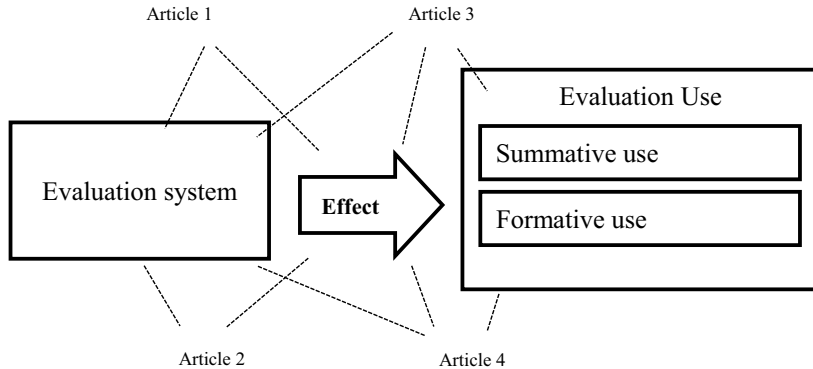
Table 1-2 Overview of articles

Number	Title	Research questions	Findings	Publication status
Article 1	'Evaluation use in the organisational context – changing focus to improve theory'	How can evaluation use be explained by factors that are contextual to the evaluating organisation rather than contextual to the evaluation? How can non-use and justificatory use types be integrated in a model of evaluation use?	The article demonstrates how organisational institutionalism can explain justificatory uses of evaluation as well as non-use. The article provides a short review of the application of institutional explanations in the evaluation literature. We learn that: 1) justificatory use types and non-use need to be better integrated into a theory of evaluation use; 2) that a theory of evaluation use should take into consideration the organisational and institutional context of the evaluating organisation; and 3) that the assumptions of organisational institutionalism can explain non-use and justificatory uses.	Published in <i>Evaluation</i> , 20(1)
Article 2	'Evaluation in the European Commission – for accountability'	Evaluation in the European Commission – For accountability or learning?	The article accounts for the development of the European Commission's evaluation system. By adopting a historical approach, the article shows how internal and external developments that shape the evaluation system also have consequences for accountability concerns and policy learning – understood as two justifications for the evaluation system. We learn that policy learning stood a better chance when associated with certain types of accountability in the early and later stages of the evaluation system's development, and that justifications	Under review for publication in a special issue of <i>European Journal of</i>

	or learning?'		related to accountability are not necessarily opposed to justifications regarding policy learning.	<i>Risk Regulation</i>
Article 3	'Evaluation use in evaluation systems – the case of the European Union'	Are evaluation systems conducive to evaluation use?	The article finds that the European Commission's evaluation system is conducive to evaluation use of programme evaluations in several ways. Evaluations are used instrumentally to improve the programme, albeit mainly at the programme management level and only within the narrow limitations of the LIFE programme Regulation and the general legal framework the Commission functions by. The alignment of programme evaluations to the policy cycle often makes evaluations partly or completely redundant for programme management and policy preparation. We learn that there is a de facto de-politicisation of programme evaluations in the EU evaluation system, where evaluation information conforms to the administrative context of the programme management in the Commission instead of the political context of policy-makers.	Published in <i>Evaluation</i> , 20(4)
Article 4	'Evaluation and Policy Learning – the Learner's Perspective'	Are evaluation systems conducive to policy learning?	This article examines how evaluation in the context of the evaluation system induces policy learning. Taking the case of three programme evaluations in the European Commission, the article examines the patterns of policy learning that emanate from evaluations and their context in the evaluation system. We learn that programme units and external evaluators learn from programme evaluations, and that learners learn different things including things related to a programme overview, small-scale programme adjustments, policy change, and evaluation methods.	Published in <i>'European Journal of Political Research'</i> , 53(5)

Figure 1 illustrates the coherence between the four articles. It presents the causal relationship between evaluation system and evaluation uses (summative and formative). The dotted lines show which phenomena each article analyses.

Figure 1 Coherence of the four thesis articles



The four articles investigate different aspects of the relationship between the evaluation system and evaluation use. Thus, articles 1 and 2 investigate the context of the evaluation system theoretically (article 1) and empirically (article 2), taking the Commission as its case. Articles 3 and 4 investigate evaluation use from evaluations in the context of the Commission's evaluation system. The following expands on the content of each of the articles.

The *first article* is a theoretical contribution that investigates the relationship between evaluation use and the evaluation's organisational context. It lays down theoretical foundations for the other three articles, which are more directly focused on the concept of evaluation system. The article addresses gaps 1 and 2 by proposing a theoretical framework by which to understand non-use of evaluation, while also accounting for the evaluative context in which evaluation takes place. The article finds that the literature on evaluation use has been very good at describing the evaluation, its conditioning factors, while neglecting the organisational context in which the evaluating organisation operates, as well as the organisation's ability to evaluate. This is done more specifically by introducing

institutional theoretical explanations to explain organisational behaviour and different types of evaluation use.

The *second article* extends this by analysing the motives behind the implementation of evaluation practices and the creation of an evaluation system. In doing so, the article shows how the development of an evaluation system is a response to an organisational need to increase primarily accountability but also learning. Therefore, article 2 explains learning with the evaluation system while also explaining under which circumstances non-evaluative outcomes are likely. Thus, the article addresses gaps 2 and 3.

The *third article* provides empirical evidence of evaluation use in the European Union's evaluation system. It demonstrates empirically how important the organisational context (in the form of the evaluation system) is when determining types of evaluation use. The article addresses all three gaps. Both the third and fourth articles are empirical investigations of the effect that the evaluation system has on evaluation use in the context of the EU evaluation system.

Finally, the *fourth article* investigates how evaluation systems affect policy learning when it is understood as a type of evaluation use. Similarly to article 3, article 4 also demonstrates empirically how important the evaluation system is in determining types of evaluative outcomes. The article addresses gaps 1 and 3.

1.4.2 CONTRIBUTIONS

The thesis fills three gaps in the literature and addresses a lack of empirical research on the EU's evaluation system. The contributions are described in the subsections below.

1.4.2.1 EXPLAINING NON-USE AND JUSTIFICATORY USES

The contribution of this thesis is the introduction of organisational institutionalism to the evaluation literature in order to explain the phenomenon of non-use of evaluation. The thesis (mainly article 1) points out and illustrates the painful paradox of evaluation theory's inability to explain why evaluations are not used, despite the fact that the purpose of conducting evaluations is to use them to improve policies (Højlund, 2014b). Empirically, non-use of evaluation is already supported by evidence, but the literature lacks a theory that can explain it. With

the contribution of this thesis, non-use of evaluation is theoretically explained by focusing on the evaluation in its systemic context and non-use of evaluation is assumed a priori.

This contribution is made theoretically in article 1 (Højlund, 2014b) and supported by empirical research in articles 3 (Højlund, 2014a) and 4 (Borrás and Højlund, 2014). The theoretical contribution is the explanation of non-use of evaluation with the explanatory framework and theoretical assumptions of organisational institutionalism. By assuming a need for organisations to legitimise themselves, evaluation is understood in its context of the evaluation system. In the evaluation system, organisations (the Commission) use evaluations to seek accountability in the system. Accountability-seeking behaviour does not exclude the organisation from using evaluations per se, but in the evaluation system an organisation can be obliged to evaluate in order to appear legitimate and survive in the long-term in that system.

1.4.2.2 EXPLAINING EVALUATION USE WITH THE ‘EVALUATION SYSTEM’

The thesis also addresses a second gap in the evaluation literature: that the literature does not account for the context of evaluation when explaining evaluation use. More specifically, previous contributions on evaluation systems do not account for the effect these systems have on evaluation use.

When explaining evaluation use, the literature has focused on the evaluation itself much more than its context. The contribution of this thesis is to introduce empirically tested assumptions of organisational institutionalism to the evaluation literature, thereby illustrating that a theory of organisation is better at explaining evaluation uses than evaluation theory. In particular, these assumptions are added to existing concepts of evaluation systems to increase their theoretical depth. The concepts were first formulated by Michael Scriven (Scriven, 1967) and have since been a common reference point in the literature. However, the summative and formative debate in the evaluation literature has long been under-theorised. This thesis is a credible attempt to add to this debate by emphasising evaluation’s summative role through the assumptions of organisational behaviour drawn from organisational institutionalism.

It is a major contribution of this thesis to expand the concept of ‘evaluation system’ so that it can explain evaluation use. This contribution is important, in that it translates the assumptions and concepts of organisational institutionalism into existing evaluation theory, thus theoretically invigorating the longstanding conceptual dichotomy in the evaluation literature between accountability and learning.

The concept ‘evaluation system’ existed already in the evaluation literature, but it was merely conceptual and lacked theoretical depth (see section 3.2 for an introduction). By linking the assumptions of organisational institutionalism with the evaluation system, the concept gets the necessary ‘theoretical depth’ to explain phenomena such as evaluation use. This is done theoretically in article 1 (Højlund, 2014b) and empirically in articles 2 (Højlund, Forthcoming), 3 (Højlund, 2014a) and 4 (Borrás and Højlund, 2014).

1.4.2.3 EMPIRICAL EVIDENCE OF EVALUATION SYSTEMS’ EFFECT ON EVALUATION USE

The third gap relates to lack of empirical evidence of the effects of evaluation systems on evaluation use. Articles 2 (Højlund, Forthcoming), 3 (Højlund, 2014a) and 4 (Borrás and Højlund, 2014) each describe in their own way the effect of the EU’s evaluation system on evaluation use including learning. Articles 3 and 4 provide a rare insight into the micro-level of users of evaluations in the European Commission and beyond. Focusing on the users and the uses of evaluations illustrates the effects of the evaluation system’s formal and informal institutionalisation. Thus, article 4 finds that different types of actors learn differently from evaluations, depending on their position in the evaluation system.

1.4.2.4 EMPIRICAL CONTRIBUTIONS

The thesis also makes several very important empirical contributions, described below.

First, the literature on evaluation systems is primarily conceptual and not particularly empirical (Leeuw and Furubo, 2008; Mendez and Bachtler, 2011). This thesis analyses the EU evaluation system as its case in order to remedy this empirical gap in the literature. Articles 2, 3 and 4 take the Commission – the most

important organisation in the system – as their starting point in the EU evaluation system. For example, the third article improves our understanding of the implications of an evaluation system on evaluation use, and finds that formal and informal institutions both impede and enable the use of evaluation.

Second, the EU evaluation system has not previously been scrutinised academically. The thesis focuses on the EU evaluation system and, in particular, the European Commission, where most of the system's evaluation activity takes place. The case is interesting for its novelty. The Commission has had a significant impact on evaluation practices and the setting up of evaluation systems in public administrations across Europe through EU programme conditionality (Toulemonde, 2000; Furubo et al., 2002; Toulemonde et al., 2005; Hoerner and Stephenson, 2012; Mendez and Bachtler, 2011; Eser and Nussmueller, 2006). Therefore, it is important to study how and why the Commission evaluates. The project also accounts for the development and implementation of the Commission's evaluation system (see article 2).

Third, evaluation practices in the Commission have been addressed only by consultancy reports and not by systematic academic inquiry. This thesis analyses both the Commission's adoption of evaluation as practice as well as the effects of the practices on evaluation use. This is done primarily in articles 2 and 3, but also in article 4.

1.5 DATA

The thesis is based on empirical analysis of 58 recorded interviews, two group interviews and one conference on evaluation in the EU, along with numerous informal talks with experts and Commission desk officers, as well as personal observations including evaluation steering committee meetings. Interviewees were sampled purposefully and according to availability, and 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 who played key roles in the early implementation of the evaluation system and who thus had a good historical overview of evaluation in the Commission. Interview data were validated with document data comprising more than a

hundred public and non-public documents, such as internal evaluation policy papers, guidelines, minutes of meetings in the evaluation network and so on.

1.6 INDUSTRIAL PHD THESIS

This thesis is funded by the industrial PhD programme. In Denmark, an industrial PhD thesis must meet the same requirements as a conventional PhD thesis. However, the industrial PhD candidate is employed by a company, which is both an advantage and a challenge. In this case, it was an advantage because it facilitated access to the European Commission subsequent to a large ex post evaluation conducted for DG Environment by COWI, where I worked as a consultant before shifting position to become a PhD candidate in the same company. Without this contact the third article would never have been possible. Moreover, my previous experience in the field and with the Commission made it possible to write a better thesis. On the other hand, this contact raises conflict of interest issues, both in relation to science and the Commission.

The issue of conflict of interest is standard to the industrial PhD programme as a whole. The only way to mitigate it in practice is through openness and management of this risk throughout the PhD process. Fortunately, it was never necessary to manage the risk of conflicts of interest during this project. It is obvious from the articles produced in this thesis that they favour neither COWI nor the Commission through gratuitous mentioning, branding or positive framing. Interviewees in the Commission and other stakeholders (such as COWI) did not have any interest in answering questions differently because of my affiliation, which was always made clear.

The Commission is naturally keen to appear in a favourable light in the articles, but that is not relevant to my affiliation with COWI. COWI, on the other hand, would have an interest in protecting its brand before the Commission. However, association with a PhD candidate persistently asking questions is not necessarily the best way to do that. Likewise, the findings of this thesis are, on numerous occasions, quite critical towards the EU evaluation system and the Commission's role in this system.

For these reasons, I do not feel that there was a conflict of interest at any time or between me and any of the parties in this project or that it had an effect on the

thesis' findings. I was aware of the potential for conflict from the beginning, but taking measures to mitigate against was fortunately never necessary.

2 THEORETICAL FRAMEWORK

The assumptions and hypotheses in this thesis are drawn from organisational theory. Organisational theory is, broadly speaking, the logical first choice of theory for this thesis, because both accountability and evaluation are linked to public administration and organisation. Practices of accountability, evaluation practices and evaluation use take place almost exclusively in organisational contexts. Typically, evaluations are procured by organisations, carried out by organisations (enterprises or teams of individuals) and read and acted upon by organisations.

According to Dahler-Larsen (2012: 34), ‘evaluation is strongly dependent on its social and organizational context’. In parallel with Dahler-Larsen, Boswell (2008: 473) argues that ‘Any account of how organizations use knowledge will inevitably be premised on a theory of organizations: a set of claims about the sources and nature of organizational interests, and how these translate into organizational action.’ Hence, to answer the research question (*What effect do evaluation systems have on evaluation use?*), we should rely on a theory of organisation with assumptions about accountability and organisational behaviour to explain evaluation use in evaluation systems.

This section explains the theoretical framework of this thesis leading up to the formulation of the research question and the sub-questions. In the first subsection, I give an overview of the theoretical background and assumptions in the evaluation literature. Second, I introduce the theoretical framework of this thesis that is a precondition for formulating the underlying assumptions of the thesis described in subsection three. Subsection four describes the research question and three sub-questions that follow logically from the theoretical framework and the gaps in the literature.

2.1 COMMON ASSUMPTIONS IN THE EVALUATION LITERATURE

Evaluation has two main functions in democratic governance; namely, enhancing accountability and facilitating learning (van der Meer and Edelenbos, 2006; Scriven, 1991). The former type is commonly referred to in the evaluation literature as ‘summative evaluation’, while the latter is called ‘formative

evaluation', a dichotomy invented by Michael Scriven (Scriven, 1967) that is often the starting point for discussion about evaluation use and learning.

In the literature, it is commonly suggested that evaluations do not always generate learning, even though they are implemented to support the learning from past experience (Scriven, 1991; Albæk, 1995). Also, it is suggested that control systems such as evaluation and audit are implemented at a cost to general performance (Boven, 2005). For that reason, the dichotomous relationship is often seen as a contradictory relationship between learning and accountability understood as outcomes. This is also the case in several other areas of research, such as audit studies (Power, 2005; Boven, 2005; Dubnick, 2005) and studies on knowledge use (Boswell, 2008; Böhling, 2013; Mendez and Bachtler, 2011).

Evaluation itself was born at a time when policy-makers and administrators believed in constructing a better society through rational interventions and social engineering based on the lessons of past successes and failures (Vedung, 2010). Therefore, from the outset the underlying assumptions of evaluation's role in public administration were rationalist, causal and evolutionary in nature (Sanderson, 2000). In relation to public policy, evaluation was one of the feedback mechanisms described in the system theory framework (Easton, 1965) that informs policy-makers and civil servants of what works and what does not. Thus, most evaluation models and assumptions are derived from this primary rationale and function according to the logic of cause and effect (see, for example, Mark and Henry (2004: 38); Cousins and Leithwood (1986); and Pawson and Tilley (1997). When it was introduced in North America in the 1950s, evaluation was associated primarily with policy learning.

In the evaluation literature, a number of general assumptions underlie evaluation and its use: first, public organisations are assumed to rationally pursue improvements to public policy; second, public organisations are assumed to be able to objectively interpret feedback such as evaluations; and finally, it is assumed that evaluations are, in fact, used to improve public policies. Ontologically and epistemologically, these assumptions are linked to realist and positivist understandings commonly found in classical economic thinking and rational choice, where policy-makers are perceived to be rational and able to

objectively acknowledge and assess the social world around them in a way comparable to other people (Sanderson, 2000). In evaluation, this positive assumption also dominates through the general perception of ‘social betterment’ as the ultimate objective of evaluation (Henry, 2004). Thus, if one can talk about a ‘general logic’ of evaluation, this logic is inherently realist and rational (Albæk, 1995; Van der Knaap, 1995; Schwandt, 1997; Sanderson, 2000).

Despite the fact that the initial purpose of evaluation was formative, evaluation was also institutionalised to support organisations’ accountability in relation to public spending programmes. This was particularly so as evaluation was introduced in many countries as part of New Public Management (Pollitt and Bouckaert, 2004). The institutionalisation of evaluation in public administrations trained a lens on evaluation systems, when criticism arose from practitioners in fields where evaluation had been introduced as an accountability tool rather than as a learning tool. Nevertheless, evaluation became an institutionalised and ritualistic practice that was closely identified with good governance and accountable government (Dahler-Larsen, 2012). The introduction of evaluation into public administration thus meant a shift from a formative to a summative purpose of evaluation that translated into a dichotomous relationship between the two in the evaluation literature. This also gave rise to the paradox that evaluation inherently was a learning tool, but that it in practice was rarely used as such.

This non-use is exceptionally difficult for the existing evaluation theory to explain, with its underlying assumptions of rationality and belief in feedback to the political system. Instead, this thesis proposes to look to organisational theory and borrow another set of assumptions from organisational institutionalism to improve this explanation. Organisational institutionalism is introduced below, together with the assumptions that will lead the research question.

2.2 ORGANISATIONAL INSTITUTIONALISM

Fundamental to the theoretical contribution of organisational institutionalism is the distinction between organisational efficiency and organisational legitimacy. One famous contribution that spelled out this dichotomy was Meyer and Rowan’s (1977) distinction between output legitimacy and normative legitimacy. They wrote: ‘Some organizations use routine, clearly defined technologies to produce

outputs. When output can be easily evaluated... . . . efficiency often determines success.’ (Meyer and Rowan, 1977: 353-54). They continued: ‘Increasingly, such organizations as schools, R&D units, and governmental bureaucracies use variable, ambiguous technologies to produce outputs that are difficult to appraise’ (ibid). What Meyer and Rowan allude to in their ground-breaking article is that some organisations (such as public organisations) seek legitimacy rather than efficiency. Thus Meyer and Rowan laid the foundation for organisational institutionalism by adopting the view that allows us to understand organisations as legitimacy-maximising rather than efficiency- or output-maximising.

While organisational theory for several decades has been dominated by organisational institutionalism (especially in Scandinavia), the field of evaluation has not taken serious note of these theoretical developments (Højlund, 2014b). A few evaluation studies use institutional theory and consider adoption of evaluation practices in organisations and countries (Hansen and Borum, 1999; Borum and Hansen, 2000; Radaelli, 2009; Sager and Rissi, 2011; Varone et al., 2005; Benjamin, 2012). Linked to the use of knowledge, Boswell (2008: 485-6) uncovers evidence that an organisation uses knowledge to legitimise itself in an unstable and uncertain environment, as predicted by institutional theory. Dahler-Larsen argues that ‘institutional theory can contribute to describe the relatively mechanical diffusion of a phenomenon like evaluation, where the adopted standard in the respective organisation, however, remains a symbolic facade, which does not affect the organisation’s core activities significantly.’ (Dahler-Larsen and Krogstrup, 2006: 297; Meyer and Rowan, 1977). However, the theory has not been linked explicitly with evaluation systems and their use of evaluation.

By developing a broader theory of organisational behaviour, organisational institutionalism is probably the organisational theory that focuses most directly on accountability. Early on, Selznik (1949) discovered that organisations act contrary to the rational utility-maximising behaviour that was expected from organisations in classic organisational theory. He discovered that organisational behaviour often worked contrary to the objectives organisations set for themselves. Instead he found that organisations mostly act according to norms and values in their environment and with the aim of legitimising themselves (Zald and Denton, 1963; Scott and Meyer, 1994). An interesting conclusion from these early studies was

that organisations adapt to the practices that are legitimised by the organisational environment (Meyer and Rowan, 1977). This is particularly the case if the organisation perceives itself to be operating in an unstable environment, where there is uncertainty about the distribution of resources between itself and rival organisations (DiMaggio and Powell, 1991: 30-31). Meyer and Rowan (1991: 190) argue that organisations are concerned with reducing uncertainty and stabilizing social relations (exemplified in Selznik (1984) and thus not primarily with maximising either power or outputs.

At the heart of institutional theory is the definition of institution. First, institutions should not be confused with organisations. Rather, they are the frames for action of organisations, or, simply put, the taken-for-granted formal and informal rules of the game. Scott (1995: 49) writes: ‘...institutions are multifaceted, durable, social structures, made up of symbolic elements, social activities, and material resources’. From this definition, it is important to stress that institutions are resilient to change and that they consist of cultured-cognitive, normative, and regulative elements that provide meaning to social life (Scott (1995: 48). Thus, formal and informal rules shape social action, constraining actors and limiting their choice of action, and might be the subject of conflict between competing interests (DiMaggio and Powell, 1991). An institution requires a clear typification (e.g., ‘evaluation’), social knowledge or sedimentation.

The institution of evaluation is taken for granted in many organisations. A number of relatively stable interpretations of concepts, practices, principles, norms, ethics, values and artefacts are associated with evaluation. In other words, evaluation fits well with Scott’s previously mentioned definition of an institution. Dahler-Larsen and Krogstrup (2006) argue that evaluation has become a protected discourse and that ‘Evaluation has stepped into character as a mandatory political and administrative ritual, which priorities and decisions need to adhere to in order to be socially valid’ [my translation]. Schwandt (2009) refers to the ‘evaluation imaginary’. In other words, evaluation has become a de facto legitimising institution – in many cases a practice we accept without question.

The notion of ‘organisational field’ is also important in institutional theory. Scott (1995: 207-208) defines it in the following way: ‘... the notion of field connotes

the existence of a community of organizations that partakes of a common meaning system and whose participants interact more frequently and fatefully with one another than with actors outside of the field.’ The most important actors in the field of evaluation are arguably the World Bank, OECD, UN, the American Evaluation Association and the European Evaluation Society, as well as regional supranational political organisations such as the EU and some national donor agencies. (Furubo et al., 2002; Varone et al., 2005; Stern, 2006). They mutually reinforce and constantly negotiate the evaluation institution through their production of guidelines and practices, as well as through exchanges of opinions in public debates and at evaluation seminars, courses and conferences. The field is an important concept, because this is what an organisation adapts to.

In this thesis, I argue that the notion of field is too broad to enrich the debate in the evaluation literature on evaluation use. However, it contains assumptions that are nonetheless useful. These assumptions will be developed further in the next section. Also, while organisational institutionalism is concerned with *legitimacy* in a large organisational field, evaluation systems are smaller clusters of interdependent public organisations where legitimacy is gained through accountability. Thus, the evaluation system is a community of organisations that reinforce shared evaluation practices put in place to secure accountability. The evaluation system is what the organisation adapts to, in order to seek legitimacy and accountability and survive in the long run.

Thereby, the concept of evaluation system is not as extended as ‘organisational field’ and therefore more operational and with a more easily defined scope in empirical terms. The translation from ‘field’ to ‘evaluation system’ is also important because of the existing debate on evaluation systems in the evaluation literature. By proposing to apply the assumptions of organisational institutionalism to evaluation systems, this thesis adds more theoretical depth to the concept of evaluation system. This follows from the key assumptions that are made in organisational institutionalism below.

2.3 KEY THEORETICAL ASSUMPTIONS

The theoretical vantage point of organisational institutionalism is organisational legitimacy. The concept of accountability is strongly related to the legitimacy and

theoretical assumptions made by organisational institutionalism in relation to legitimacy.

Accountability is the hallmark of democratic governance and linked almost exclusively with public governance, administration and management.

Accountability is broadly defined in this thesis 'as a social relationship in which an actor feels an obligation to explain and to justify his or her conduct to some significant other' (Boven, 2005). This definition implies that accountability in a setting of public governance ultimately relies on the norms, values and cognitive frameworks of actors. Thereby accountability becomes almost synonymous with legitimacy, as the definition of accountability also stresses the link between the organisation and its environment and the assumption of justificatory objectives associated with the organisational behaviour in organisational institutionalism.

Following from the above presentation of organisational institutionalism, the assumptions of this thesis are the following:

- 1. Organisations survive by legitimising themselves in the organisational field.**
- 2. Therefore, organisations aim to legitimise themselves.**
- 3. One way organisations legitimise themselves is by pursuing practices of accountability.**

According to organisational institutionalism, organisations survive in the organisational field by legitimising themselves in that field. As we saw earlier, the organisational field is a community of organisations sharing norms such as evaluation. In this thesis, field is replaced by evaluation system (see section 3.2).

According to organisational institutionalism, the main aim of organisations is to seek legitimacy. One way of doing so is to establish practices of accountability (in the evaluation system). The fourth assumption of this article is that evaluation is a practice that establishes accountability:

- 4. Evaluation is used in public administration to establish accountability.**

This assumption finds support in the evaluation literature. Evaluation is commonly understood to have two main functions in democratic governance; namely, enhancing accountability and securing policy learning (van der Meer and Edelenbos, 2006; Scriven, 1991). Whereas the former emphasises retrospective assessment in view of attaining political legitimacy relative to the evaluated object (summative evaluation), the latter emphasises improved policy-making (formative evaluation). In the evaluation literature, overwhelming empirical evidence supports the claim that evaluation is indeed an administrative tool used to secure organisational legitimacy through accountability.

Seen through the lens of organisational institutionalism, organisations adopt evaluation practices in order to legitimise themselves in the organisational field. Legitimacy is preferred over the efficiency associated with learning from experience through evaluation. This is even more so when the organisation is public and operating under great uncertainty related to governmental budget allocations and public trust.

Since organisational institutionalism posits legitimacy as the most important survival strategy for public organisations, we can assume that the use of evaluations to improve policies is a secondary objective for them. This assumption is largely shared by the evaluation literature, as explained above. Hence, the following assumption relevant to this thesis follows logically from organisational institutionalism:

5. Evaluation is implemented and used mainly to establish organisational accountability within the evaluation system.

Now we have associated a theory of organisation with evaluation use and the evaluation system. We have used organisational institutionalism to produce five assumptions of organisational behaviour and organisational use of evaluation in the evaluation system.

2.4 RESEARCH QUESTION

From the above, we assume that especially public organisations are likely to maximise their legitimacy by appearing accountable in the evaluation system. One

way public organisations can increase their accountability is by evaluating so that other organisations and citizens can secure oversight of the organisation.

However, if organisations are determined to pursue accountability through evaluation, where does that leave learning and evaluation use that results in improvements to public policies? To increase our understanding of both non-use as well as evaluation use in general under these assumptions, we need to ask and answer the following research question:

What effect do evaluation systems have on evaluation use?

The research question is directly linked to the original puzzle that my colleagues and I grappled with one day around the table in the Danish consultancy COWI. However, the question also links with some of the latest developments in the evaluation literature; namely, that of evaluation systems.

As we see more and more evaluation systems come into being, it is relevant to ask what the effects of these systems are on evaluation use. Looking to the evaluation literature, evaluation systems are generally assumed to have a negative effect on the use of information and knowledge in policy-making (Power, 1997; Leeuw and Furubo, 2008; Dahler-Larsen, 2012; Pollitt et al., 1999; Furubo, 2006).

To answer the research question, three sub-questions have directed the research in the four articles that constitute the main empirical work of this thesis. The sub-questions link up with the gaps identified in the literature.

- 1) *How can non-use and justificatory uses of evaluation be explained?*
(addressed in articles 1, 3 and 4)
- 2) *How can evaluation use be explained in its systemic organisational context?*
(addressed in articles 1, 2, 3 and 4)
- 3) *How are evaluations used in evaluation systems?* (addressed in articles 2, 3 and 4)

The research question and the three sub-questions set the scene for an interesting investigation of the inner dynamics of inter-organisational dependencies and of the implementation of evaluation practices and the institutionalisation of evaluative norms and values. It is here that I think we as researchers of evaluation can make the greatest contribution today.

3 CONCEPTUAL FRAMEWORK

The overall theoretical inspiration is found in organisational institutionalism as described in the previous chapter. But the thesis also draws on conceptual frameworks from evaluation use theory. In the following subsections, I will elaborate on the conceptual background of evaluation and evaluation use theory.

3.1 EVALUATION

In this section, I cover the definition of evaluation, the purpose of evaluation and the evaluation system.

3.1.1 DEFINITION OF EVALUATION

The usage of the term ‘evaluation’ in relation to assessments of government interventions started in the 1950s and gradually gained momentum in different sectors at different times. Hence, government interventions within development aid, health, education and research were scrutinised relatively early by practices and termed ‘evaluation’. Only in the 1990s did evaluation have its ‘linguistic breakthrough’, to the point where Evert Vedung called it ‘a linguistic magnet’ (Vedung, 1997). This was after evaluation from the 1980s and onwards had become part of the NPM-wave that permeated North American and European public administrations (Hood, 1991).

There are numerous definitions of evaluation. As an example of a common definition of evaluation, Vedung (1997) offers the following:

‘Evaluation is 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.’

In this general definition it is worth noting that evaluation is ‘retrospective’. Evaluation therefore does not include prospective planning tools, such as impact assessment or ex ante appraisals (also called ex ante evaluations). Further, evaluation is related to ‘government interventions’, which links the term exclusively with public governance but excludes its use in relation to for-profit and NGOs. Finally, stating that evaluation is ‘intended to play a role in future practical action situations’ implies some use of evaluation findings, learning or

some other kind of information processing with a chronology or causality and with the objective of impacting decision-making. It is important here to note that evaluation according to Vedung is *intended* to play this role in decision-making. Hereby, Vedung implies that evaluations do not always do so, which is a relevant distinction for studies on evaluation use and learning from evaluation.

The role of evaluation in relation to decision-making is particularly relevant to this thesis. To illustrate this importance, other scholarly definitions also refer to the subsequent use of evaluation. For example, Cousins et al. (2004) also define evaluation as a tool to improve future decision-making. They define evaluation as a

‘systematic inquiry leading to judgements about program (or organisation) merit, worth, and significance, and support for program (or organisational) decision making’.

Cousins et al.’s definition takes for granted that evaluations support decision-making, which we will see later is often not the case. Both definitions have much in common with Scriven’s early, somewhat minimalist, definition: ‘Evaluation is the process of determining the merit, worth, and value of things’ (Scriven, 1991: 1). However, Scriven does not define the purpose of evaluation (intended support for decision-making) and his definition is indeed very broad.

In this thesis, Vedung’s definition is used due to its reservations regarding the usefulness of evaluation to decision-making. Another important element of this definition is its uncompromising stance on the retrospective nature of evaluation. Though it seems counter-intuitive that evaluation can be anything but retrospective, common planning tools such as impact assessment and ex ante appraisals are often labelled ‘evaluation’. The distinction was useful for this thesis in order to frame the cases to be mainly retrospective programme evaluation in the European Commission and not ex ante evaluation nor impact assessment, which are also important but very different tools in the Commission.¹ Therefore, the

¹ The Commission defines evaluation as “a judgment of interventions according to their results, impacts and needs they aim to satisfy.” http://ec.europa.eu/dgs/internal_market/evaluation/evaluation/index_en.htm (1st November, 2014).

definition allows for more generalisability of the findings than would be the case if completely different tools such as impact assessment and ex ante appraisals were also contained in the definition.

3.1.2 PURPOSE OF EVALUATION

Evaluation has two main functions in democratic governance; namely, enhancing accountability and facilitating learning (van der Meer and Edelenbos, 2006; Scriven, 1991). Whereas the former emphasises retrospective assessment in order to secure accountability and oversight of the programme in question, the latter emphasises improved policy-making. The former type is commonly referred to in the evaluation literature as ‘summative evaluation’, while the latter is called ‘formative evaluation’. This thesis starts from this distinction and dichotomy, as it focuses on the implications of evaluation systems for the formative role of evaluations.

Scriven’s dichotomy is often taken as a starting point for discussion about evaluation use and learning. However, in this case it should be noted that the literature on use is far more developed and extensive than the evaluation literature on learning. For the same reason, this thesis uses the term ‘evaluation use’ to cover both evaluation use and learning.

Scriven introduced the distinction between summative and formative evaluation (Scriven, 1967), but most evaluation literature takes it as a base for the elaboration of standpoints. Alkin and Christie (2004) have made several attempts to map the evaluation literature. The result is that one important part of the literature investigates the formative aspects of evaluation under the heading ‘evaluation use’ (Shulha and Cousins, 1997: 199) and essentially focuses on participatory evaluation and the delicate relationship and participation of stakeholder – and particularly intended users – in the evaluation process. This type of evaluation can also focus on personal and organisational learning, as well as ownership, engagement, self-determination and emancipation. Hence, organisation theory on learning and processes was also integrated in the theory building (Torres and Preskill, 2001; Preskill and Torres, 1999) as well as in cognitive theories on learning and knowledge sharing (Dolowitz and Marsh, 1996; Owen and Lambert, 1995). In collaborative modes of evaluation, the evaluands or stakeholders are

involved in the evaluation practices, and the evaluator assumes the role of teacher while the idea of the objective and independent evaluator is down-played. Collaboration has been promoted as a vehicle for capacity building and organisational learning (Cousins and Earl, 1995b; Cousins and Earl, 1995a) as well as stakeholder empowerment (Fetterman, 1994).

The proponents of 'evaluation use' include Daniel Stufflebeam, Marvin Alkin, Michael Q. Patton, Hallie Preskill, Jean King, David Fetterman and Brad Cousins. Michael Q Patton is probably the most notorious and the reason 'evaluation use' or 'utilization' plays such an important role in evaluation literature. His key contribution, 'Utilization-focused evaluation' (Patton, 1997), is a well-written textbook that is mainstream in evaluation curricula, despite Patton's extreme views that make him the most prominent representative for the formative branch of the evaluation literature. According to Patton, the evaluator should cast away any methodological considerations of reliability and validity if they stand in the way of evaluation use. In other words, the use by intended users of the evaluation takes priority over any other consideration. This view is also sometimes referred to as 'pragmatic' and is criticised for being too close to management consulting and too far from objective independent evaluation.

In contrast to the views of Michael Q Patton, a number of authors emphasise the summative role of evaluation by focusing on methodology, objectivity and reliability. This branch in the evaluation literature is represented by Thomas Cook, Gary Henry and Melvin Mark, Huey Chen, Carol Weiss, Ray Pawson and Rick Tilley, who are sometimes referred to as the 'positivists'. Here we also find the adherents to the so-called 'realist' type of evaluation. According to Alkin and Christie (2004), proponents of the positivist evaluation are also supporters of a value-free type of evaluation. They focus particularly on high reliability, and emphasise that evaluation should be as close as possible to objectivist social science and that truth about a social phenomenon can be value-free and studied out of its political context.

Very dear to the 'positivist' views is the discussion of evaluation's place in relation to science and research. Research and evaluation are both related to the systematic collection of data and production of knowledge. According to Owen

(2002), traditional academic social research, as well as policy analysis and program evaluation, is based on systematic social inquiry. They all use the same scientific methods and can analyse the same study objects. However, they have different objectives. Whereas science aims for universality and generalisability, evaluation aims to study a particular phenomenon and does not apply methods and research designs in order to generalise the findings to the population or to seek a universal truth. In other words, the purposes and uses of the knowledge produced are different (Alkin and Taut, 2003). In the case of research, the goal is generalisable knowledge that holds true across all settings, times, and for all individuals represented by the produced knowledge. Science aims to produce general theories through rigorous testing of hypotheses. The collection of systematic knowledge is based on information gathered by systematic social scientific procedures. The systematic creation of knowledge is based on linking empirical observations and conclusions in a coherent, shareable, and persuasive way with the aim of ensuring objectivity, reliability and validity of findings.

It is important to focus on the distinctions between knowledge produced through research and knowledge produced through evaluation. In most cases, the focus of an evaluation is more specific. Evaluation can thus be compared to a case study: its purpose is context-specific and applicable only within a particular setting at a particular point in time, and it is intended for use by a particular group of people (Cronbach and Suppes, 1969). According to Scriven (1967) and the definition used earlier by Vedung, evaluation is different from research in terms of its unique purpose of judging the worth, merit, or quality of the evaluated object, as well as in terms of the context-specific uses of the knowledge produced to make these judgments.

3.2 EVALUATION SYSTEMS

In the evaluation literature, the attention on evaluation systems increased on publication of *From Studies to Streams*, edited by Ray C Rist and Nicoletta Stame (Rist and Stame, 2006) and *Systems concepts in evaluation: an expert anthology*, edited by Bob Williams and Iraj Imam (2007). In the evaluation literature, the concept of 'evaluation system' is used in two different ways. First, it is used mainly as a reference to a 'system' of organisations or organisational entities in which evaluation practice is institutionalised. This view on evaluation systems

developed to describe the rapid growth of systematic and institutionalised evaluation practice observed in public agencies in the 90's and 00's. Second, the concept is used to describe the 'evaluation object' (Imam et al., 2007). In this view, the emphasis is on a more contextual approach to evaluation studies and practice. The argument is that traditional evaluations focus on organisations, projects or programmes as the object of evaluation but that this is not adequate in order to understand (and evaluate) the complexity of the organisation, project or programme. Instead, the focus should be on the system-level to allow for more complex analyses of cause and effect.

In this thesis, 'evaluation system' refers to the organisational entities forming a collective (the system) that shares a common understanding of evaluation practices in that collective, including the purpose of evaluation. Based on Leeuw and Furubo (2008), an evaluation system can be understood as 'permanent and systematic evaluation practices taking place and institutionalised in several interdependent organisational entities with the purpose of informing decision-making and securing oversight' (Højlund, 2014a).

This definition draws on Leeuw and Furubo (2008) as well as (Dahler-Larsen, 2012; Dahler-Larsen, 2006). Following Leeuw and Furubo (2008), the four constitutive elements of an evaluation system are: 1) common practice; 2) organisational entity; 3) permanence; and 4) purpose of use.

First, evaluation needs to be institutionalised to the extent that the actors involved in the evaluation activity share a common understanding about evaluation, including the way evaluation tasks are performed and the potential outcomes. Thus, for the participants in an evaluation within an evaluation system, there are common expectations, and conscious and unconscious normative values about evaluation and the evaluative practices associated with it. These might not be identical, but evaluation practices should be generally accepted throughout the system for the system to be institutionalised.

Further, Leeuw and Furubo (2008: 159) argue that evaluation systems need to be constituted by several entities. There has to be not only an evaluator but also a commissioner of the evaluation. The evaluation system is thus institutionalised formally in at least one organisational structure which separates it from the

operational structure of this organisation. Hence, the system has at least one formal institutionalised organisational element (for example, ‘an evaluation unit’) that, typically, is in charge of planning, tendering, implementing, quality-checking and follow-up of evaluations. The term ‘evaluation machine’ refers to this particular entity, which is a constitutive element of the evaluation system but should not be confused with the entire evaluation system (Dahler-Larsen, 2006). An evaluation system can be constituted by, for example, an evaluation unit, policy units, external evaluators, politicians and maybe some beneficiaries, NGOs and industry.

As a third characteristic, Leeuw and Furubo (2008) emphasise that an evaluation system is permanent and not ad hoc. Evaluation activities are thus constantly ongoing and reproductive and there is a history of systematic evaluation in the organisation. It is important to stress that permanence does not necessarily imply stasis. The system is likely to change but usually according to dynamics defined in the system or changes in the environment of the system.

Finally, Leeuw and Furubo (2008) claim that evaluation systems need to include some sort of practice with which evaluations are used. Thus, evaluation practices need to be related to decision-making practices as well. Evaluation cannot stand alone as practice, and needs to be related to the practice of decision-making or oversight as these are practices of public organisations. Thus in the evaluation system, evaluations are organised and planned so that they relate to the activity cycle of the organisation or the evaluand (e.g., budget or policy cycle).

To summarise, an evaluation system has a purpose (learning and/or accountability) and a boundary that is defined by the institutionalisation of evaluation practice. This is very much related to the purpose of the system. So, the evaluation system extends to actors who take evaluation practices for granted, either as consumers or producers of evaluations. Consumers are typically political masters such as the European Parliament or European Council, who expect evaluations to be produced by the European Commission with the aim of learning and oversight (accountability). Outside the EU evaluation system (described in more detail in section 4.2), we find NGOs, industry, the press and citizens, who might consult single evaluation reports for information although they do not internalise

evaluation practices in a systematic way either as consumers or as producers. The difference is that the European Parliament expects to consume evaluations from the European Commission systematically and over time, while citizens would not do so.

One of the main arguments of this thesis is that the evaluation system is a good framework for understanding evaluation use. There are two reasons for this:

1) The evaluation system provides an organisational framework in which evaluation and evaluation use can be understood. An evaluation system is constituted by several organisations and therefore the concept also captures the relationship between these organisations instead of focusing solely on the evaluating organisation. This is important because evaluation use, and in particular non-use, cannot be understood if we do not understand the interdependent relationships between the evaluating organisation and those organisations that oversee and check the work of that organisation.

2) This is because in democratic systems of governance public organisations control each other through systems of checks and balances. According to organisational institutionalism, this control determines the legitimacy of the organisation and ultimately its survival. Evaluation is one tool among several that is used to establish control and oversight in determining the legitimacy of an organisation. However, the concept of evaluation system captures the organisational interdependence very well among those organisations that are part of the system; that is, those organisations that share a common understanding of the evaluation practices put in place to control one or more organisations.

Following from these two points, it is important to stress that the notion of ‘field’ from organisational institutionalism and ‘evaluation system’ are related but should not be confused. The organisational field as described by organisational institutionalism is a very broad community of organisations that uphold and update the institution of evaluation (e.g., professional associations, international organisations and states). An evaluation system is much more specific and related to a defined set of organisations defined by a common understanding of evaluation practices within the system, and thereby also by a common understanding of the

control and oversight associated with evaluation as well as the potential for policy learning and improved decision-making.

For this reason, it makes more sense to talk about ‘accountability’ rather than ‘legitimacy’ in the context of the evaluation system, because accountability relates more specifically to the relation of power and control that is the key premise of the evaluation system. Evaluation is institutionalised in an evaluation system in order to secure this accountability. That is not the case in the organisational field that organisational institutionalism refers to. In the organisational field, organisations that share the institution (for example, evaluation) might be totally independent of each other in terms of power, control and accountability.

Therefore, I argue in this thesis that the notion of ‘field’ is too broad to enrich the debate in the evaluation literature on evaluation use. However, it contains some important and useful attributes. While organisational institutionalism is concerned with *legitimacy* in large organisational fields consisting of many organisations, evaluations systems are smaller clusters of interdependent public organisations, where control and power is distributed and accountability is important for the organisation’s survival. Thus, the evaluation system is a community of organisations that reinforce shared evaluation practices that have been put into place in order to secure accountability. Borrowing the logic from institutional theory that organisations adapt to the external environment (the field), this is also the case for the organisations in the evaluation system. In the evaluation system, the evaluated organisation also adapts in order to seek accountability and survival.

Table 3-1 Difference between evaluation system and organisational field

	Evaluation System	Organisational field
Scope	Defined cluster of public organisations sharing norms of evaluation practice in relation to distributed power and control	Shared norms and values among a large and often undefined number of organisations
Survival factor	Accountability through formal evaluation practices and evaluation output	Legitimacy through adopting institutions (norms, values, practices of evaluation)
Example	The EU evaluation system comprises the Commission, the EP, the Council, MSs and consultants who share common evaluation practices and norms of control of the Commission's expenditures.	In the evaluation 'field' the institution of evaluation is shared among hundreds of organisations around the world, with the most important being states and international organisations.

Thus, the concept of evaluation system is not as extended as 'organisational field' and its scope is therefore more operational and easier to define in empirical terms. In the literature on evaluation systems (Olejniczak, 2013), other concepts are applied to describe the 'flows' of things, resources and information ('stacks') and feedback loops within multi-layered systems. This level of ontological complexity was not sought in this thesis as the objective of study primarily was evaluation use within the organisational boundries of the EU evaluation system.

The translation from field to evaluation system is also important because of the existing debate on evaluation systems in the evaluation literature. However, this

debate is not founded in any particular theory that allows for hypotheses to be formulated in relation to evaluation use. By proposing to apply the assumptions of organisational institutionalism to evaluation systems, this thesis adds theoretical depth to the concept of evaluation system.

In relation to evaluation use, evaluation systems are generally assumed to have a negative effect on information and knowledge use in policy-making (Power, 1997; Leeuw and Furubo, 2008; Dahler-Larsen, 2012; Pollitt et al., 1999; Furubo, 2006). Previous studies suggest that evaluative knowledge tends to be made relevant primarily for administrators and not for policy-makers, and that use in administrations will be linked to procedural assurance and legitimisation of the organisation rather than policy-making (see also Furubo, 2006; Langley, 1998). The purpose of this thesis is to continue the research on the effect of evaluation systems on evaluation use, and to provide empirical evidence where presently there is little. The next section elaborates on the phenomenon of evaluation use.

3.3 EVALUATION USE

The 1950s and 1960s saw a massive expansion of public interventions in the construction of welfare states in Western Europe and North America. A positivist approach to social engineering was prevalent and policy-makers believed in social betterment through targeted interventions that changed people's behaviours (Porter, 1995). Part of this paradigm was a belief in the ability of policy-makers and administrators to learn lessons from previous interventions and thus to constantly improve the quality, efficiency and effectiveness of public spending and interventions. In Easton's terminology (Easton, 1965), evaluation relates to the feedback that policy-makers receive as input for improved policy making.

However, this positive view on policy-making and public interventions was soon challenged in both organisational studies on decision-making and implementation studies (Cohen et al., 1972). Scholars and practitioners of evaluation and science started to question whether evaluative and scientific knowledge was, in fact, used to improve policies (Lazarsfeld et al., 1967). Though the paradigm to which Easton belonged assumed feedback happens and improvements are made, very little evidence existed to support this theory. In fact, subsequent research into the use of scientific knowledge and evaluation use often illustrated how existing

knowledge was not used to improve policies. This phenomenon was referred to by some scholars as a 'utilization crisis' (Patton, 1997).

Because of concern over evaluation findings not being used, a large body of literature related to evaluation use developed from literatures on the use of scientific results in policy making (Lazarsfeld et al., 1967; Porter, 1995; Vedung, 2010; Weiss, 1998; Weiss and Bucuvalas, 1980). At present, the substantial literature on 'evaluation use' exists independently of the literature on the use of scientific results and knowledge, mainly because the field of evaluation is a relatively specific practitioner field. Science and knowledge use are much broader categorisations and investigated mainly by organisational sociologists.

Nevertheless, 'evaluation use' can be viewed as a sub-field in the literature on knowledge use. The knowledge-use literature and the evaluation-use literature are already very closely related, though cross-referencing does not always take place. Knowledge use is obviously broader than evaluation use; however, evaluation use does have particular attributes due to the specificity of evaluation itself. First of all, evaluation is an administrative tool invented in order to improve feedback to policy-makers and to secure accountability. One purpose of evaluation is thus to be used to improve the knowledge of policy-makers so that policies will also improve. This is different from, for example, scientific knowledge production, which does not aim to guide policy-makers but instead to increase our understanding and knowledge about our lives and the world we inhabit. A second objective of evaluation is related to accountability, which is again related to democratic control and oversight of public administration. This also is not shared by scientific knowledge production.

In the wake of disenchantment over the scarce evidence of use of evaluations, the literature turned to the question of why evaluations were, or were not, used (Leviton and Hughes, 1981; Cousins and Leithwood, 1986). After the evaluation literature more or less had converged on a typology of four categories of use (see below), research was then dedicated mostly to explanatory variables and to answering the question of interest to all evaluators: what makes my evaluation useful? This resulted in numerous case studies that inferred relationships between contextual variables or conditions and the four evaluation use categories or

variations based on them. The literature focused on factors related to the attributes of the evaluation (e.g., methodology, quality, and relevance of findings) or the immediate contextual factors pertaining to the organisation, in which the evaluation is implemented (e.g., political climate, timing of the evaluation relative to decision-making and so on). These categories were empirically informed from the late 1970s and onwards (see for example Leviton and Hughes, 1981). Other streams of literature look into the building of evaluation capacity in organisations or formative, process, empowerment and developmental evaluations, which are attempts to enhance the use of evaluation through a change in evaluation practice (Preskill and Torres, 1999; Patton, 1997). With Michael Q. Patton's 'discovery' of process use (Patton, 1997), questions related to the users of evaluation also became more central.

3.3.1 DEFINITION OF USE

Leviton and Hughes (1981) argue that two 'bottom-line' criteria must be satisfied in order for something to be called use:

- 1) There must be a conscious attempt to relate the evaluation findings to the policy or program issue under study in order for evaluations to be considered used (Cook and Pollard, 1977: 161). Merely reading the evaluation report is not considered use. Weiss (1978) calls this 'information processing'.
- 2) In order for evaluation use to be observable, there must be evidence that in the absence of the evaluation, those actors and stakeholders involved with the policy or programme activities would have thought or acted differently. That is to say, there must be an impact on the programme or policy activity of the stakeholders involved. This evaluation-use criterion is obviously more difficult to satisfy.

A practical definition of evaluation use is provided by Kelli Johnson: Evaluation use is '*any application of evaluation processes, products, or findings to produce an effect*' (Johnson et al., 2009a: 378). In relation to Johnson's definition, a number of critical points are raised. One could argue that use of, for example, findings, does not – strictly speaking – need to produce an effect to be classified as use, as is the case with Johnson's definition. Therefore, one could argue that the definition of evaluation use could be limited to '*any application of evaluation*

processes, products or findings'. Such a definition would expand to incorporate more instances of application.

For example, this implies that simply reading the evaluation would also classify as evaluation use. This thesis leans towards this broader definition that does not entail an effect or policy impact of the evaluation. It does so, because effects and impacts of evaluations are very hard to research empirically. To attribute a change in policy to one document is often hard, if not impossible. Second, it is arguably a type of use, if policy-makers read and acquire knowledge from evaluations. Consciously deciding to read or otherwise engage with the evaluation or the implementation of the evaluation is also using it, just as reading a book is 'using' that book. Broadening the definition of evaluation use in this way allows for a much more explorative approach to evaluation use, which was necessary in the context of this thesis due to the complexity of the study object.

Nevertheless, broadening the definition of evaluation use too much is also a problem. Johnson already commits the fallacy of redundancy, when she defines use as application. This merely says that use is use, which is not a strong definition. In this thesis, I keep the concept of use entirely open to interpretation by my interviewees. When conducting the data collection for the thesis, this approach allowed interviewees to speak freely. Opening 'use' to interpretation during the data collection also mitigated the risk of pre-defining the term for the interviewees. Also, the focus of the thesis is to explain the effect of the evaluation system on use. Therefore, the definition of use should be left open to allow for as much variance in the data as possible in order to gauge all types of effects that the evaluation system has on use. This also includes learning, which some interviewees considered a type of use.

While openness to divergent interpretations of use was chosen for the data collection, the analysis and coding of these data was guided by a typology of use. The next section elaborates on these types that have emerged over the decades of research in evaluation use.

3.3.2 TYPES OF EVALUATION USE

When data was collected for this thesis, interview guides were designed to be open towards different interpretations of use. However, during the analysis of the data,

several pre-existing concepts of use were used to categorise and code the interview data. This section explains the origins of these use concepts.

Evaluation use was traditionally divided into three categories: instrumental, conceptual, and symbolic (Leviton and Hughes, 1981). Instrumental use refers to instances where someone has used evaluation knowledge directly. Conceptual use refers to cases when no direct action has been taken, but where people's understanding has been affected. Symbolic use refers to examples where a person uses the mere existence of the evaluation, rather than any aspect of its results, to persuade or to convince. These three categories are still used as the basis for much research, though newer and more elaborate conceptual frameworks have been suggested (Alkin and Taut, 2003; Kirkhart, 2000; Henry and Mark, 2003; Weiss, 1998).

The three categories are summarised here:

1. **Instrumental use** was defined by Rich (1977: 200) as '*cases where respondents cited and could document ... the specific way in which research was being used for decision-making or problem-solving purposes*'. 'Instrumental use' signifies a direct and concrete transfer of evaluation findings and recommendations into the reshaping or reconfiguration at any level of policy or programme. This type of use is expected to be traceable through concrete signs of findings in the subsequent policy or programme, though in practice this is often very difficult.
2. **Conceptual use** of research information was defined by Rich as '*influencing a policy maker's thinking about an issue without putting information to any specific, documentable use*' (Rich, 1977: 200). 'Conceptual use', which is more indirect and relates to 'enlightenment', signifies a more general knowledge use or conceptual development (Stufflebeam, 2001; Owen and Rogers, 1999; Weiss, 1979). For example, programmers could make conceptual use of evaluation when applying theories of change and related concepts to their programme through interaction with the evaluators of the programme.
3. **Persuasive use/symbolic use/legitimising use** involves drawing on evaluation evidence to convince others to support a political position, or to defend such a position. (This category is termed 'symbolic' by Pelz (1978); and Young and Comptois, 1979). Owen and Rogers (1999) distinguish between legitimising use and symbolic use. The former involves justifying a

prior decision and the latter concerns enhancing the reputation of the organisation or programme. In the latter case, it is the evaluation process – the simple fact that an evaluation is being conducted – that legitimises the organisation. For example, the organisation commissioning the evaluation will use the evaluation as a means to justify its own existence in order to avoid changes such as budget cuts. Legitimising use is when an evaluation is used to justify or legitimise the policy or programme being evaluated (Owen and Rogers, 1999; Patton, 1997; Weiss, 1979). The legitimising use of evaluation results is linked to persuasion on compliance issues and accountability, and therefore it was originally called ‘persuasive’ use.

Another category of evaluation use appeared not much after the other three. It was termed ‘process use’. Process use was defined by Patton (1997: 90) as *‘individual changes in thinking and behaviour and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process’* (Preskill et al., 2003). Process use happens in the evaluation process when stakeholders and the evaluation commissioner are drawn into the process of evaluation with the evaluator. Process use is therefore not dependent on a final evaluation report or other output per se, but happens subsequently to the evaluation process.

These four use categories dominated (and still do dominate) the conceptualisation of evaluation use in the literature. However, several researchers also grew dissatisfied with them. They were displeased with the multitudes of potential evaluation uses and influences that did not fit into these four vaguely defined categories of use (Alkin and Taut, 2003; Henry and Mark, 2003; Kirkhart, 2000; Weiss, 1998). Some criticism was directed at the operationalisation and measurement of the concepts. The appropriate typologies of use are still debated, most probably because assessing the influences and effects of evaluations continues to be difficult (Alkin and Taut, 2003; Hofstetter and Alkin, 2003; Johnson et al., 2009a; Kirkhart, 2000; Mark and Henry, 2004).

There is little doubt that the four categories of evaluation use represent the backbone of the theoretical conceptualisation in the literature on evaluation use. Nevertheless, there are a number of problems with the typology, particularly in relation to the operationalisation of them. It has been argued by several scholars that there is an overlap between the categories due to the very broad definition of

these concepts. Especially conceptual use can be confused with both process use and instrumental use if an evaluation does not result in immediate instrumental use or process use.

Also, it has been argued that the causality of the evaluation use types is not clear. The normal inferred causality and order of events from evaluation to use can be challenged if, for example, instrumental use is determined by external factors such as the policy cycle or other events. So, instead of being put to use immediately by the administration, evaluation use might be determined by the policy cycle or other events much later. If a decision is pending, then instrumental use is more likely. If not, then conceptual use or persuasive use from that evaluation might be more likely. So, it might not be the qualities of the evaluation that determine use, but rather the external context and timing. Also, conceptual use might lead to symbolic use or instrumental use of the evaluation. So, it is not clear if one category of use acts as an intermediate variable to other categories of use and how this contribution problem should be tackled by researchers.

Finally, it has been argued that symbolic use does not satisfy Leviton and Hughes (1981) criteria of use described earlier. Using evaluations to delay action, to allow policy-makers to appear concerned about a problem, or to compete for political positions (Pelz, 1978; Rich, 1977) are not instances of use, if there is no evidence that evaluations were seriously discussed or consciously used this way. However, researchers such as Leviton and Hughes (1981) would argue against this. They would argue that symbolic use should be understood more broadly to cover the activity of evaluation as a legitimising act.

The literature also points to more general problems and difficulties related to research on evaluation use (Kirkhart, 2000):

First, all of the categories are difficult to measure and operationalise in a research project. This is a common problem in studies on learning and knowledge use. The categories are often intangible and hard to measure and understand, both for the researcher as well as the interviewees. This is particularly the case with the symbolic category, because it is perceived as a 'negative' or 'strategic' type of use that data subjects are not easily willing to admit to the researcher. Also, the conceptual category is difficult to measure, as conceptual 'enlightenment' or

learning is notoriously hard to measure. Only rarely is it possible to test policy-makers to see if they understood the concepts and information they have been given.

Second, there is a time aspect to the use of evaluation, which makes research on evaluation use harder. Evaluation use does not necessarily take place immediately after the evaluation, and respondents cannot necessarily remember or link evaluation results to actual decisions or other types of use. Retrospective analysis depends heavily on the respondents' memories. Only cases of explicit use might be recalled, and not less tangible ones. Also, respondents might be biased toward saying that the public organisation they represent actually used the evaluation. Thus, it is harder to attribute evaluation use with policy-making. Also, a political or administrative decision to, for example, change a programme usually evolves over time and depends on many different inputs. The actual effect of the evaluation is therefore almost impossible to gauge.

Third, instrumental use and conceptual use often take place without any documentation (such as oral conclusions in a meeting). This makes it hard for the researcher to identify the decisions made and then to attribute them to the evaluation. Conceptual use is particularly difficult to measure, since it concerns the thoughts and ideas of the administrators or organisation as a whole.

Fourth, most existing studies on evaluation use are case studies. Their strength is to show linkage between evaluation findings and their use in qualitative studies. Their weakness is lack of generalisability and confounding of situational factors. Thus, because case studies are the main approach found in the evaluation literature due to the uniqueness of most evaluation interventions, evaluation use is often also treated as unique cases. This increases variability in the way concepts are applied in each case to match the unique features of the case under study.

3.3.3 ALTERNATIVE FRAMEWORKS OF EVALUATION USE

To remedy the problems with the old framework based on the four categories of evaluation use, several approaches have been proposed. Some key shortcomings were identified by, in particular, Kirkhart (2000), which inspired Henry and Mark (2003) and Alkin and Taut (2003) to make similarly important contributions. In addition to the criticisms summarised in the previous section, Henry and Mark

also argued that the conceptual framework of evaluation use failed to focus on important change processes caused by evaluation. In other words, the literature gives limited guidance about the underlying processes from an evaluation output (a report) to its eventual impact on social betterment through policy improvements. This is a reasonable criticism, as empirical studies have found several categories of use at different points in time. To many researchers it seemed unsatisfactory that the multitudes of evaluation uses should simply be made to fit into four vague categories.

Kirkhart (2000) for her part favoured the concept of ‘influence’ over that of use. Her definition of influence is *‘the capacity or power of persons or things to produce effects on others by intangible or indirect means’* (Kirkhart, 2000: 7). Influence, according to her, includes *‘effects that are multidirectional, incremental, unintentional, and non-instrumental, alongside those that are unidirectional, episodic, intended, and instrumental (which are well represented by the term use)’* (Kirkhart, 2000: 7). She thereby marks the conceptual boundaries of use and points to a much wider concept of evaluation effects that takes into account *‘non-results-based applications, the production of unintended effects, and the gradual emergence of impact over time’* (Kirkhart, 2000: 6).

Kirkhart (2000) theory of influence encompasses three dimensions:

1. Intention (unintended vs. intended): Evaluation influence is usually guided in an intended direction for the ultimate creation of social betterment. However, the effects of an evaluation cannot always be predicted, and thus evaluation influence can be unintended as well.
2. Source (process vs. results): Evaluation influence might stem from both evaluation results and the evaluation process.
3. Time (immediate, end-of-cycle, long-term): The time variable is made up of three categories, which Kirkhart admits are not perfect. Nevertheless, they allow for long-term impacts of an evaluation study.

Even though Kirkhart adds more precision, I will argue that her concept of influence on top of the ‘old’ use categories muddies the picture even more. For that reason, this thesis relies on the old categories but distinguishes between process use and findings use as two ‘meta’-types of evaluation use, in which three ‘sub’-categories of evaluation use pertains. This is described in the section below.

3.3.4 CHOICE OF CONCEPTUAL FRAMEWORK FOR THIS THESIS

If evaluators want to answer the one basic evaluation question of interest ('What works?'), then use is the dependent variable and thus not the primary object of study. In order not to make claims about inference too complex, 'less is more' in this regard and a few subtle categories allow researchers to present results more clearly and simply. Too many independent and dependent categories will make it impossible to reach any valid conclusion about the causality and generalise across different organisational contexts. A strong theory is generic and simple and for that reason this thesis uses a broad definition of evaluation use as described earlier.

In an important contribution, Alkin and Taut (2003) argue that process use is in fact not a fourth category of evaluation use, but rather a conceptual frame for other types of use. They argue that there are two categories of evaluation use; namely, 'findings use' and 'process use'. The difference between these two categories is that findings use takes place after the evaluation has been conducted, while process use takes place during the process of evaluation.

In both process use and findings use, instrumental, conceptual and symbolic/legitimising use exist. An instrumental change might thus happen during the evaluation process as the result of, for example, learning in the organisation and thus before the evaluation report materialises at the end of the evaluation. Process use may also occur conceptually. The evaluation process may change the attitudes of users about the importance of evaluation or about the potential role of multiple stakeholders in decision-making. This distinction between process use and findings use is important for this thesis and in particular in article 3, where it is applied.

Moreover, Alkin and Taut (2003) use Owen and Rogers (1999) distinction between symbolic use and legitimising use. However, they argue that legitimising use pertains to the findings use type, because it happens after the evaluation result to legitimise the intervention. Symbolic use, however, is more likely to happen during the evaluation process, because the evaluation process is used to justify the existence of the organisation.

In addition to instrumental, conceptual, legitimising and symbolic uses, the evaluation literature has also proposed two other categories that relate to the use of evaluation understood simply as a source of information – a type of use that often precedes other use forms (Alkin and Stecher, 1983; Finne et al., 1995). An instance of information use would be using evaluation information in a presentation or simply reading the evaluation to acquire knowledge. ‘Information use’ can take place both before and after the completion of the evaluation and is thus related to both ‘findings use’ and ‘process use’.

Finally, scholars have pointed to a fifth type of use often referred to as ‘strategic use’. Strategic use is distinguished from symbolic and other use types as it is not related to securing organisational or programme legitimacy, but rather to advocacy in relation to decision- or policy-making (Weiss, 1992; Pröpper, 1987: cited in Van der Knaap, 1995: 211). Strategic use needs to be included, because legitimising use originally proposed by Rich (1977) does not appropriately cover the strategic and political use of arguments found in evaluations and which is used to justify political arguments and decisions. Legitimising use is the justification by the evaluating organisation of the programme or policy being evaluated. However, an evaluation system involves other actors, who have an interest in using the evaluation as a source of legitimacy to back their positions and political arguments. This type of use I call ‘strategic use’, as it is not necessarily related to legitimising the programme (legitimising use) or the justification of the evaluating organisation (symbolic use). Instead it is related to other issues, such the use of facts from the evaluation to back a certain position in the renegotiation of a new programme.

3.3.5 LEARNING UNDERSTOOD AS EVALUATION USE

The broad approach of this thesis to evaluation use includes uses related to learning. Empirically, use and learning are difficult to differentiate, mainly because learning indirectly is assumed in several of the use categories (conceptual, instrumental, and information use). To simplify, in this thesis learning is considered an outcome of evaluation, similar to evaluation uses. This is done in order to broaden the scope of evaluation use and evaluation outcomes in the context of the evaluation system. Overall, learning is therefore considered a type of evaluation use.

Over the years, the knowledge-use literature overlapped the policy-learning literature, when policy learning theory developed in the 1960s as an alternative explanation of policy change. Previously, the stress had been on power and conflict, particularly among social classes. At that time, cognition and knowledge utilisation appeared as alternative explanations (Grin and Loeber, 2007). Simon (1957), Deutsch (1966) and Heclo (1974) were the first to take up these new ideas. Learning studies then gained a particular prominence in the 1980s and 1990s, when several different approaches were developed (Grin and Loeber, 2007). However, empirical studies in the field of policy learning are scarce (Dobbin et al., 2007; Freeman, 2006; Grin and Loeber, 2007; Dunlop and Radaelli, 2013).

According to Dunlop and Radaelli (2013: 3), there is a general feeling of disappointment (see also Egan, 2009; James and Lodge, 2003; Volden et al., 2008), and the field is still *'struggling to produce truly cumulative knowledge'*. It is the same case in the evaluation literature, where the focus on 'evaluation use' has dominated the focus on learning. This is another important reason for including learning in the conceptual framework related to evaluation use.

Following the general views on learning of James March (March, 2010), and more specifically on policy learning (Bennett and Howlett, 1992; Dunlop and Radaelli, 2013), learning is defined in this thesis as the adaptation of beliefs and views that result from the learner's sense-making of past experience. Therefore, the learning that emanates from a specific evaluation exercise is the adaptation of that specific policy area, programme or other government intervention that results from making sense of past experience systematically gathered through that evaluation. Naturally, for authors interested in studying the conditions for the use of evaluation results, the assumption is that organisational learning and the use of evaluation results are intrinsically linked (Preskill, 2008; Radaelli and Meuwese, 2010).

3.3.6 CONCEPT OF MISUSE

The concept of misuse is very much related to non-use (Cousins, 2004). Cousins (2004) distinguishes between two types of misuse; namely, 'mistaken use' and 'mischievous use'. The former relates to a situation in which a user's misuse of evaluation findings is not deliberate, typically because the user lacks an

understanding of evaluation methodology and analysis. The user may be guided by personal interest or organisational interests, which can lead to ‘cherry picking’ or biased stress on positive recommendations while leaving negative findings out of communications and subsequent decision-making. At its worst, misuse can be related to manipulation and behaviour that thwarts the improvement of policies or administrative practices.

This thesis does not make use of the term ‘misuse’ because of several drawbacks related to it. First, it is a normative term resting on a normative and often subjective foundation. Misuse is strongly related to the rationalist paradigm that evaluation is based on (see earlier sections). However, misuse is ‘the evil twin’ when evaluation findings are not used by a strategic and politicised organisation. The theoretical framework of this thesis assumes that organisations will do most things to stay legitimate and survive in the organisational field. In other words, the assumption of organisational institutionalism is that policy-makers are likely to use evaluations in any way they see fit to attain this objective. The concept of misuse is therefore laden with controversy and normativity. Evaluators are likely to be concerned about the ‘proper’ use of their evaluations, while administrators and politicians would take a more ‘strategic’ approach to using evaluation findings.

Second, and related to the first point, misuse looks like strategic or legitimising use, depending who the judge is. This thesis represents research that is as objective as possible within the scope of this thesis and its premises. For this reason, the term misuse is ill-suited for use in academic research, as it contains value-laden judgement about very intangible human intentions and actions. It is appropriate to use a more value-free conceptual framework that describes evaluation use better, such as strategic use and legitimising use. Whether these types of use are misuses would have to be discussed on a case-by-case basis.

Third, misusing an evaluation is not an end in itself and for that reason the term is not a strong concept in empirical investigation. Strategic use and legitimising uses have defined ends, while no administrator would admit to misusing anything. In most cases what some would term ‘misuse’, an administrator would describe differently.

Fourth and finally, evaluation misuse also relates to the empirical case investigated. If an evaluation was made in order to be used in a particular way, then it would be possible to identify misuse in relation to the actual way it was used. However, the evaluations investigated in this case (and in most cases) did not emphasise any particular subesquent use or intention of use. In practical terms, it is therefore up to the commissioner of evaluations to use evaluations in the way they see fit.

4 OBJECT OF STUDY

The object of study in this thesis is the European Union's evaluation system, with a particular focus on the special role that the European Commission plays in this system. The case is the European Commission, but the EU's evaluation system is the systemic context in which the Commission operates. The object of study is both the case and the context of the case. This is important to stress, because the theoretical assumptions made previously about how organisations seek legitimacy from their organisational field relates to the context of the organisation. Thus, in order to answer the research question and gauge whether evaluation use takes place in the system, we need to understand the systemic context of the Commission to understand what the demand is for accountability in this field.

4.1 CHOICE OF CASE

The object of study is the European Union's evaluation system and the case of interest is the European Commission. This case is interesting for the following reasons:

First, several studies have shown the Commission to be very preoccupied with accountability vis-à-vis the other EU organisations, such as the European Parliament, the Council and also EU Member States (Grøn, 2009). The Commission is a supranational organisation with powers often reluctantly transferred by the EU Member States. This, as well as the Sander Commission's demission in 1999, have increased scrutiny of the Commission and the Commission's interest in increasing its accountability. With the focus on accountability so pronounced, the Commission makes an interesting and extreme case to investigate.

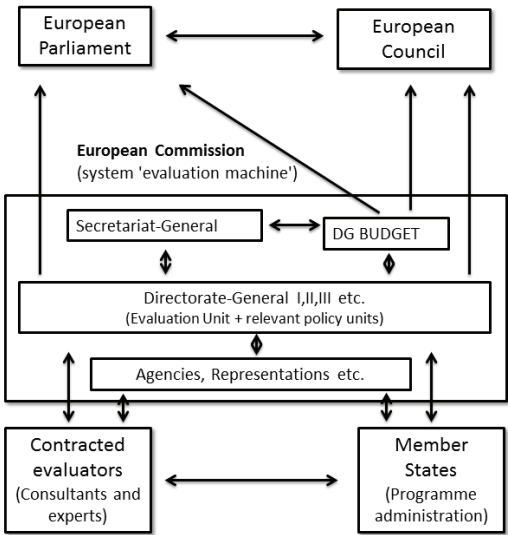
Second, the focus on the EU's evaluation system and the Commission's evaluation practices are novel in the academic literature. In fact the literature has very few empirical investigations of evaluation systems or of the evaluative practices in the European Commission (Stern, 2009). The literature on evaluation systems took a leap forward in 2006, as I explained above. Since then, contributions have been mainly theoretical. This thesis represents an empirical investigation of the effects of an evaluation system.

Third, the Commission has not only been absent as a case in the evaluation literature, but it is also highly relevant to study due to its importance in the global field of evaluation. The Commission’s evaluation practices are applied all over the world through EU Aid. Through EU enlargements, evaluative practices have spread across Europe as a prerequisite for getting Structural Funds or IPA-funding (candidate countries). It is therefore immensely important to study the European Commission’s evaluative practices, the extent to which evaluations are used, and their contribution to improving European policies.

4.2 THE EU EVALUATION SYSTEM

The European Union’s evaluation system is a case in point. Figure 2 illustrates the main constitutive actors and reproductive patterns in the system. The arrows signify the primary direction of communication in the system.

Figure 2 EU evaluation system



The EU evaluation system constitutes a relatively well-defined system within the EU institutions, as well as the EU Member States. In Figure 2, only the evaluators

are not (always) public organisations. In the EU evaluation system, the European Commission is the most active organisation in terms of numbers of evaluations. Over the last 30 years, the Commission institutionalised evaluation practices in response to internal and external pressures (see article 2). Practices and ideas were drawn from other organisations in the field and an EU evaluation system slowly emerged. Evaluation is a common practice in each Directorate General (DG) and is institutionalised through evaluation policy, guidelines and standards auditable through internal and external audits. In the Commission, the practices of evaluation are commonly referred to as the 'evaluation system'. This system encompasses all aspects of evaluation activity in the Commission, including formal institutions such as DG evaluation policies, evaluation functions, communications, the internal control standards, the Financial Regulation and so on. Also, evaluation is an integral part of the activity-based management and budgeting system of the Commission and thus formally related to decision-making regarding EU budgetary allocations. The system's core consists of the European Commission, the European Parliament (EP) and the European Council (Council).

In the Commission DGs, designated evaluation units supervise and guide evaluation activity with support from the Secretariat-General. The evaluation units are subject to internal audits as described in the Internal Control Standards of the Commission services. The Commission undertakes most evaluations in the system, but the EP and Member States also carry out or commission evaluations, usually subject to EU evaluation standards and supervised by the Commission (in the case of Member States). About 80 per cent of all evaluations are externalised to consultants or groups of experts (Commission and Jacobsen, 2007) and the consultancies are thus also part of the system. Overall, the EU evaluation system constitutes a relatively easily defined evaluation system. The example adheres to the definition proposed previously and thus was selected as a valuable and interesting case for this thesis.

5 RESEARCH DESIGN AND METHODOLOGY

This chapter elaborates on the research design and methodology applied in the four articles that constitute the empirical and theoretical contribution of this thesis. The articles vary in terms of type of data used and methodology applied. Apart from the first article, which is purely theoretical, the other three articles apply a case-based research design and a qualitative methodology. Since the first article is theoretical, only the methodology of the three empirical articles will be explained in more detail in the following sections. A distinction is made between research design, methodology and method. The following sections will elaborate on all three elements of this study.

5.1 CASE STUDIES

This thesis uses case studies in all three empirical articles. However, the case studies are different and constitute several types of case study research designs. This will be elaborated below.

Case studies are widely used in organisational studies and across the social sciences (Hartley, 2004). According to Stake (2000), case studies have become ‘one of the most common ways to do qualitative inquiry’. As in organisational sociology, case studies are also among the preferred research designs in the evaluation literature. For example, Easterby-Smith et al. (Easterby-Smith et al. (2000) note that *‘case study examples have always played an important role in the field. In the public sector, with quantitative measures much less well developed than for private firms, case studies have been especially prominent’*. The reason for this is straightforward, as organisations and evaluations alike are very contingent on complex environments and complex internal dynamics that are hard to systematise, generalise and define. Another important feature of at least political organisations as well as of evaluations is their tendency to be unique. In other words, there is rarely a population of political organisations or evaluations from where a sample can be extracted and inference made.

5.1.1 DEFINING A CASE STUDY

Some disagreement exists on exactly what a case study is. In this thesis, a case study is a research design that serves mainly to refine the focus of research. A case study is not a method (as indicated by Bryman), as case studies are not data

collection tools like interviews or surveys. Case studies are not methodologies either, because a methodology is a system of procedures and assumptions on how to analyse the data that have been collected. Consequently, case studies may be conducted either quantitatively, qualitatively or both, with several methods, data types and even several methodologies (Yin, 2003).

Stake (2000) defines a case to be a 'bounded system' that operates in a coherent and systematic way. In other words, a case is a stable system and its features can be observed and defined in time and space. The social phenomenon, which cannot be defined in time and space, is not a case, because it is undefined and indistinguishable from its environment. To draw closer to a precise definition, Stake argues that a case study 'is not a methodological choice but a choice of what is to be studied'. (Stake, 2000: 435). The approach in this thesis agrees with Stake in relation to the latter.

Therefore, this thesis is closer to the point of view of Yin (2003: 13-14), who places less emphasis on the boundedness of cases: 'A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context.' In fact, Yin stresses the case study's ability to investigate the complexity of cases and their environments, making it the point of case studies to study complex social phenomena with complex contexts. Bent Flyvbjerg (2006) concurs with Yin, stressing that case studies are very context dependent and provide concrete and tangible understandings of social phenomena in all their complexity. He adds that case studies provide value to social science exactly because of the complexities of social phenomena and the consequent lack of *universal and predictive* theories inherent to the social sciences (Flyvbjerg, 2006). Following Yin and Flyvbjerg on this point means accepting that a system will never be completely bounded or isolated in the social world, and that the researcher has to be open to contingencies on the 'out of system' context.

The case studies in articles 2, 3 and 4 are all *explanatory* (as opposed to descriptive and exploratory (see Yin (2003)). They focus on a given phenomenon (use, learning and the evaluation system) and describe it while also trying to explain it. Article 2 is a single longitudinal case study and article 3 is a single case study. Article 4 is a multiple case study involving three programme cases. Stake

(2006: 437) distinguishes between *intrinsic case studies*, *instrumental case studies* and *collective case studies*. The first two are both one case study, which is either the object of study itself or an instance of the case study being used to study another object of study and where the case is *instrumentalised* to shed light on this very phenomenon. In this thesis, article 4 uses several cases to research the phenomenon of policy learning from evaluations in the context of the evaluation system. This is therefore a collective case study. Article 3 is an instrumental case study, where four LIFE evaluations are used to understand how the Commission uses evaluation in the context of the EU's evaluation system. Article 2 is an intrinsic case study of the European Commission's evaluation system.

5.1.2 CONSIDERATIONS ON SCIENTIFIC CRITERIA

The strength of a case study design is its richness in terms of understanding the case itself and its relations with its environment. Because case studies are very suited to analyse complexity, 'how' and 'why' are the questions most commonly answered through case studies (Yin, 2003). A case study develops deep knowledge of a particular case and all or most of its features and attributes, such as its history, development, internal processes, appearance, reasons for actions, institutions and so on. On the other hand, case studies are not suited for studies that aim to generate generalisable knowledge in time and space (Bryman, 2012; Yin, 2003: 10). Yin (2003) describes it like this: 'case studies [...] are generalizable to theoretical propositions and not to populations or universes. In this sense, the case study [...] does not represent a 'sample', and in doing a case study, your goal will be to generalize theories (analytical generalization) and not to enumerate frequencies (statistical generalization)' (p.10).

In other words, the external validity (or generalisability) of case studies is considered to be low (see Flyvbjerg (2006) for an interesting discussion on this point). On the other hand, the internal validity of case studies is high. The case study allows the researcher to understand causal relationships because of the in-depth understanding of the case. Moreover, the construct validity (or conceptual or measurement validity) is also high (Bryman, 2012; Flyvbjerg, 2006). The case study is an excellent approach to testing or understanding social science concepts, because it goes into depth with the particular phenomenon.

Another important feature of case studies is their ecological validity, which is the research design's level of application to the real world. Ecological validity exists if the data subjects and practitioners find the research findings relevant and reliable. This is particularly relevant for an industrial PhD thesis such as this one, which relies heavily on data from practitioners of one particular organisation (the European Commission).

In the field of knowledge utilisation, case studies have been widely used (Cousins et al., 2004) because not many cases or observations are available. By far the most common unit of analysis is an organisation. Organisations are plentiful, but within a given field and context such as evaluation use, the sample of organisations available to the researcher is often limited. Collecting quantifiable data from organisations is another difficulty, as organisations as well as their evaluations differ. More importantly, knowledge utilisation is from the outset a qualitative judgement from the observer (the researcher) or the interviewees in the organisation. A quantification is difficult, because the phenomenon analysed in this study is qualitative in nature.

5.1.3 CASE SELECTION

The case study is the methodological framework of this study. First, the sample of organisations is small and too small to warrant sampling DGs to produce inferential statistics. Second, to answer the research question, it is necessary to gain qualitative knowledge of the evaluation institution and the use of evaluation knowledge in the DGs in question, and to understand the complex organisational dynamics at play within the organisations as well as between the DGs in the European Commission. A case study provides a better understanding of the qualitative relationships, processes and causal effects of the study object and the factors that affect it and how it affects evaluation use. Moreover, a case study typically produces high conceptual validity, which is needed in the literature on evaluation use, and which is well suited for generating hypotheses relevant to a larger population (political organisations more generally). The weakness of a case study is its particularity (the difficulty of generalising the findings to a population) and its weak insight into the population as a whole (width) as well as the dynamics that can be found there. However, as the EU (and the Commission) claims to be a

sui generis organisation, from the outset it is difficult to talk about a genuine ‘population’ of complex international organisations such as the EU Commission.

5.2 ARTICLES

The research designs applied in the three empirical articles are all case-based, including longitudinal case study (2nd), single case study (3rd) and multiple case study (4th). An overview of the four articles and their respective research designs, methodologies, methods and data sources is provided below.

Table 5-1 Overview of article methodologies

No	Title	Research design	Methodology	Methods	Data sources
1	Evaluation use in the organisational context – changing focus to improve theory	NA	Literature review	Bibliographic search	Scientific articles
2	Evaluation in the European Commission – for learning or accountability?	Longitudinal case study (The European Commission’s evaluation system)	Qualitative content analysis Historic analysis	Unstructured expert interviewing Document review	Transscribed interviews Documents Homepages Literature
3	Evaluation use in evaluation systems – the case of the European Commission and its LIFE programme	Single case study (LIFE programme)	Qualitative content analysis	Semi-structured interviewing	Transscribed interviews Documents
4	Evaluation and Policy Learning –the learner’s perspective	Multiple case studies(LIFE programme, Intelligent-Energy Europe II programme and FP7)	Qualitative content analysis	Structured interviewing	Transscribed interviews

5.2.1 FIRST ARTICLE

The first article is purely theoretical, and uses organisational institutionalism to create a model for different types of evaluation use. It is based on a literature review that focuses on the application of organisational institutionalism in the evaluation literature. Therefore, this article is the theoretical starting point for this thesis, as it discovers that organisational institutionalism has not been used to

investigate evaluation use in the evaluation literature. The article thus discovers a gap that is filled theoretically by itself and the other articles in this thesis.

5.2.2 SECOND ARTICLE

The second article is based on a case study of the Commission's implementation of evaluation practices over time. Thereby, the article investigates how an evaluation system is implemented and how the issues of accountability and learning are reflected in the practices implemented.

The article is based on data generated from 58 recorded interviews, two group interviews and one conference on evaluation in the EU, along with numerous informal talks with experts and Commission desk officers and personal observations including evaluation steering committee meetings. Interviewees were sampled purposefully and according to availability, and 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 who had performed key roles in the early implementation of the evaluation system and thus had a good historical overview of evaluation in the Commission. Interview data were validated with document data comprising more than a hundred public and non-public documents, such as internal evaluation policy papers, guidelines, and minutes of meetings in the evaluation network (on document analysis see Stewart and Kamins, 1993).

Data were analysed according to the principles of qualitative content analysis (Schreier, 2012; Mayring, 2002; Mayring, 2000; Mayring, 2004; Kohlbacher, 2006) and by using the NVIVO software package (Bazeley, 2013). Qualitative content analysis focuses on both latent and manifest content through stressing the visualisation of patterns in the data. The data were coded according to the interviewees' perceptions of accountability and learning, as well as their rationalisation of the Commission's evaluation practices. Findings were triangulated and validated with document data, and interviewees had the opportunity to comment on the findings. Finally, the reliability of the findings was strengthened by the author's prolonged engagement in the field over a period of

two years as well as five years of professional experience with evaluation of EU policies.

5.2.3 THIRD ARTICLE

The third article analyses the use of four evaluations of the Commission's Programme for the Environment and Climate Action (LIFE) over a ten-year period. The LIFE programme was chosen because it is a 'classic' centrally managed EU expenditure programme (parts of the programme was externalised to EACI in 2014). The programme has experienced a full Commission evaluation cycle (ex ante, midterm, final and ex post), and therefore represents a complete picture of evaluation use over an entire policy cycle as well as an entire evaluation cycle. Moreover, by looking at four evaluations over time, we get a more diversified idea of the effects of the evaluation system on the use of evaluations.

The analysis is based on 16 semi-structured in-depth interviews and eight follow-up interviews. The informants were sampled purposefully according to relevance and availability and consisted primarily of staff from the Directorate General for the Environment (DG ENV), consultants who performed the evaluations, representatives of Members of the EP's Committee for the Environment (member assistants) and Council members. The data collection continued until 'saturation' was achieved. Document data were also coded and used primarily to validate interview data through triangulation (Bryman, 2012). The document analysis comprised four retrospective evaluations of the LIFE programme (midterm, 2003; ex post, 2009; midterm, 2010; final, 2013) and several other documents, including DG ENV presentations to the Committee of Regions and the EP, internal Commission documents, the combined ex ante and impact assessment along with fiches and Commission position papers for the new LIFE programme 2014–2020.

The methodology applied in the paper is based on the principles of qualitative content analysis, and the coding and analysis of data was carried out using the NVIVO software package. The first 16 semi-structured interviews were analysed with a view to existing conceptual frameworks developed in the evaluation literature and described earlier, and the eight follow-up interviews were conducted to check for saturation. Coder reliability was sought by using the existing conceptual frameworks and subsequently running three rounds of coding on the

interview data (Kohlbacher, 2006). The credibility of the findings was further strengthened by a prolonged engagement in the field, with the researcher conducting interviews in four consecutive waves over a period of one year. In addition, the researcher has several years of experience with evaluation of EU programmes, including working as evaluator on the ex post evaluation of LIFE from 2009. Findings were triangulated and validated with document data, and follow-up interviews and interpretations were checked against interview data. Interviewees were debriefed and were given the opportunity to comment on the findings of the article, and peers with comprehensive knowledge of the subject gave important comments to the draft article before submission.

5.2.4 FOURTH ARTICLE

The research design of the fourth article is an example of a collective of three cases. The case evaluations are midterm or interim evaluations commissioned by the European Commission and produced by external parties using a methodology comparable to that of all European Commission midterm evaluations. The cases are of similar budget sizes for the programme cycle 2007–2013. Moreover, the programmes work in similar ways through project support and are all managed by the Commission or by a Commission executive agency. 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 programme 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).

The fourth article was based on data generated from 25 structured in-depth interviews from the three case evaluations. The interviews were structured around 20 questions that related general features of the evaluation system to learning from evaluation. The questions probed the design and implementation phases of the evaluation, as well as the learning from the evaluation. Interviewees were sampled purposefully and according to availability, and included Commission officers in the programming unit (the principal learners) as well as Commission officers in the steering committee that managed the evaluation implementation. External experts and consultants directly involved in the evaluation implementation were

also interviewed. 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 with the two previous articles, the methodology applied in this article was qualitative content analysis, and the coding and analysis of data was carried out using NVIVO. As described by Mayring, the data was classified (coded) along the lines of the existing conceptual framework from Dunlop and Radaelli (Dunlop and Radaelli, 2013), basically following the interview guide.

5.3 METHODOLOGY

The data analysis in the three empirical articles is qualitative. The choice of methodology is contingent on the nature of the research question(s) and the available data, two factors that are mutually dependent as well. For this study, quantitative data were not available in relation to evaluation use and learning in the European Commission. Therefore, the articles in this thesis apply qualitative methods to generate qualitative data. Consequently, a qualitative methodology was chosen as analytical framework for the data analysis. Qualitative methodology is very suitable for asking ‘why’ and ‘how’ questions, which are the questions asked in the articles of this study. According to Cassell and Symon (1994: 1), qualitative methods are the most appropriate to research questions focusing on organisational processes and outcomes.

According to Cassell and Symon (1994: 4), qualitative research is ‘less likely to impose restrictive a priori classifications on the collection of data’. Consequently, qualitative research is more likely to be inductive, explorative and concerned with emergent concepts and themes rather than with the testing of hypotheses. However, several qualitative methodologies can also be used with existing conceptual frameworks deduced from state-of-the-art literature. Qualitative methodologies generally emerge from phenomenological and interpretive epistemologies where social reality is inter-subjective and negotiated (Cassell and Symon, 1994: 2). The assumptions behind the choice of methodology are also described in the next section. Within qualitative methodologies, the most prominent in organisation studies are ethnography, symbolic interactionism, historic methodologies, discourse analysis and qualitative content analysis. This

study applies qualitative content analysis as described more in detail in the next section.

The core of qualitative research is the ‘search for patterns’, where the researcher looks for regularities and associations between different parts of the data categories. Bernard and Ryan (2010) describe qualitative data analysis in the following way: ‘Analysis is the search for patterns in the data and for ideas that help explain why those patterns are there in the first place.’ In other words, analysis involves disintegrating data parts and bringing them together again to look for patterns. Similarly, Dey (1993: 42) argues that the essential part of qualitative analysis is breaking up data and compiling it in new ways. This classification is often conducted through coding text or transcripts of interviews. Graphic representation of results is important in qualitative data analysis for communicative reasons as well as for replicability.

Qualitative data analysis tends to be inductive and exploratory, but it can also integrate theoretical or research-based concepts or assumptions about the way classifications are constructed. Bernard (2013) emphasises the importance of building assumptions from tacit knowledge or research, and of exploring the associations between these assumptions. In relation to analysing qualitative data, Miles and Huberman (1994) summarises the process as three steps: 1) data reduction; 2) data display; and 3) conclusion. Data display involves reducing the complexity of data to ‘manageable sizes’ that can be meaningfully communicated to the reader to account for the analytical processes. Displaying data practically and visually illustrates the patterns in the data in such a way that the reader can relate to the findings of the research. This was particularly relevant but also difficult to do in this thesis, due to the article-based format chosen.

The ‘groundedness’ of, or ‘closeness’ to, the observed phenomenon is a key asset of qualitative studies. Qualitative research is also excellent for collecting meanings, perceptions and assumptions that people might have about a given phenomenon. The main weakness of qualitative research is the lack of standardised procedures in the field (Miles and Huberman, 1994). The consequence of a lack of clear procedures is a low level of replicability, and thus a weakened trustworthiness or reliability. In qualitative research it is generally not

reported how data are reduced and classified, analysed and interpreted (Miles and Huberman, 1994). This weakens not only the transparency of a study, but also its replicability or reproducibility, and has consequences in relation to trustworthiness and reliability. In fact, the problem of reliability is central to the critique of qualitative social science. The ‘deconstruction’ of data into classifications or codes is based on subjective interpretations made by the researcher. Since it is often hard to replicate or reconstruct qualitative social science studies, inter-coder reliability during the study becomes more important. This can be established in several ways but, usually, to ensure consistency of coding several coders will review the same material. Also, results can be validated by experts or interviewees or triangulated with findings from other data analyses or documents (Mayring, 2004). Inter-coder reliability is probably the most effective way of establishing internal consistency and reliability. However, it is also demanding in terms of time and resources.

Another criticism often raised about qualitative research is the generalisability or the inference of results. Generalisability refers to ‘the degree to which the findings are applicable to other populations or samples’ (Ryan and Bernard, 2000: 786). Thus, it draws on the degree to which the original data were representative of a larger population (*ibid*). Narratives, interviews and documents can have varying degrees of idiosyncrasy and particularity that is hard to generalise.

5.3.1.1 QUALITATIVE CONTENT ANALYSIS

Qualitative content analysis develops from quantitative approaches to content analysis, where words and other attributes of texts were counted to say something about the texts. As the name implies, qualitative content analysis focuses on the qualitative aspects of text and thus its explicit and implicit meaning. Four types of qualitative content analysis have emerged. They are cross-case analysis, thematic analysis, grounded theory and simple qualitative content analysis. While grounded theory is commonly regarded as an inductive theory-generating methodology, qualitative content analysis is appropriate and relevant to answer how, why and when questions and also questions involving existing theoretical constructs. Qualitative content analysis does entail some degree of interpretation, but not at the level of methodologies used in hermeneutic traditions. At the same time, it does not quantify as much as thematic analysis and cross-case analysis. Qualitative content analysis involves numbers and the counting of ordered

classifications, categories or data break-downs, but also allows for the interpretation of text. In fact, the strength of qualitative content analysis lies in its combination of qualitative interpretation with quantitative elements according to techniques used by classic content analysis (Titscher et al., 2000: 64; Remenyi et al., 2002: 6).

The data-reduction techniques in qualitative content analysis reduce large amounts of qualitative data to categories of manageable sizes that can be analysed quantitatively. The systematic approach gives the methodology increased validity and reliability. At the same time, the methodology retains the reflexivity and interpretation that are the main characteristics of qualitative methodologies in general. Schreier (2012) defines qualitative content analysis as follows:

‘Qualitative content analysis is a method for systematically reducing data and describing the meaning of categories through latent examination of their context.’

Bryman (2012: 542) offers a rather similar definition: ‘[Qualitative content analysis is] an approach to documents that emphasizes the role of the investigator in the construction of the meaning of and in texts. There is an emphasis on allowing categories to emerge out of data and on recognizing the significance for understanding the meaning of the context in which an item being analysed (and the categories derived from it) appeared.’ Mayring (2000), who is the main contributor to qualitative content analysis (Kohlbacher, 2006), offers the following definition of qualitative content analysis: ‘An approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytical rules and step by step models, without rash quantification’. Mayring stresses four elements of qualitative content analysis. First, data need to be reduced and segmented into meaningful units. Second, the researcher looks for patterns in the segmented data. Third, the results are verified by checking results with other explanations, contradictory data or ‘outliers’. Finally, the results are displayed using tables or graphs to illustrate the distribution of categories, etc. Data display is an important point in illustrating the break-downs made by the researcher and thus in giving the reader access to the interpretations of the researcher and a better foundation for questioning the results.

The key feature stressed by Mayring in relation to qualitative content analysis is ‘structuring’. Structuring is the classification and categorisation explained earlier

that allows for a quantification of the data material. This procedure is related to classic content analysis that is mainly quantitative. The classifications and the categories emanate from the unit of analysis and might or might not be established on a theoretical basis. During the analysis, the categories and classifications are evaluated and altered and, if necessary, the text is recoded accordingly. Finally, qualitative examples are extracted together with a data presentation.

It follows from the above that qualitative content analysis can both be inductive and be based on assumptions or existing categories developed from theory. The methodology in the work of Mayring emphasises mainly theory-driven research, but inductive categories can also be developed with the procedures of qualitative content analysis (Mayring, 2003: 74-76). For the purpose of this study, qualitative content analysis is applied differently in the three articles. Article 2 is the most inductive, with open questions and little prior theorising. Article 3 uses established concepts and semi-structured interviews that allow for more quantification of categories. Finally, the structured interviews of article 4 that made it possible to do an even more quantifiable version of qualitative content analysis is much in line with Mayring's writing.

5.3.2 METHODS AND DATA

The main method applied in the three articles is in-depth interviews (Legard et al., 2003). In the three empirical articles, 58 interviews were conducted. The interviews were conducted as semi-structured interviews taped and transcribed. Article 2 relies on all the interviews and, in particular, on 35 open semi-structured interviews conducted primarily with desk officers and consultants who work (or have worked) with evaluation in the Commission. Articles 3 and 4 rely on 16 and 25 interviews respectively. In article 3, the 16 interviews were semi-structured, whereas article 4 used a more structured approach with 20 interview items. However, both approaches qualify as 'semi-structured', as they both allowed the interviewee to deviate to some extent from the interview questions and because all the questions were open.

Interviews were conducted with desk officers in the respective evaluation units, heads of unit and desk officers responsible for policy, as well as with some of the key consultants who carried out externalised evaluations, and relevant researchers

and other stakeholders. The interview questions followed directly from the hypotheses and were open to allow the interviewee to describe processes and attitudes in detail. In this regard, the thesis was driven by conceptual deduction, but at the same time also encompassed conceptualisations and insights from the interviewees that might finally be constitutive for theoretical improvements. The interview data were analysed in NVIVO and coded. The line of coding followed from the typologies developed earlier on evaluation institution, learning and knowledge utilisation.

6 CONCLUSION

This thesis investigates the effect of evaluation systems on evaluation use. The articles constitute the theoretical and empirical work of the thesis as they respond to the three sub-questions that support the overall research question. The three sub-questions relate to the three gaps identified in the literature. Table 6-1 summarises the research question and the sub-questions.

Table 6-1 Research question and sub-questions

<i>What effect do evaluation systems have on the use of evaluation?</i>
1) <i>How can non-use and justificatory uses of evaluation be explained?</i> (addressed in articles 1, 3 and 4)
2) <i>How can evaluation use be explained in its systemic organisational context?</i> (addressed in articles 1, 2, 3 and 4)
3) <i>How are evaluations used in evaluation systems?</i> (addressed in articles 2, 3 and 4)

6.1 ANSWERING THE SUB-QUESTIONS

The following sub-sections answer the sub-questions of this thesis.

6.1.1 HOW CAN NON-USE AND JUSTIFICATORY USES OF EVALUATION BE EXPLAINED?

In article 1, the phenomena of non-use and justificatory use are investigated theoretically through the lens of organisational institutional theory. By assuming that organisations will implement technologies such as evaluation in order to legitimise themselves in the face of other organisations in the system, this article reinterprets the *raison d’être* of evaluation completely, when compared to previous assumptions in the evaluation literature that regard evaluation as a feedback mechanism to improve policy- making.

In this new theoretical framework, non-use of evaluations and justificatory uses are explained by the internal propensity to evaluate as well as by the external pressures on the organisation. Article 1 illustrates four outcomes of adaptation to the system that organisations will follow depending on these two variables.

In articles 3 and 4, the EU evaluation system is investigated and several empirical explanations for the non-use and justificatory use of evaluations were found. Articles 3 and 4 do not use the theoretical framework developed in article 1. Nevertheless, the overall findings fit very well to the ‘mimetic’ model of adaptation envisaged in the theoretical framework for organisations that have a high propensity to evaluate as well as high external pressure. The Commission has both, and thus evaluation practices have been institutionalised both formally and informally in the Commission, as predicted by the theory developed in article 1.

Articles 3 and 4 also illustrate how important the evaluation system is when explaining non-use and justificatory uses. The evaluation system is designed to deliver sound, credible and independent evaluations of high quality for use at particular decision points at the end of programme periods. This is to increase bargaining power for the Commission and legitimise the programme at the time decision-makers’ attention is likely to be on the programme; that is, just before renegotiation of the new programme budget in the programming cycle of the Commission. However, this focus on evaluation use in the decision point reduces use (particularly process use) during the programming period before the decision point is reached.

Moreover, if evaluations miss important decision points, the relevance of evaluations is affected negatively, potentially leading to non-use. Further, evaluation use is hampered by competing information in the system such as other studies and audits. Due to mistiming of evaluations, other competing information and non-innovative findings, the relevance of evaluations’ to decision-making is very limited. These two findings imply a de facto de-politicisation of programme evaluations in the EU evaluation system, where evaluation information conforms to the administrative context of programme management in the Commission instead of the political context of policy-makers.

For these reasons, programme evaluations are rarely used for policy-making in the evaluation system. They are used for justificatory purposes but, according to one interviewee in the Commission, the legitimacy is already built into the system itself.

Also, the legal structures that govern the Commission's work reduce the possibilities of innovative and surprising recommendations in the evaluations, because consultants' recommendations are steered first of all by the evaluation questions in the tender material that is typically in accordance with what the Commission can influence within the scope of its legal competences. Thus, the formal structures of the evaluation system steers the evaluation use in a very particular direction, which is almost always within the boundaries of what is possible for the Commission to implement given the present legal framework it is working under. This effectively narrows the potential scope for evaluation use.

Article 3 finds that the 'de-politicisation' represents a paradox, since it is the EP and the EU Member States that compelled the Commission to develop its evaluation practices in the first place. But this article shows that policy-makers rarely use the evaluations; nor does the Commission maximise its use of evaluations although it allocates considerable resources to evaluations and numerous other internal and external assessments and audits.

Overall, articles 3 and 4 describe how the potential scope of evaluation use is reduced considerably by the formal institutionalisation of the evaluation system. The relevance to policy-makers and administrators of evaluation is reduced in different ways. For policy-makers, there is a lot of competing information and evaluations are too general and broad-scoped to be used strategically. If evaluations are not critically negative, then they will not be used by policy-makers in the Council and the EP. At the same time, evaluations do not bring to light a lot of new knowledge for programme administrators in the Commission, who are already very knowledgeable about the programme they administer. Often, evaluations are mistimed in relation to decision points, or they are too close to a previous evaluation to make a significant difference.

6.1.2 HOW CAN EVALUATION USE BE EXPLAINED IN ITS SYSTEMIC ORGANISATIONAL CONTEXT?

All the articles of this thesis contribute to answering this question. Article 2 explains the implementation over time of the EU evaluation system. Through this analysis, it is clear that the evaluation system is tied closely with the activity-based management system of the Commission and introduced mainly to hold the Commission accountable for the large expenditure programmes that the Commission manages.

The linking of evaluations and their timing to the programming cycle has consequences both negative (non-use) and positive on evaluation use. Article 3 finds that the Commission's programming cycle (or policy cycle) enables findings use because the evaluation system is designed to deliver feedback into a particular decision point at the time for programme renegotiating. This decision point enables findings uses including instrumental, legitimising and strategic use. These are enabled because the decision point allows for potential programme change and because the Commission needs programme legitimacy and overall accountability when the focus of policy-makers in the EP and Council is on the programme.

Because of the coherence between the evaluation system and the programming cycle in the Commission, the evaluation system is designed to deliver independent quality evaluations that are timely, well communicated, and credible in order to secure the legitimacy of the programme and the strategic position of the Commission in such negotiations. The institutionalisation of evaluation practices (including the highly standardised evaluation process, staff training and guidelines) almost completely negates process uses, as explained previously. However, it also enables a smoother execution of evaluation processes and ultimately delivers a more standardised evaluation product with minimal 'risks' to the Commission (Højlund, 2014a).

Overall and as explained above, the scope for evaluation use is considerably limited by the formal structures of the evaluation system. However, within the narrow framework of the evaluation system, evaluations are produced to be used and to increase learning in the Commission as well as outside the Commission. Articles 3 and 4 find that the Commission's programme units are the main users of

the evaluations for small-scale adjustments to the programme or conceptual or information use. Evaluators also learn from conducting the evaluations for the Commission. However, apart from these two stakeholder groups that are both involved in evaluation implementation, very few people use the evaluations directly.

As explained, the evaluation system is aligned with the programming cycle as well as the programming and planning of new interventions. Thus, impact assessments are required to take into consideration previous evaluations. Also, the draft budget needs to be based (to the extent that it is possible) on previous evaluation of budget lines. These initiatives do testify to a willingness to use evaluations actively for decision-making, but the impact on actual decision-making is very limited.

6.1.3 HOW ARE EVALUATIONS USED IN EVALUATION SYSTEMS?

Articles 3 and 4 in particular address this question through empirical investigation. They find that evaluations are used in a number of ways, ranging from making small-scale instrumental changes to the programme to more conceptual learning about the programme. The programme unit and the evaluators are the main users of the evaluations, although some evidence exists that policy-makers in the EP also use evaluations strategically and to get information.

Article 2 finds that the Commission's evaluation practices are implemented to increase controls and oversight of the Commission by other organisations in the evaluation system. The evaluation system is introduced for purposes of accountability over those of learning. However, a focus on learning does coexist with accountability and the two are therefore not mutually exclusive. Article 3 finds that the EU evaluation system is conducive to evaluation use of programme evaluations in several ways. Evaluations are used instrumentally to improve the programme and strategically to acquire knowledge or for advocacy purposes. Evaluation use takes place in windows of decision-making and typically in relation to an impact assessment or ex ante evaluation.

In relation to use, the most striking finding was probably that evaluations are not used during the evaluation process. The evaluation system dictates a very formal and independent type of evaluation where the evaluator is relatively remote from

the evaluation management (steering committee) in the Commission. This is because the EU evaluation system is primarily designed to feed information into the EU decision-making procedure every seventh year, before the beginning of a new programme cycle. Thus, due to the institutionalisation of evaluation practices that are fitted to the activity-based management system in the Commission, the use of evaluations takes place mainly just before important decision points. The limited process use is a consequence of deliberate choices to secure and improve findings use for decision-making, in particular decision points before programme renegotiation. Thus, the loss of process use is a direct consequence of systemic factors related to policy-making practices in the EU political system that influence the Commission's decisions in relation to implementation of evaluation practices.

6.2 ANSWERING THE RESEARCH QUESTION

Based on the findings of the articles that reply to the three sub-questions, I direct my attention to the main research question. The conclusion from the case study of the EU evaluation system is that evaluation systems have an effect on evaluation use and that this effect is significant.

6.2.1 EVALUATION SYSTEMS HAVE AN EFFECT ON EVALUATION USE

The articles find that evaluation systems have an effect on evaluation use. Theoretically as well as empirically, the articles demonstrate how formal and informal structures such as evaluation guidelines, standards and administrative practices and rules affect evaluation use in various ways.

The EU evaluation system is constituted by the Commission, the European Parliament, the Council and several other organisations. Together they partake in the system by playing different roles. The Commission plays the most active role due to pressures from the Member States and the European Parliament for more oversight and control with EU-spending. As a consequence of the external pressure, the Commission adopted and institutionalised evaluation practices with guidelines and standards that included formally and legally committing the DGs to evaluation. Through this process, the evaluation system was created as a common normative framework, and formal guidelines for evaluation were put in place. Evaluation was delegated to particular bodies within the Commission and was made a repeated praxis through the policy cycle and the activity-based

management system in the Commission. Hence, the system achieved permanence and all the attributes required for it to meet the definition of an evaluation system.

By analysing the EU evaluation system with a particular focus on the European Commission, the thesis demonstrates how formal structures are introduced to increase oversight of the Commission by other organisations in the system and how evaluation is used to increase accountability in the Commission. Article 2 finds that evaluation is, in fact, primarily institutionalised in the Commission for accountability purposes. The evaluation system is thus set up with a main aim of securing the legitimacy through accountability for the Commission. Nevertheless, articles 2, 3 and 4 all show that, despite this aim of the evaluation system, there is still room for evaluation use within the framework of the evaluation system's rules and standards.

The three main effects of the evaluation system on evaluation use are: 1) the 'sacrifice' of process use for findings use and accountability in decision points (explained above); 2) a very narrow scope for evaluation use, due to the formal institutionalisation of evaluation; 3) a de-politicisation of evaluation.

First, and as a consequence of the policy cycle, the possibility of evaluation use in the evaluation process is decreased because of the tightly managed and standardised evaluation process and the stress on evaluator independence that ultimately secures the legitimacy of the evaluation output and the Commission. Also, **process use** is sacrificed as a logical consequence of the fact that programme changes are usually attainable only in the design phase of a new programme (and not during its implementation), at which time the Commission needs credible, trustworthy and independent evaluations to increase its own legitimacy as well as that of the new proposal. In other words, process use as envisaged by, for example, Michael Q Patton, with a very interactive and engaging evaluation process to bring about significant programme changes immediately during the evaluation process, is in the EU context unlikely to yield more programme changes because of the rigidity of the decision-making and the legal constraints the Commission works under.

Second, evaluation recommendations tend to suggest small procedural programme changes rather than large-scale programme changes that only the EP and Council

could decide upon. Evaluation findings and recommendations for programme change are usually not surprising or innovative because political and legal structures limit the implementation of recommendations. For example, the terms of reference, the evaluation questions and the Commission staff guide evaluators towards recommendations that are feasible on a managerial level and that do not require a political decision-making procedure or that go against the legal structures of the Commission.

Third, the two previous findings imply a de facto **de-politicisation** of programme evaluations in the EU evaluation system, where evaluation information conforms to the administrative context of programme management in the Commission instead of the political context of policy-makers. Evaluation does not challenge the policies because evaluation is institutionalised, formally and informally, so that evaluation content and outputs conform to programme-specific recommendations. As an example, out of the three case evaluations of article 4, only one recommendation was not possible for the Commission to implement on its own. The recommendation was to increase the budget of the programme in question, which might be the reason why it was not objected to by the Commission during the implementation of the evaluation.

We have now seen how the evaluation system has several effects on evaluation use both inside and outside the Commission. The evaluation system limits the potential evaluation use, but also enables or improves some uses at specific times. In other words, the formal and informal structures of the system create a playing field where only a certain range of outcomes is possible. However, at the same time the evaluation system's standards and procedures, as well as staff commitment to evaluation and experience with evaluation, also enable evaluation use by securing a usable evaluation product. Evaluations are highly complex and potentially very political. The Commission's standards and procedures for evaluation implementation limits the possible evaluation outcomes, as explained above, but at the same time they increase the chance of a legitimate, sound and relevant evaluation product that is more likely to be used by the Commission.

6.2.2 OTHER FINDINGS

A number of other findings from the four articles are indirectly linked to the research question. First, the articles together show the importance of analysing phenomena such as evaluation and evaluation use in their systemic organisational context. When trying to explain evaluation use, the evaluation literature has focused on the evaluation much more than on the context of the evaluation. The main contribution of this thesis is to introduce to the evaluation literature empirically tested assumptions of organisational institutionalism, thereby illustrating that a theory of organisation is better at explaining evaluation uses than evaluation theory. The purpose of the evaluation system is to secure the Commission's accountability. Justificatory use is therefore the most important type of use for the Commission and this *raison d'être* can explain why process uses are not made possible in the evaluation system and why findings uses are significantly limited to mainly small-scale programme changes.

An important finding of this thesis is that the concept 'evaluation system' needs more theoretical depth. If an 'evaluation system' is defined only in terms of its boundedness, units and institutionalisation then we fail to understand how accountability and organisational effectiveness affects evaluation practices and evaluation use. This thesis shows very clearly how organisational accountability plays an important role in determining how evaluations are used.

6.3 PERSPECTIVES ON FINDINGS

The conclusions of this thesis point in several directions and have implications for the setting up of evaluation systems. Some of these implications have been raised before, but are nevertheless all the more relevant, given the conclusions in this thesis regarding non-use.

6.3.1 STUDIES RATHER THAN SYSTEMS

One of the main points in the book *From studies to streams* (Rist and Stame, 2006) is that evaluations used to be implemented as ad hoc studies rather than as streams of evaluations implemented in evaluation systems. One of the defining features of an evaluation system is that evaluations take place continuously and not on an ad hoc basis.

However, during the data collection for this thesis, several interviewees argued that evaluations were too broad in scope and did not bring about anything new because they were produced so frequently that nothing new had really happened since the last evaluation. Also, some interviewees argued that the targeted and specific studies they commissioned were much more useful, as they typically helped mitigate a problem in programme implementation and because they could be commissioned without adhering to the formal criteria of evaluation dictated by the evaluation system's rules and guidelines.

Therefore, administrators looking to set up an evaluation system should consider if learning objectives are not more easily met through ad hoc studies rather than through streams of standardised evaluations. There is a trade-off between, on one hand, learning and evaluation use and, on the other, the accountability that is vested in the system.

6.3.2 PERFORMANCE AUDIT INSTEAD OF EVALUATION

As this thesis concludes, evaluation systems are set up mainly to secure accountability in relation to programme spending. However, audits are already performed on spending programmes in the EU evaluation system and performance audits also investigate programme effectiveness and efficiency. Therefore, it is worth considering the *raison d'être* of evaluation in the face of the development of performance audits.

Audit is vested with much more credibility than evaluation due to its strong institutionalisation and accreditation. But it is also externalised and difficult to control for the evaluating organisation (e.g., the Commission). Thus, it is not in the interest of the organisation to give up control to external auditors who do not follow the rules of the evaluating organisation (e.g., the terms of reference). On the other hand, considerable resources are directed into evaluation management and outsourcing, and EU expenditure programmes are already audited on multiple levels. Also, audit is by its nature better equipped to investigate spending programmes.

Finally, since external audits (from the European Court of Auditors) are vested with more credibility, they are also more useful to policy-makers in the EP and

Council. For overall considerations of use, in relation to expenditure programmes performance audits might be a better investment than evaluation systems.

6.3.3 POLICY EVALUATION OVER PROGRAMME EVALUATION

Evaluation systems are typically set up to secure accountability in relation to expenditure programmes. However, as I argued before, expenditure programmes are likely to be assessed better by auditors than evaluators. And as this thesis concludes, the use of evaluations of expenditure programmes is very limited in relation to significant programme change.

Instead, evaluation is more likely to be much better suited for policy evaluation. Policies (understood as legal measures) rarely include a spending operation, where auditors should be involved, and policies do not work in cycles as do programmes. Many of the reasons for non-use of evaluations are linked to the close ties between the evaluation system and the programme cycle. Therefore, it is likely that policy evaluations will generate a much larger potential for evaluation use as it slowly becomes a reality as systematic praxis within the EU evaluation system.

- Albæk E. (1995) Between Knowledge and Power: Utilization of Social Science in Public Policy Making. *Policy Sciences* 28: 79-100.
- Alkin M and Christie CA. (2004) An evaluation theory tree revisited. In: Alkin M (ed) *Evaluation roots*. Thousand Oaks: Sage.
- Alkin M and Taut S. (2003) Unbundling Evaluation Use. *Studies in Educational Evaluation* 29: 1-12.
- Alkin MC and Stecher B. (1983) Evaluation in Context: Information Use in Elementary School Decision Making. *Studies in Educational Evaluation* 9: 23-32.
- Auditors ECo. (2005) Presidential Letter.
- Balthasar A and Rieder S. (2000) Learning from Evaluations: Effects of the Evaluation of the Swiss Energy 2000 Programme. *Evaluation* 6: 245-260.
- Barnes M, Matka E and Sullivan H. (2003) Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems. *Evaluation* 9: 265-284.
- Batterbury SCE. (2006) Principles and purposes of European Union policy evaluation. *Regional Studies* 40: 179-188.
- Bauer M. (2008) Introduction: Organizational change, management reform and EU policy-making. *Journal of European Public Policy* 15.
- Bazeley P. (2013) *Qualitative Data Analysis with NVIVO*, London: Sage Publications Ltd.
- Benjamin L. (2012) Nonprofit Organizations and Outcome Measurement: From Tracking Program Activities to Focusing on Frontline Work. *American Journal of Evaluation* 33.
- Bennett CJ and Howlett M. (1992) The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences* 25: 275-294.
- Bernard HR. (2013) *Social Research Methods*, Thousand Oaks: Sage.
- Bernard HR and Ryan GW. (2010) *Analyzing Qualitative Data: Systemic Approaches*, Thousand Oaks: Sage.
- Beyer jM and Trice HM. (1982) The unilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly* 27: 591-622.

- Bienias S and Iwona L. (2009) Evaluation Systems in the Visegrad Member States. Warsaw: Polish Ministry of Regional Development.
- Borrás S. (2011) Policy learning and organizational capacities in innovation policies. *Science and Public Policy* 38: 725-734.
- Borrás S and Højlund S. (2014) Evaluation and Policy Learning - The Learners' Perspective. *European Journal of Political Research* 53.
- Borum F and Hansen HF. (2000) The Local Construction and Enactment of Standards for Research Evaluation - The Case of the Copenhagen Business School. *Evaluation* 6: 281-299.
- 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.
- Boven M. (2005) Public Accountability. In: Ferlie E, Lynn LE and Pollitt C (eds) *The Oxford Handbook of Public Management*. Oxford: Oxford University Press.
- Bryman A. (2012) *Social Research Methods*, Oxford: Oxford University Press.
- Burns C. (2012) How and When Did We Get Here? An Historical Institutional Analysis of EU Biotechnology Policy. *Journal of European Integration* 34: 341-357.
- Burry j, Alkin M and Ruskus J. (1985) Organizing evaluations for use as a management tool. *Studies in Educational Evaluation* 11: 131-157.
- Böhling K. (2013) Sidelined Member States: Commission learning from Experts in the Face of Comitology. *Journal of European Integration*.
- Cassell C and Symon G. (1994) *Qualitative Methods in Organizational Research*, London: Sage.
- Christiansen T. (1997) Tensions of European governance: politicized bureaucracy and multiple accountability in the European Commission. *Journal of European Public Policy* 4: 73-90.
- Christie CA. (2007) Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation* 28: 8-25.
- Cohen MD, March JG and Olsen J. (1972) A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly* 17.

- Commission E. (1996) Concrete steps towards best practice across the Commission. In: Secretariat-General (ed). Brussels.
- Commission E. (1999) Spending more wisely: Implementation of the Commission's evaluation policy. In: Secretariat-General (ed). Brussels.
- Commission E. (2000a) Focus on results: Strengthening evaluation of Commission activities. In: Commission E (ed) *SEC(2000)1051*. Brussels.
- Commission E. (2000b) Reforming the Commission: A White Paper. Brussels.
- Commission E. (2004) Evaluating EU Activities - A practical guide for the Commission services. Brussels: European Commission.
- Commission E. (2007a) Responding to Strategic Needs: Reinforcing the use of evaluation. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2007b) Revision of the Internal Control Standards and Underlying Framework - Strengthening Control Effectiveness. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2010) Smart Regulation in the European Union. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2013) Public Consultation on Commission Guidelines for Evaluation. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E and Jacobsen S. (2007) Evalueringsaktiviteter & -resultater på tværs af Europa-Kommissionens tjenestegrene. *Dansk Evalueringselskabs konference*. Kolding, Denmark.
- Cook TD and Pollard WE. (1977) Guidelines: how to recognize and avoid some common problems of mis-utilization of evaluation research findings. *Evaluation* 4: 161-164.
- Council E. (1995) Financial Regulation Amendment. In: Union TCotE (ed).
- Cousins B. (2004) Commentary: Minimizing Evaluation Misuse as Principled Practice. *American Journal of Evaluation* 25: 391-397.
- Cousins JB and Earl LM. (1995a) *Participatory evaluation in education: Studies in evaluation use and organizational learning*, London: Falmer Press.
- Cousins JB and Earl LM. (1995b) Participatory evaluation: Enhancing evaluation use and organizational learning capacity. *The Evaluation Exchange* 1: 2-3.
- Cousins JB, Goh SC, Clark S, et al. (2004) Integrating evaluative inquiry into the organisational culture: A review and synthesis of the knowledge base. *The*

- Canadian Journal of Program Evaluation* 19: 99–141.
- Cousins JB and Leithwood KA. (1986) Current empirical research on evaluation utilization. *Review of Educational Research* 56: 331-364.
- Cronbach LJ and Suppes P. (1969) *Research for tomorrow's schools: Disciplined inquiry for education*, New York: Macmillan.
- Dahler-Larsen P. (2006) Organizing Knowledge: Evidence and the Construction of Evaluative Information Systems. In: Rist R and Stame N (eds) *From Studies to Streams*. New Brunswick: Transaction Publishers.
- Dahler-Larsen P. (2012) *The Evaluation Society*, Stanford, Ca.: Stanford University Press.
- Dahler-Larsen P and Krogstrup HK. (1998) *Den Rituelle Refleksion – om evaluering i organisationer*, Odense: Odense Universitetsforlag.
- Dahler-Larsen P and Krogstrup HK. (2006) Evaluering og institutionelle standarder: Nyinstitutionelle betragtninger af evaluering som vor tids ritual. *Tidsskrift.dk*: 283-299.
- Deutsch KW. (1966) *The Nerves of Government*, New York: The Free Press.
- Dey I. (1993) *Qualitative Data Analysis: A User-Friendly Guide for Social Scientists*, London: Routledge.
- DiMaggio PJ and Powell W. (1991) Introduction. In: Powell WW and DiMaggio PJ (eds) *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press, 1-38.
- Dobbin F, Simmons B and Garrett G. (2007) The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning. *Annual Review of Sociology* 33: 449-472.
- Dolowitz D and Marsh D. (1996) Who learns what from whom: A review of the policy transfer literature. *Political Studies* 44: 343-357.
- Dubnick M. (2005) Accountability and the promise of performance - In search of the Mechanisms. *Public Performance & Management Review* 28: 376-417.
- Dunlop CA and Radaelli CM. (2013) Systematising Policy Learning: From Monolith to Dimensions. *Political Studies* 61: 599-619.
- Easterby-Smith M, Crossan M and Nicolini D. (2000) Organizational Learning: Debates Past, Present and Future. *Journal of Management Studies* 37: 783-796.

- Easton D. (1965) *A Framework for Political Analysis*, Englewood Cliffs: Prentice-Hall.
- Eckerd A and Moulton S. (2011) Heterogeneous Roles and Heterogeneous Practices: Understanding the Adoption and Uses of Nonprofit Performance Evaluations. *American Journal of Evaluation* 32: 98-117.
- Ecorys and COWI. (2008) Study on the state of implementation of Activity Based Management in the European Commission. Rotterdam.
- Egan M. (2009) Governance and learning in the post-Maastricht era? *Journal of European Public Policy* 16: 1244-1253.
- Ellinas A and Suleiman E. (2008) Reforming the Commission: between modernization and bureaucratization. *Journal of European Public Policy* 15: 708-725.
- Eser TW and 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.
- Eureval-C3E. (2006) Study on the Use of Cost-effectiveness Analysis in EC's Evaluations. Lyon: Eureval-C3E.
- Ferry M and Olejniczak K. (2008) The use of evaluation in the management of EU programmes in Poland. Warsaw: Ernst & Young.
- Fetterman DM. (1994) Steps of empowerment evaluation: From California to Cape Town. *Evaluation and Program Planning* 17: 305_313.
- Finne H, Levine M and Nilssen T. (1995) Trailing research: A model for useful program evaluation. *Evaluation* 1: 11-31.
- Floden RE and Weiner SS. (1978) Rationality to Ritual: The Multiple Roles of Evaluation in Governmental Processes. *Policy Sciences* 9: 9-18.
- Flyvbjerg B. (2006) Five Misunderstandings About Case-Study Research. *Qualitative Inquiry* 12: 219-245.
- Francesco FD, Radaelli CM and Troeger VE. (2011) Implementing regulatory innovations in Europe: the case of impact assessment. *Journal of European Public Policy* 19: 491-511.
- Freeman R. (2006) Learning in public policy. In: Moran M, Rein M and Goodin RE (eds) *Oxford Handbook of Public Policy*. Oxford: Oxford University Press, 367-389.

- Furubo J-E. (2006) Why Evaluations Sometimes Can't be Used - and Why They Shouldn't. In: Rist R and Stame N (eds) *From Studies to Streams*. New Jersey: Transaction Publishers, 147-165.
- Furubo J-E, Rist RC and Sandahl R. (2002) *International Atlas of Evaluation*, London: Transaction Publishers.
- Gornitzka Å and Sverdrup U. (2011) Access of experts: information and EU decision-making. *West European Politics* 34: 1452-1474.
- Grin J and Loeber A. (2007) Theories of policy learning: agency, structure, and change. In: Fischer F, Miller GJ and Sidney MS (eds) *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. Boca Raton: CRC Press, 201-219.
- Grøn C. (2009) Same Procedure as Last Year? An analysis of constellations of trust and control in management in the European Commission. *Department of Political Science*. Copenhagen: University of Copenhagen.
- Hanberger A and Schild I. (2004) Strategies to Evaluate a University–Industry Knowledge-exchange Programme. *Evaluation* 10: 475-492.
- Hansen HF and Borum F. (1999) The Construction and Standardization of Evaluation. The Case of the Danish University Sector. *Evaluation* 5: 303-329.
- Hansen M, Alkin MC and Wallace TL. (2012) Depicting the logic of three evaluation theories. *Evaluation and Program Planning*.
- Harlow C. (2002) *Accountability in the European Union*, Oxford: Oxford University Press.
- Hartley J. (2004) Case studies in organizational research. In: Cassell C and Symon G (eds) *Qualitative methods in organizational research, a practical guide*. London: Sage, 208-229.
- Hecló H. (1974) *Modern Social Politics in Britain and Sweden*, New Haven: Yale University Press.
- Henry GT. (2004) Why not use? In: Caracelli VJ and Preskill H (eds) *The expanding scope of evaluation use. New Directions for Evaluation*. San Francisco, CA: Jossey-Bass, 85–98.
- Henry GT and Mark MM. (2003) Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation* 24: 293.

- Hoerner J and Stephenson P. (2012) Theoretical Perspectives on Approaches to Policy Evaluation in the EU: The Case of Cohesion Policy. *Public Administration* 90: 699-715.
- Hofstetter C and Alkin M. (2003) Evaluation use revisited. In: Kellaghan T, Stufflebeam D and Wingate L (eds) *International Handbook of Educational Evaluation*. London: Kluwer Academic.
- Hood C. (1991) A public management for all seasons? *Public Administration* 69.
- Højlund S. (2014a) Evaluation use in evaluation systems - the case of the European Commission. *Evaluation* 20: 428-446.
- Højlund S. (2014b) Evaluation use in the organisational context – Changing focus to improve theory. *Evaluation* 20: 26-43.
- Højlund S. (Forthcoming) Between Accountability and Learning - The Case of the European Commission's Evaluation System.
- Imam I, LaGoy A and Williams B. (2007) Introduction. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes: EdgePress of Inverness.
- Jacob S. (2005) *Institutionnaliser l'évaluation des politiques publiques. Étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*, Bruxelles: PIE-Peter Lang.
- James O and Lodge M. (2003) The limitations of "policy transfer" and "lesson drawing" for public research. *Political Studies Review* 1: 179-193.
- Johnson K, Greenseid LO, Toal SA, et al. (2009a) Research on Evaluation Use – A review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation* 30: 377-410.
- Johnson K, Greenseid LO, Toal SA, et al. (2009b) Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation* 30: 377-410.
- Kassim H. (2008) 'Mission impossible', but mission accomplished: the Kinnock reforms and the European Commission. *Journal of European Public Policy* 15: 648-668.
- King JA and Thompson B. (1983) Research on school use of program evaluation: A literature review and research agenda. *Studies in Educational Evaluation* 9: 5-21.

- Kirkhart KE. (2000) Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation* 88: 5-23.
- Kohlbacher F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research* 7.
- Laffan B. (1997a) *The Finances of the European Union*, Basingstoke: Macmillan.
- Laffan B. (1997b) From policy entrepreneur to policy manager: the challenge facing the European Commission. *Journal of European Public Policy* 4: 422-438.
- Laffan B. (2003) Auditing and accountability in the European Union. *Journal of European Public Policy* 10: 762-777.
- Langley A. (1998) In Search of Rationality - The Purpose Behind the Use of Formal Analysis in Organizations. In: Maanen JV (ed) *Qualitative Studies of Organizations*. Thousand Oaks: SAGE Publications, 51-90.
- Lazarsfeld PF, Sewell WH and Wilensky HL. (1967) *The uses of Sociology*, New York: Basic Books.
- Leeuw FL and Furubo J-E. (2008) Evaluation Systems : What Are They and Why Study Them? *Evaluation* 14: 157-169.
- Legard R, Keegan J and Ward K. (2003) In-depth Interviews. In: Ritchie J and Lewis J (eds) *Qualitative Research Practice*. London: Sage Publications.
- Levin B. (1987) The uses of research: A case study in research and policy. *The Canadian Journal of Program Evaluation* 2: 43-55.
- Leviton LC. (2003) Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation* 24: 525-535.
- Leviton LC and Hughes EFX. (1981) Research on the utilization of evaluations: A review and synthesis. *Evaluation Review* 5: 525-548.
- Laat Bd. (2005) Study on the Use of Evaluation Results in the Commission. In: Laat Bd (ed). Paris: Technopolis.
- March J. (2010) *The ambiguities of experience* Ithaca, N.Y.: Cornell University Press.
- Mark MM and Henry GT. (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation* 10: 35-57.
- Mayring P. (2000) Qualitative Content Analysis. *Forum: Qualitative Social Research* 1.

- Mayring P. (2002) Qualitative content analysis - research instrument or mode of interpretation. The Role of the REsearcher in Qualitative Psychology. In: Kiegelmann M (ed) *Nexus Qualitative Psychology*. Tübingen: Verlag Ingeborg Huber, 139-148.
- Mayring P. (2003) *Qualitative Inhaltsanalyse, Grundlage und Techniken*, Weinheim: Beltz, UTB.
- Mayring P. (2004) Qualitative Content Analysis. In: Flick U, Von Kardorff E and Steinke I (eds) *A Companion to Qualitative Research*. London: SAGE, 266-269.
- Mendez C and Bachtler J. (2011) Administrative reform and unintended consequences: an assessment of the EU Cohesion policy 'audit explosion'. *Journal of European Public Policy* 18: 746-765.
- Meyer JW and Rowan B. (1977) Institutionalized organizations: Formal structure as Myth and Ceremony. *American Sociological Review* 83: 340-363.
- Meyer JW and Rowan B. (1991) Institutionalized Organizations: Formal Structure as Myth and Ceremony. In: Powell WW and DiMaggio PJ (eds) *The New Institutionalism in Organizational Analysis*. Chicago: The University Chicago Press, 41-62.
- Miles MB and Huberman AM. (1994) *Qualitative Data Analysis*, Thousand Oaks, CA: Sage Publications.
- Mocker DW and Spear GE. (1982) *Lifelong Learning: Formal, Nonformal, Informal, Informal, and Self-Directed.*, Columbus OH: Eric.
- OECD. (2002) Regulatory policies in OECD Countires: From Interventionism to Regulatory Governance. Paris: OECD.
- Olejniczak K. (2013) Mechanisms Shaping an Evaluation System - A Case Study of Poland 1999-2010. *Europe-Asia Studies* 65: 1642-1666.
- Owen JM. (2002) Linking evaluation use to the research utilisation literature. *Annual Meeting of the American Evaluation Association*. Washington DC.
- Owen JM and Lambert Fc. (1995) Roles for Evaluation in Learning Organizations. *Evaluation* 1: 237-250.
- Owen JM and Rogers P. (1999) *Program evaluation: Forms and approaches*, London: Sage.
- Owens S, Rayner T and Bina O. (2004) New agendas for appraisal: reflections on

- theory, practice, and research. *Environnement and Planning* 36: 1943-1959.
- Patton MQ. (1997) *Utilization-focused evaluation: The new century text*, Thousand Oaks: Sage.
- Pawson R and Tilley N. (1997) *Realistic Evaluation*, London: SAGE Publication Ltd.
- Pelz DC. (1978) Some expanded perspectives on use of social science in public policy. In: Yinger JM and Cutler SJ (eds) *Major Social Issues: A Multidisciplinary View*. New York: Macmillan, 346-357.
- Pollitt C and Bouckaert G. (2004) *Public Management Reform: A Comparative Analysis*, Oxford University Press.
- Pollitt C, Girre X, Lonsdale J, et al. (1999) *Performance Audit and Public Management in Five Countries* Oxford and New York: Oxford University Press.
- Porter TM. (1995) *Trust in Numbers: the pursuit of objectivity in science and public life*, UK: Princeton University Press.
- Power M. (1997) *The Audit Society - Rituals of Verification*, Oxford: Oxford University Press.
- Power M. (2005) The Theory of Audit Explosion. In: Ferlie E, Lynn LE and Pollitt C (eds) *Oxford Handbook of Public Management*. Oxford: Oxford University Press.
- Preskill H. (2008) Evaluation's Second Act: A Spotlight on Learning. *American Journal of Evaluation* 29: 127.
- Preskill H and Torres RT. (1999) Building Capacity for Organizational Learning Through Evaluative Inquiry. *Evaluation* 5: 42-60.
- Preskill H, Zuckerman B and Matthews B. (2003) An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation* 24.
- Pröpper IMAM. (1987) Beleidsevaluatie als argumentatie. *Beleidswetenschap* 2.
- Radaelli CM. (2009) Measuring policy learning: regulatory impact assessment in Europe. *Journal of European Public Policy* 16: 1145-1164.
- Radaelli CM and Dunlop CA. (2013) Learning in the European Union: theoretical lenses and meta-theory. *Journal of European Public Policy* 20: 923-940.
- Radaelli CM and Meuwese ACM. (2010) Hard Questions, Hard Solutions:

- Proceduralisation through Impact Assessment in the EU. *West European Politics* 20: 923-940.
- Remenyi D, Money A, Price D, et al. (2002) The creation of knowledge through case study research. *Irish Journal of Management* 23: 1-17.
- Rich RF. (1977) Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding. In: Weiss CH (ed) *using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Rist R and Stame N. (2006) *From Studies to Streams: Managing Evaluative Systems*, London: Transaction Publishers.
- Ritchie J, Lewis J and Elam G. (2003) Designing and Selecting Samples. In: Ritchie J and Lewis J (eds) *Qualitative Research Practice - A Guide for Social Science Students and Researchers*. London: SAGE.
- Ryan GW and Bernard HR. (2000) Data management and analysis methods. In: Denzin N and Lincoln YS (eds) *Handbook for qualitative research*. Thousand Oaks: Sage.
- Sager F and 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.
- Schofield J. (2004) A model of learned implementation. *Public Administration* 82: 283-308.
- Schout A. (2009) Organizational learning in the EU's multilevel-governance system. *Journal of European Public Policy* 16: 1124-1144.
- Schreier M. (2012) *Qualitative Content Analysis in Practice*, Thousand Oaks: SAGE Publications Ltd.
- Schwandt TA. (1997) Evaluation as Practical Hermeneutics. *Evaluation* 3: 69-83.
- Schwandt TA. (2009) Globalizing influences on the Western evaluation imaginary. In: Ryan KE and Cousins JB (eds) *Sage international handbook on educational evaluation*. Los Angeles: Sage, 119-138.
- Schön-Quinlivan E. (2008) Implementing organizational change - the case of the Kinnock reforms. *Journal of European Public Policy* 15: 726-742.

- Scott RW. (1995) *Institutions and organizations*, Thousand Oaks: SAGE.
- Scott RW and Meyer JW. (1994) *Institutional environments and organizations: Structural complexity and individualism*, Thousand Oaks: SAGE Publications.
- Scriven M. (1967) The methodology of evaluation. In: Stake RE (ed) *Curriculum evaluation*. Chicago: Rand McNully.
- Scriven M. (1991) Beyond Formative and Summative Evaluation. In: McLaughlin M and Philips DC (eds) *Evaluation and Education: At Quarter Century*. 3 ed. Chicago: University of Chicago Press.
- Selznik P. (1949) *TVA and the Grass Roots*, Berkeley: University of California Press.
- Selznik P. (1984) *Leadership in administration: A sociological interpretation*, Berkeley: University of California Press.
- Shulha LM and Cousins B. (1997) Evaluation Use: Theory, Research and Practice Since 1986. *Evaluation Practice* 18: 195-208.
- Simon H. (1957) *Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization*, New York: Macmillan.
- Smith AS. (2013) How the European Commission's Policies Are Made: Problematization, Instrumentation and Legitimation. *Journal of European Integration*.
- Spence D. (2000) Plus ca change, plus c'est la meme chose? Attempting to reform the European Commission. *Journal of European Public Policy* 7: 1-25.
- Stake RE. (2000) Case studies. In: Denzin N and Lincoln YS (eds) *Handbook of qualitative research*. Thousand Oaks: Sage, 435-453.
- Stake RE. (2006) *Multiple Case Study Analysis*, Thousand Oaks, CA: Sage Publications.
- Stern E. (2006) Contextual challenges for evaluation practice. In: Shaw I, Green J and M. M (eds) *The SAGE Handbook of Evaluation*. London: SAGE.
- Stern E. (2009) *Evaluation policy in the European Union and its institutions*.
- Stewart DW and Kamins MA. (1993) *Secondary research: information sources and methods*, Thousand Oaks, CA: Sage Publications.
- Stufflebeam D. (2001) Evaluation Models. *New Directions for Evaluation* 89: 7-98.

- Summa H and Toulemonde J. (2002) Evaluation in the European Union: addressing complexity and ambiguity. In: Furubo J-E, Sandahl R and Rist R (eds) *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.
- Thompson B and King JA. (1981) Evaluation utilization: A literature review and research agenda. *Annual Meeting of the American Educational Research Association*. Los Angeles, CA.
- Titscher S, Meyer M, Wodak R, et al. (2000) *Methods of Text and Discourse Analysis*, Thousand Oaks: Sage.
- Torres RT and Preskill H. (2001) Evaluation and Organizational Learning: Past, Present, and Future. *American Journal of Evaluation* 22: 387-395.
- Toulemonde J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *Vierteljahrshefte zur Wirtschaftsforschung* 69: 350-357.
- Toulemonde J, Summa H and Usher N. (2005) Assessing EU Evaluations. In: Schwartz R and Mayne J (eds) *Quality matters: seeking confidence in evaluation, auditing and performance reporting*. New Brunswick: Transaction Publishers.
- Van der Knaap P. (1995) Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation? *Evaluation* 1: 189-216.
- van der Meer F-B and Edelenbos J. (2006) Evaluation in Multi-Actor Policy Process. Accountability, Learning and Co-operation. *Evaluation* 12: 201-218.
- Varone F, Jacob S and Winter LD. (2005) Polity, Politics and Policy Evaluation in Belgium. *Evaluation* 11: 253.
- Vedung E. (1997) *Public Policy and Program Evaluation*, New Brunswick: NJ: Transaction Publishers.
- Vedung E. (2010) Four Waves of Evaluation Diffusion. *Evaluation* 16: 263-277.
- Volden C, Ting MM and Carpenter DP. (2008) A Formal Model of Learning and Policy Diffusion. *American Political Science Review* 102: 319-332.
- Weiss CH. (1978) Improving the linkage between social research and public policy. In: Lynn LE (ed) *Knowledge and Policy: The Uncertain Connection*. Washington DC: National Research Council.

- Weiss CH. (1979) The many meanings of research utilization. *Public Administration* 39: 426-431.
- Weiss CH. (1992) *Organizations for Policy Analysis - Helping Government Think*, Newbury Park, CA: SAGE.
- Weiss CH. (1998) Have We Learned Anything New About the Use of Evaluation. *American Journal of Evaluation* 19: 21-33.
- Weiss CH and Bucuvalas M. (1980) *Social Science Research and Decision Making*, New York: Columbia University Press.
- Wille A. (2010) Political-Bureaucratic Accountability in the EU Commission: Modernising the Executive. *West European Politics* 33: 1093-1116.
- Williams B and Imam I. (2007) *Systems concepts in evaluation - An Expert Anthology*, Point Reyes: EdgePress.
- Williams K, Laat Bd and Stern E. (2002) *The Use of Evaluation in the Commission Services*. Paris: Technopolis.
- Wonka A. (2008) Decision-making dynamics in the European Commission: partisan, national or sectoral? *Journal of European Public Policy* 15: 1145-1163.
- Yin RK. (2003) *Case study research: Design and methods*, Thousand Oaks, CA: Sage.
- Zald MN and Denton P. (1963) From Evangelism to General Services: the Transformation of the YMCA. *Administrative Science Quarterly* 8: 214-234.
- Zito A and Schout A. (2009) Learning theory reconsidered: EU integration theories and learning. *Journal of European Public Policy* 16: 1103-1123.

8.1 ARTICLE 1 – EVALUATION USE IN THE ORGANIZATIONAL CONTEXT – CHANGING FOCUS TO IMPROVE THEORY

Reference:

Højlund, S. (2014) *Evaluation use in the organizational context – changing focus to improve theory*. Evaluation 20(1), 26-43.

<http://evi.sagepub.com/content/20/1/26.abstract>

*(To avoid any copyright issues in relation to the final print of this thesis, this article is conveyed in its **‘pre-script’ version**. A pre-script version is a near-published version of the article as it was after peer-review and language check by the author as well as after the final acceptance of the article by the journal editor. The pre-script version does not include the final editing and formatting of the article by the journal in which the article was published. The author of this thesis recommends readers of this thesis published by the Library of Copenhagen Business School to seek and purchase this article in its original version at the journal in which it was published.)*

Evaluation Use in the Organisational Context - Changing Focus to Improve Theory

This article is about evaluation use. It focuses on the well-known paradox that evaluation is undertaken to improve policy, but in fact rarely does so. Additionally, the article also finds that justificatory uses of evaluation do not fit with evaluation's objective of policy-improvement and social betterment. The article explains why the paradox exists and suggests applying organisational institutional theory to explain evaluation use. The key argument is that - in order to explain all types of evaluation uses including non-use and justificatory uses - focus needs to be on the evaluating organisation and its conditioning factors, rather than the evaluation itself.

Keywords: neo-institutionalism, justificatory use, non-use, rationality, organisational theory

Introduction

Evaluation use is probably the most researched theme in the literature on evaluation (Christie 2007, 8; Henry and Mark 2003, 294). That is because administrators, politicians and citizens have an interest in knowing to what extent evaluations are used to improve policies. Despite this great interest in evaluation use, it is well described that evaluations rarely do change policies (Patton, 1997; Pawson and Tilley, 1997: 2–3). This constitutes a paradox, since the very objective of evaluation is to improve policy. And so, one could ask; why evaluate with the objective of improving policy, if the evaluation is not used after all? This article addresses this paradox in new ways.

First, the article argues that the paradox described above is inherent in the nature of evaluation and that this has ramifications for our understanding of types of evaluation uses such as legitimising- and symbolic use as well as non-use.

Second, the article argues that the literature on evaluation use insufficiently addresses evaluation's organisational context and therefore is unable to resolve the paradox.

Finally, to remedy this gap in the literature concerning the lack of contextual explanations of evaluation use, the article argues along the lines of Sanderson (2000) and Dahler-Larsen (2012) that it would be beneficial to complement core assumptions of causality and rationality with a theory of organisation. A model is proposed that; 1) explains evaluation use by factors that are contextual to the evaluating organisation rather than contextual to the evaluation; and 2) integrates non-use (as well as justificatory use types) to resolve the paradox of non-use explained above.

Clarifying evaluation use

The literature on evaluation use is strongly linked with a much larger stream of literature on the use of scientific results and knowledge use in policy-making (Porter 1995; Vedung, 2010; Weiss, 1998; Weiss and Bucuvalas, 1980; Lazarsfeld and Sewell et al., 1967). The literature on evaluation use has developed primarily in the US (Saunders 2012) and is often associated with the work of Michael Quinn Patton and Carol Weiss. A definition of evaluation use is provided by Johnson et

al. (2009: 378); any application of evaluation processes, products, or findings to produce an effect.'

The literature on evaluation use converged early on four main types of evaluation use: Instrumental-, conceptual-, process- and symbolic use. However, over the years, many researchers grew dissatisfied with these categories. Alkin and Taut were displeased with the multitudes of potential evaluation uses and influences that did not fit into these four vaguely defined categories of use (Alkin and Taut, 2003; Kirkhart, 2000; Henry and Mark, 2003; Weiss, 1998). Some of the criticism concerned the operationalisation and measurement of the concepts: used when and used by whom? The appropriate typologies of use (or influence) are still debated most likely because it continues to be difficult to assess the influences and effects of evaluations (Alkin and Taut, 2003; Kirkhart, 2000; Mark and Henry, 2004; Johnson et al., 2009; Hofstetter and Alkin, 2003).

In this article, a simple typology of evaluation use is applied. Evaluation use is either referred to as findings use (instrumental, conceptual, legitimising) or process use (instrumental, conceptual, symbolic) in line with Alkin and Taut's (2003) refinement of the old categories. They recognised that process use (use during the evaluation process) could be instrumental, conceptual and legitimising (e.g. evaluation is legitimising the organisation). Moreover, the evaluation result (typically an evaluation report), could also be used instrumentally, conceptually and symbolically (e.g. evaluation legitimises the evaluated programme). In this article, legitimising- and symbolic use are referred to as 'justificatory' uses.

There are good reasons for applying newer frameworks (e.g. Kirkhart, 2000; Mark and Henry, 2004), but the previously explained concepts serve adequately the purpose of reference in this article that is mainly concerned with the factors explaining evaluation use and the legitimising- and symbolic use types.

Understanding the paradox

This section elaborates on the nature of evaluation, evaluation use and the paradox associated with it. The problem for the literature is defined and a possible solution to this problem is found in organisational institutional theory.

The nature of evaluation

If one can talk about a 'general logic' of evaluation, this logic is inherently realist and rational (Albaek, 1995; Van der Knaap, 1995; Schwandt, 1997; Sanderson, 2000). The often-stated purpose and idea behind evaluation is rooted in assumptions of rationality and causality. Evaluation is commonly understood as a tool providing a feedback mechanism informing policy-makers and civil servants of what works and what does not. Most evaluation models and assumptions of evaluation use are derived from this primary rationale and function along logics of cause and effect (see e.g. Mark and Henry, 2004: 38; Cousins and Leithwood, 1986; Pawson and Tilley, 1997). This is ontologically and epistemologically linked to realist and positivist understandings commonly found in classical economic thinking and rational choice, where the actor is perceived to be rational and pursuit goal-attainment and utility-maximising behavioural patterns of action (Sanderson, 2000). These assumptions contain an underlying positive and evolutionary assumption of progress and betterment, whether of the bureaucracy or the policy. In evaluation, this positive assumption also dominates through the general conception of 'social betterment' as the ultimate objective of evaluation (Henry and Mark, 2004; Dahler-Larsen, 2012).

According to Albaek (1995), the positivist assumptions behind evaluation has been weakened somewhat over the last decades, as positivism was challenged by phenomenological and hermeneutic traditions as well as critical theory. However, Sanderson (2000) questions whether evaluation has completely escaped positivist logics of cause and effect as they are still dominating evaluation models even if methodological plurality is accepted widely.

The inborn paradox

Evaluation itself was born in a time of believe in a better world through rational interventions and social engineering (Vedung, 2010) and therefore evaluation is inherently rationalist, causal and evolutionary in nature (Sanderson, 2000). This is apparent considering the *raison d'être* of evaluation that is included in mainstream definitions of evaluation: Evaluation is commonly defined as a 'systematic inquiry leading to judgements about program (or organisation) merit, worth, and significance, and support for program (or organisational) decision making' (Cousins et al., 2004: 105; see also Vedung, 1997). It is clear from the above that the use of an evaluation is an intrinsic part of the evaluation itself. However, this

also constitutes a logical paradox since an un-used evaluation would then seize to be an evaluation according to the definition.

The example above serves well to illustrate the point. However, in all fairness, all definitions of evaluation do not include a stated purpose of evaluation. Nevertheless, the paradox remains insofar as the stated purpose of evaluation is to improve political interventions, develop knowledge or assess merit and worth, which is almost always the case (see e.g. Henry et al., 2000).

Justificatory uses 'tacked on'

The fundamental assumption of linear cause and effect is intrinsically linked with the instrumental use category: Ideally, evaluation improves policy through the instrumental appliance of an evaluation's results (conclusions and recommendations) to improve policy. However, evaluation does not deliver on its promise to inform policy-makers (Patton, 1997; Pawson and Tilley, 1997: 2–3). This was realised quite early in the evaluation literature (Shulha and Cousins, 1997) and more sophisticated categories of evaluation use were devised so that evaluation use was not only instrumental but also related to conceptual enlightenment, process, symbolic use and legitimising types of uses (Patton, 1997; Pelz, 1978; Alkin and Taut, 2003).

Nevertheless, there is a logical problem between the general assumption of linearity towards social betterment and the inclusion of symbolic and legitimising uses (as well as non-use) in the causal chain. The two types of use were recognised by Pelz (1978) quite early, but nevertheless they have not been appropriately integrated into the causal logic prevailing in models of evaluation and evaluation use. Kirkhart notes: "Other types of use [than 'results based use'] are secondary, 'tacked on' or seen as important primarily in the service of results based use." (Kirkhart, 2000: 6). There is no doubt that justificatory uses exist and should be kept as separate categories of use apart from instrumental uses, but it is very hard to argue that symbolic and legitimising use of evaluation lead to social betterment; e.g. water quality is not improved because an aid organisation use its evaluation process to legitimise itself. The argument of rationality would need to be stretched very far to argue that symbolic and legitimising evaluation use can actually improve living conditions for human beings (i.e. social betterment). So

contrary to the main rational assumption of the purpose of evaluation, evaluating organisations or actors are likely to have a very different objective with evaluation. It seems that the evaluation literature empirically have recognised justificatory uses, but theoretically have disregarded this acknowledgement's implications for the overall rationalist and causal foundation of evaluation thinking.

Paradox – and so what?

The paradox explained above is well described by the literature on evaluation use. It remains to be seen how justificatory use types can be properly integrated into a model of evaluation use without just being 'tacked on' as Kirkhart notes (Ibid.). It seems unsatisfactory to empirically acknowledge justificatory uses of evaluation and widespread non-use of evaluations – and to call it a 'utilization crisis' (Patton, 1997) – while not having a good explanation for the phenomena.

The problem might be that the evaluation use literature often focuses on the evaluation itself; its implementation, its outputs, its conditioning factors etc. and largely ignores the organisational context. However, on the organisational level, there might be forces at play that are so strong that they overrule a persuasive evaluation result and cancels instrumental use. On this note, Christina Boswell states that:

'Despite a huge expansion of literature on knowledge utilization from the mid-1970s onwards, surprisingly little attention has been paid to the symbolic functions of knowledge. (...) Moreover, contributions in organizational sociology have shown how organizations derive legitimacy through signalling their commitment to knowledge utilization (...). However, there has been little attempt to develop a theory setting out the conditions under which symbolic knowledge utilization may be expected to occur, or testing these claims through empirical enquiry.' (Boswell, 2008: 471-472).

In this article, the claim is that organisational factors are more important, in terms of explaining evaluation uses and non-uses, than the literature so far has acknowledged. It is argued that an organisation's concerns of its external legitimacy are likely to have priority over evaluation findings and thus in certain cases lead to non-use of these findings no matter how relevant they might be to the organisation. This argument will be further developed in the following section.

Looking for a solution

Drawing inspiration from Boswell (2008), organisational institutionalism seems like a good place to start in order to explain justificatory uses. Selznick (1949) and later a large number of studies observed that organisations act contrary to the rational utility-maximizing behaviour that is expected *and contrary* to the objectives organisations set themselves. Instead, organisations mostly act according to norms and values in their environment in order to legitimise themselves (Zald and Denton, 1963; Scott and Meyer, 1994). An interesting conclusion from these early studies on institutions was that organisations adapt to the practices that are legitimised by the organisational environment (Meyer and Rowan, 1977). This is particularly the case if the organisation perceives itself to be operating in an unstable environment, where there is uncertainty about the distribution of resources between itself and rival organisations (DiMaggio and Powell, 1991: 30–1). DiMaggio and Powell (1991: 190) argue that organisations are concerned with reducing uncertainty and stabilizing social relations (exemplified in Selznick, 1984) and thus not primarily with maximizing power or maximizing outputs as e.g. a rational-choice perspective would be arguing (see also Scott, 1995: 21; Weick, 1995: 86–7). Hence, symbolic and legitimizing uses of evaluation are in line with the most basic argument of organisational institutionalism that organisations have a need to legitimise themselves in order to survive as organisations (Dahler-Larsen, 2012; Powell and DiMaggio, 1991; Boswell, 2008).

The argument put forward in this section is that the institutional context influences evaluation use and that context has not received the attention I claim it should have with regard to explaining evaluation use. According to Peter Dahler-Larsen (2012: 34), ‘evaluation is strongly dependent on its social and organizational context.’ Parallel to Dahler-Larsen, Boswell (2008: 473) argues that: ‘Any account of how organizations use knowledge will inevitably be premised on a theory of organizations: a set of claims about the sources and nature of organizational interests, and how these translate into organizational action.’ Along the lines of that argumentation, the argument put forward here is that evaluations are almost exclusively embedded in organisational contexts one way or the other. Typically, evaluations are procured by organisations, carried out by organisations (enterprises or teams of individuals), maybe read and acted upon by a third organisation etc.

So far, the majority of literature on evaluation use focuses on identification of uses and influences and factors and conditions related to the evaluation. The three reviews on the evaluation use literature by Leviton and Hughes (1981), Cousins and Leithwood (1986) and Johnson et al. (2009) covering the period from 1971 to 2005 do not find or mention institutional explanations of evaluation use. Cousins and Leithwood's findings were later challenged by Levin (1987), who found contextual factors to be highly important. Shulha and Cousins (1997) later produced another important review concluding that contextual factors and organisational contexts were gaining more prominence in the literature. But the review did not mention institutional theory either (Saunders 2012). Neither did two later reviews of the literature by Cousins et al. (2004) and Amo and Cousins (2007). Not so long ago, Ledermann (2011: 160) argued that '...it is time to abandon the ambition of finding 'the important' characteristic for use and to adopt a focus on context-bound mechanisms of use instead'.

Turning to the evaluation literature for answers rooted in institutional theory yields few results. A bibliographical search in four leading evaluation journals (American Journal of Evaluation; Evaluation; Evaluation Review; Canadian Journal of Programme Evaluation) resulted in 24 articles and two papers referring to 'institutionalism' or 'neo-institutional*'.ⁱ The majority of the articles merely refers to or mentions the institutional context without applying the theory (e.g. Van der Knaap, 1995; Sanderson, 2000; Saunders, 2012; Barnes, 2003). However, Van der Knaap (1995) and Sanderson (2000) are two important articles in terms of articulating the need for contextual analysis of evaluation research. Another part of the articles looks to define evaluation institutions at work in organisations (Barnes et al., 2003; Hanberger and Schild, 2004; Verone and Jacob, 2005; Jacob, 2005). A third part of the articles that turned up in the search used institutional theory as a framework for understanding the evaluation institution's diffusion and adoption (Hansen and Borum, 1999, 2000; Radaelli, 2009). These articles apply a traditional organisational institutional analysis to illustrate how evaluation and impact assessment are diffused in an organisational field. Finally, two articles go further by looking into how institutionalised evaluation practices influence performance and evaluation use (Sager and Rissi, 2011; Eckerd and Moulton, 2011). However, Eckerd and Moulton (2011) are the only authors that apply institutional logics to explain evaluation use. Their valuable contribution is an

important source of inspiration and backing to the model put forward later in this article. So, despite the multitude of explanatory factors and conditions that are analysed in the literature on evaluation use, organisational institutional theory is not evoked systematically as an explanatory framework.

Since institutional theory as a means to explain evaluation is almost absent in the evaluation literature, one can either conclude that it is not significant as explanatory factor of evaluation use, or that it may have been overlooked so far as a theoretical framework for understanding evaluation use. The claim in this article is that the latter option is the case, and therefore the next chapter investigates how institutional theory can be applied to explain evaluation use.

Explaining evaluation use with organisational theory

In this section, the article proposes an explanatory model linking assumptions from organisational institutionalism with new findings from research on evaluation and knowledge use. The model includes the external and internal factors that are at play in an organisation. It is empirically supported by the findings of Boswell (2008) and Eckerd and Moulton (2011). But first, the next section describes the institution of evaluation.

The Institution of Evaluation

At heart of institutional theory is the definition of an institution. First, institutions should not be confused with organisations. Rather they are the frames for action of organisations or simply put; the taken-for-granted formal and informal rules of the game. Scott (1995: 49) writes; ‘...institutions are multifaceted, durable, social structures, made up of symbolic elements, social activities, and material resources.’ From this definition it is important to stress that institutions are resilient to change and that they consist of cultured-cognitive, normative, and regulative elements that provide meaning to social life (Scott, 1995: 48). Thus, formal and informal rules shape social action, constraining actors and limit their choice of action and might be the subject of conflict between competing interests (DiMaggio and Powell, 1991). An institution requires a clear typification (e.g. ‘evaluation’), social knowledge or sedimentation. Though it is difficult to assess the degree or power of an institution, indications of its strength are: 1) the level of

taken-for-grantedness or automation; 2) the severity of sanctions when it is not followed.

The institution of evaluation is taken for granted in many organisations. There are a number of relatively stable interpretations of concepts, practices, principles, norms, ethics, values and artefacts associated with evaluation. In other words, evaluation fits well Scott's definition of an institution previously mentioned. Peter Dahler-Larsen (2006) argues that evaluation has become a protected discourse and that 'Evaluation has stepped into character as a mandatory political and administrative ritual, which priorities and decisions need to adhere to in order to be socially valid' [my translation]. Schwandt (2009) refers to the 'evaluation imaginary'. In other words, evaluation has become a de facto legitimising institution – a practice we in many cases take for granted without questioning.

The strength of an institution can be assessed by the severity of sanctions associated with non-adherence to the institution. E.g. A foundation granting a non-profit organisation means to carry out a particular project might include a provision in the contract instructing the non-profit to evaluate the project ex-post or face the sanction of returning the money. The foundation might not face the same level of sanctions if it fails to e.g. evaluate its organisation. Hence, the evaluation institution is over time supposedly becoming stronger in the non-profit compared to the foundation.

The notion of 'field' is also important in institutional theory. Scott (1995: 207-208) defines it in the following way: '... the notion of field connotes the existence of a community of organizations that partakes of a common meaning system and whose participants interact more frequently and fatefully with one another than with actors outside of the field.' The most important actors in the field of evaluation are arguably the World Bank, OECD, UN, the American Evaluation Association and the European Evaluation Society as well as regional supranational political organisations such as the EU and some national donor agencies etc. (Furubo and Sandhal, 2002; Verone et al., 2005; Stern, 2006). They mutually reinforce and constantly negotiate the evaluation institution through their production of guidelines and practices as well as exchanges of opinions in public

debates and at evaluation seminars, -courses and -conferences etc. The field is an important concept, because this is what an organisation adapts to.

While the institution of evaluation is commonly believed to produce instrumental and rational outcomes (Sanderson, 2000), it is also a result of the legitimacy-seeking behaviour of organisations (Dahler-Larsen, 2006, 2012). Hence, according to institutional theory, the institutionalisation of evaluation depends on the pressure from the field.

External pressure

The core idea of organisational institutionalism is that organisations are fundamentally concerned with securing legitimacy by meeting societal expectations about their structure, practices, rhetoric or outputs (DiMaggio and Powell, 1991). Organisations do this in order to reduce uncertainty and stabilise social relations (DiMaggio and Powell, 1991: 19; Scott, 1995: 21; Weick, 1995: 86–7). DiMaggio and Powell (1983: 147-160) identify three main processes of institutional adaptation (they refer to it as 'isomorphism'); 'coercive' related to dominance and legitimacy in the field, 'mimetic' related to field uncertainty and 'normative' related to professionalization in the field. They define isomorphism (process of adaptation) as a "constraining process that forces an organisation in an organisational field to resemble other organisations facing similar environmental conditions" (DiMaggio and Powell, 1991: 66).

Coercive isomorphism takes place when an organisation is constrained by cultural expectations (e.g. as manifested in laws) to adopt certain standards or practices in order to sustain legitimacy and avoid sanctions. Mimetic isomorphism takes place when an organisation faces uncertainties (e.g. unstable political support, fluctuating markets etc.) in the organisational field and seeks to limit these uncertainties by emulating successful organisations' practices to render itself more successful. Normative isomorphism takes place when an organisation meets normative expectations in the organisational field. These expectations are typically linked to the professionalization and education and brings changes to organisations when e.g. the organisation use external consultants, who bring along new ideas, or by the recruitment of new managers. The table below list some of

the predictors that DiMaggio and Powell associate with the three processes of institutional adaptation (DiMaggio and Powell, 1983).

Coercive isomorphism (cultural constrains in environment)	Mimetic isomorphism (uncertainty in environment)	Normative isomorphism (normative expectations in environment)
An organisation is most likely to adapt its institution to the environment when...		
<p>...it is constrained by regulation or decisions of another organisation.</p> <p>...it is dependent on a single organisation for e.g. funding, expertise, technology etc.</p> <p>...it is dependent on one organisation for political action/decisions in order to operate.</p>	<p>...it faces great (external) uncertainty (e.g. political or financial uncertainty).</p> <p>...it is able to identify successful organisations, who are 'first movers' or have adapted the institution already .</p> <p>...the choice of institutions to adopt is limited.</p>	<p>...it is hiring on the basis of academic credentials.</p> <p>...it is very integrated in professional organisations and networks.</p> <p>...the field is very professionalised.</p>

Table 1 Predictors of institutional adaptation (adapted from DiMaggio and Powell, 1983)

A few evaluation studies use institutional theory and consider adoption of evaluation practices in organisations and countries (Hansen and Borum, 1999, 2000; Radaelli, 2009; Sager and Rissi, 2011; Verone and Jacob, 2005; Benjamin, 2012). Linked to the use of knowledge, Boswell (2008: 485-6) finds evidence that an organisation uses knowledge to legitimise itself in an unstable and uncertain environment as predicted by institutional theory. Dahler-Larsen argues that "institutional theory can contribute to describe the relatively mechanical diffusion of a phenomenon like evaluation, where the adopted standard in the respective organisation, however, remains a symbolic facade, which does not affect the organisation's core activities significantly." (Dahler-Larsen, 2004: 297; Meyer and Rowan, 1977).

On the basis of the theory and the predictors in Table 1, it is possible to roughly graduate the external pressure that an organisation is under in a high and a low category. Pressures are considered high if the organisation is strongly influenced

or dependent upon other organisations for survival in an uncertain environment (DiMaggio and Powell, 1983). In the opposite case, the pressure is low.

Internal Propensity to Evaluate

Adoption of evaluation practices does not only depend on the pressure and the condition of the organisational environment but also on internal factors in the organisation:

In her study on knowledge use in the European Commission, Boswell (2008) uses Brunsson's distinction of 'action' organisation and 'political organisation', which constitute two extremes on a continuum (Brunsson, 2002). The 'action organisation' is focused on its efficiency because it gains its legitimacy from its ability to produce outputs. Therefore, this type of organisation is likely to use knowledge instrumentally to improve its efficiency and thereby increase its legitimacy. The organisation is likely to have a culture of evaluation to support its desire to increase efficiency. An example of this type of organisation in the public domain could be executive agencies or a homecare provider. The 'political organisation' draws its legitimacy from political decisions and action-taking on issues. This type of organisation has a low propensity to evaluate, because the policy outputs and impacts are often not measurable and therefore more difficult to evaluate. Also, the political organisation does not have a culture of evaluation, and for those reasons, political organisations tend to use knowledge - such as evaluations - symbolically rather than instrumentally when legitimising themselves (Boswell, 2008: 473-474). The distinction between these two ideal types of organisations draws on Meyer and Rowan's (1977) distinction between output legitimacy and normative legitimacy. They write: 'Some organizations use routine, clearly defined technologies to produce outputs. When output can be easily evaluated... efficiency often determines success.' (Meyer and Rowan, 1977: 353-54). And they continue: 'Increasingly, such organizations as schools, R&D units, and governmental bureaucracies use variable, ambiguous technologies to produce outputs that are difficult to appraise.' (Ibid).

Boswell also finds that in the same organisation, management have a tendency to act politically and use knowledge symbolically towards the external organisational environment, while operational staff involved in core production activities tend to

be more instrumental in their use of knowledge. This also comes with different objectives of work; the operational staff having an interest in increasing outputs, while top management also has as their objective to legitimise the organisation in other ways such as demonstrating that the organisation is evaluating 'just like any respectable organisation should be doing'.

In their empirical study on non-profit organisations, Eckerd and Moulton (2011) reach a conclusion that supports (albeit cautiously) the idea in this article that Brunsson's two organisation types can be associated with both adoption of practices (such as evaluation practices) as well as the evaluation use following from these practices. Eckerd and Moulton write; '...while an increase in emphasis [by non-profits] on service provision and innovation is associated with the instrumental uses of evaluation, an increased emphasis on political advocacy or social capital creation is associated with an increase in symbolic uses' (Eckerd and Moulton, 2011: 114). In other words, if the role of the non-profit organisation is hard to measure like 'political advocacy' or 'social capital creation', then the organisation is likely to use evaluations symbolically. Hence, the idea that the propensity to evaluate should have an effect on evaluation use seems to be supported by findings by both Boswell and Eckerd and Moulton.

Applying Brunsson's two ideal type organisations, we find the action organisation in one end and the political organisation in the other. To label the continuum to the context of evaluation use, the term 'internal propensity to evaluate' seems fitting. An organisation with a high internal propensity to evaluate is seeking legitimacy by producing outputs. Therefore, it is interested in increasing its efficiency by e.g. adopting evaluation practices and using evaluation findings to improve efficiency. This type of organisation is characterised by producing measurable outputs and having a culture of measurement. The organisation with low internal propensity to evaluate is characterised by few measureable outputs and little propensity to evaluate and measurement practices. It will primarily use evaluation symbolically to legitimise itself and its activities.

Figure 1 illustrates how external pressure and internal propensity to evaluate determine the adaptation of an organisation to the environment by adoption of an

evaluation institution and subsequently determine the primary mode of evaluation use.

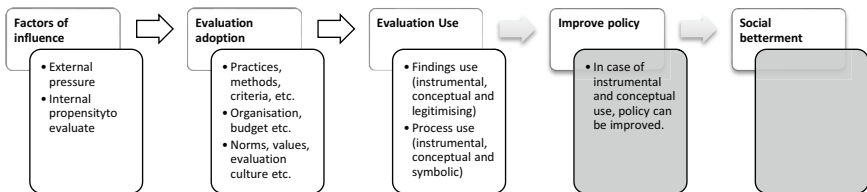


Figure 1 Overview of linkage between two explanatory factors, evaluation adoption and evaluation use

In the following section, a model is presented based on the two factors; external pressure and internal propensity to evaluate. The article considers these two factors to be important in relation to an organisation's adoption and use of evaluation.

Adoption and implications on evaluation use

On the basis of the two factors (or continuums) described previously, four 'extreme positions' or ideal types can be envisaged. These are: Coercive adoption, mimetic adoption, normative adoption and voluntary adoption. The former three are inspired by organisational institutionalism. The latter represents the pure action organisation that voluntarily adopts evaluation practices when needed in order to increase efficiency and outputs much in accordance with the 'common' idea behind evaluation (see Figure 2 below).

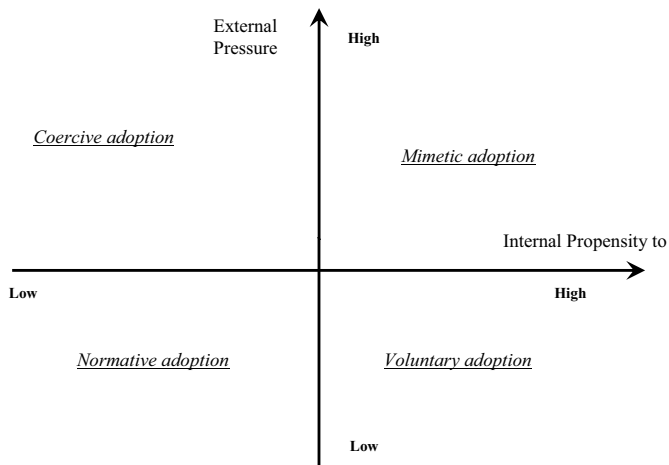


Figure 2 Four 'extreme positions'

The following sections describe and exemplifies the four extreme positions.

Coercive adoption

Coercive adoption of evaluation practices takes place in an organisation that is under, 1) high pressure from its organisational field due to e.g. funding dependency (DiMaggio and Powell, 1983); and 2) because the organisation has a low internal propensity to evaluate due to e.g. non-measurable outputs and no culture of measurement (Boswell, 2008).

This organisation will adopt evaluation practices, but will most likely 'decouple' them from the normal processes of the organisation. Meyer and Rowan (1977) notes the following:

'Sometimes institutional pressures contradict with internal efficiency needs. When that happens, organizations may claim to adapt when they in reality do not. Instead they may decouple action from structure to preserve organizational efficiency while also obtaining legitimacy.'

Formal structures and evaluation practices will be in place to 'satisfy' donors or other funding organisations (DiMaggio, 2001), but paradoxically the organisation does not seek to instrumentally use the evaluations. So, if the organisation uses its evaluations, it will most likely be symbolically to legitimise itself to the organisational field. This explains the behaviour, which is commonly considered a paradox in the evaluation use literature.

One example could be the small non-profit relief organisation working with NGO-support in a failed state. The organisation could be totally dependent on donors and is working in an unstable and politicised area, where funding can be cut almost from one day to the other for numerous reasons, which are out of the organisation's control (e.g. one supported NGO might suddenly be accused of being infiltrated by the government, of being corrupt, contributing to terror etc. and thus harming the PR of the donor organisation, which is also under pressure). Moreover, the organisation is small and overstretched depending on voluntarism and idealism of its staff. It has no internal evaluation capacity though its activities have to be evaluated according to the donor contract. Staff will most likely find evaluation a burden, counter-intuitive and non-applicable with regard to their everyday objectives and tasks related to e.g. capacity building in small NGOs.

Mimetic adoption

Mimetic adoption of evaluation practices takes place in organisations, where the propensity to evaluate and receptiveness towards evaluation is high but where the pressure to evaluate from the organisational field is also high. Because the organisation is highly receptive to the ideas and concepts of evaluation, it will adopt evaluation practices easily and not only because of coercion, but rather strategically to emulate practices from the best in the field in order to learn and

become more efficient. However, the organisation is also working in an uncertain organisational environment, where e.g. budgets are likely to be cut, and therefore the organisation also seeks to boost its normative legitimacy as well as its output legitimacy.

The result is that the organisation will not decouple but rather use evaluations actively to legitimise its programmes or policies (not the organisation itself as was the case in the coercive adoption mode). The organisation will both have formal structures (common guidelines, evaluation schedules and rules etc.) to satisfy the field but also a strong informal evaluation culture.

An example of mimetic adoption could be a school or university. Suppliers of education operate in a volatile and changing field. Teaching methods and curricula are almost permanently scrutinised and debated in public and governments are forced to take action accordingly or as reaction to international benchmarking results etc. At the same time, most outputs are measureable and the organisation has a culture of evaluation and measurement (e.g. exams). Therefore, evaluations are likely to be used both to satisfy the public by legitimising outputs as well as instrumentally to increase efficiency and effectiveness in the organisation.

Normative adoption

Normative adoption of an evaluation institution in an organisation takes place because spread of ideas in professional networks, managers changing position, consultants applying the techniques as part of an assignment in the organisation etc. The organisation is under low external pressure and has low propensity to evaluate and it is therefore not likely to adopt evaluation in the first place. When the organisation evaluates, it is thus more coincidental and linked to a particular manager's idea than an actual institutional practice resulting from a need to legitimise the organisation towards the external organisational environment. Therefore, the evaluation institution will only be weakly institutionalised - if institutionalised at all - and when evaluations are carried out, the most common learning type is likely to be conceptual use or no use. Thus, evaluation will not be a taken-for-granted practice, and consecutive evaluation output will differ much in quality and content due to organisational inexperience with evaluation.

An example of normative adoption could be a national security organisation like the secret service that is under relatively weak external pressure and that does not have internal propensity to evaluate (see study of internal evaluation in the FBI in Sonnichsen, 1999). Such an organisation would normally not evaluate regularly (it might evaluate ones under sudden pressure linked to a particular case of e.g. misuse of power) unless management alone has the idea to do so. That could be the case, if management comes from another sector, where evaluation is more common.

Voluntary adoption

Voluntary adoption of evaluation practices takes place in organisations facing little external pressure that have a high internal propensity to evaluate. The evaluation institution will most likely be adopted 'loosely' and practices will be informal and ad hoc rather than very rule-based, as formal rules and practices are only applied to the extent that they make sense for the overall purpose of increasing efficiency. Every time the organisation evaluates, the purpose is to use the evaluation instrumentally to increase efficiency and organisational outputs. Evaluations are typically designed on a case by case basis to maximise organisational utility.

This type of adoption fits the action organisation. An example of such an organisation could be a large municipality. A municipality produces a range of outputs and the municipality draws its legitimacy primarily from its delivery of these outputs to the citizens. If there are holes in the roads or no maintenance of streetlights, the Mayor will not be re-elected. At the same time (depending on the country), municipalities are relatively autonomous and are usually not required to evaluate their own programmes or initiatives (however, country-wide evaluations of schools and other comparable organisations might be carried out centrally by the state).

Below, Table 2 provides an overview of the four extreme positions of organisational adoption of evaluation institutions and evaluation use.

Adoption				
Adoption mode	Coercive Adoption	Mimetic adoption	Normative adoption	Voluntary adoption
Adoption of evaluation practices	Evaluation practices are directly or indirectly forced upon the organisation by one or more organisations in the field	Evaluation practices are adopted to emulate successful organisations in the field	Evaluation practices are adopted because of professionalization, use of consultancies or new management in the organisation and not because of pressures or because of a perceived need to do so.	Evaluation practices are adopted voluntarily to fit the organisation's needs for evaluation
Consequences for evaluation practices and use				
Purpose of evaluation in organisation	Legitimise organisation to secure its short term survival	Legitimise activities of organisation; e.g. programme or policy	Purpose can vary according to the manager / consultant.	Increase efficiency, output and output-legitimacy.
Evaluation mode	Regular summative evaluations with clear standards matching demands from principal organisation(s).	Regular summative or formative evaluations with practice matching the leader in the organisational field.	Evaluations are unpredictable and irregular.	Ad hoc formative evaluations that are adapted to needs and problem or policy.
Institutionalisation	Very formalised but with discrepancy between activities and evaluation purpose leading to decoupling.	Both formal and informal.	Neither formal nor informal institutionalisation.	Formal structures match need for evaluation. Informal culture of evaluation.
Primary use types	No use or symbolic use	Legitimising and instrumental use	No use or conceptual use	Instrumental and conceptual use
Examples				
Examples	A non-profit organisation entirely dependent on funding from one particular organisation in a highly politicised and volatile area of operation like e.g. NGO support in a failed state.	A school or university is both under strong external pressure to deliver the best education available while at the same time having measureable outputs and a culture of measurement cultivated over time.	Military or secret service organisations with non-quantifiable outputs (e.g. national security) that are so fundamental to the society that the organisation is almost beyond criticism i.e. little external pressure.	A municipality supplies services and draws its legitimacy from doing so.

Table 2 Overview of four ideal typical adoption modes and the associated expected evaluation uses.

Conclusion

This article investigates one of the old paradoxes of evaluation use, namely that evaluations are rarely contributing to policy-making despite it being the main reason for undertaking them. The paradox consists of a logical inconsistency

between, on the one hand, the causal assumption that evaluation leads to policy-improvements, and on the other hand, that the empirical literature on many occasions find non-use and justificatory uses, which do not lead to policy-improvements or social betterment.

We learn that; 1) justificatory use types and non-use need to be better integrated into a theory of evaluation use; and 2) that a theory of evaluation use should take into consideration the organisational and institutional context of the evaluating organisation.

Evaluations are almost exclusively embedded in organisational contexts one way or the other. Typically, evaluations are procured by organisations and carried out by organisations. Therefore, we need to focus more on the organisational context of evaluation and less on the evaluation and its immediate conditioning factors. In other words, evaluation use or non-use are likely to be explained better by understanding the organisational context, in which the evaluation takes place, rather than the evaluation's attributes or other 'micro-level' explanations of conditions and factors. The article finds that the literature on evaluation use has been very good at describing the evaluation, its conditioning factors etc. while neglecting the organisational context, in which the evaluating organisation operates in as well as the organisation's ability to evaluate.

The article demonstrates how organisational institutional theory can explain justificatory uses of evaluation as well as non-use. Organisational institutional theory explains well non-use of evaluation as well as justificatory uses by arguing that organisations sometimes act contrary to their objectives. Therefore, it complements core assumptions of causality and rationality, which are inherent to evaluation. The article provides a short review of the application of institutional explanations in the evaluation literature. Only very few articles applied institutional theory, and only one article really used an institutional framework to predict evaluation use.

A simple explanatory model illustrates the points of organisational institutional theory combined with elements from Brunsson's theory of organisation. The model consists of two explanatory factors (external pressure and the propensity to evaluate) and aims at demonstrating that these factors can explain; 1) the adoption

of evaluation practices that an organisation is likely to make; 2) the subsequent use of evaluations that the organisation is likely to have. By applying institutional theory to evaluation in this way, we can integrate justificatory use categories and non-use meaningfully in a genuine theoretical framework, where organisational rationality is not a priori.

So far, the literature on evaluation use has been much more focused on causal explanations on the theoretical micro-level, i.e. demonstrating that high evaluation quality leads to increases in use. This model, seeks explanations at a theoretical medium-level arguing that it is more important to understand the evaluating organisation's internal and external conditioning factors rather than focussing solely on the evaluation when explaining evaluation use.

Needless to say, further work is needed to operationalise the model. For example, how do we assess the external pressure and the internal propensity to evaluate? Those are the next questions that need to be answered in order to fully take advantage of institutional theory when predicting and explaining evaluation use.

8.2 ARTICLE 2 – EVALUATION IN THE EUROPEAN COMMISSION - FOR ACCOUNTABILITY OR LEARNING?

Reference:

Højlund, S. (2014) *Evaluation in the European Commission - For accountability or learning?* European Journal of Risk Regulation (Forthcoming).

(To avoid any copyright issues in relation to the final print of this thesis, this article is conveyed in its ‘pre-script’ version. A pre-script version is a near-published version of the article as it was after peer-review and language check by the author as well as after the final acceptance of the article by the journal editor. The pre-script version does not include the final editing and formatting of the article by the journal in which the article was published. The author of this thesis recommends readers of this thesis published by the Library of Copenhagen Business School to seek and purchase this article in its original version at the journal in which it was published.)

EVALUATION IN THE EUROPEAN COMMISSION - FOR ACCOUNTABILITY OR LEARNING?

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.

Keywords:

European Union, Evaluation, Policy learning, Accountability, European Commission

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 accountability-oriented (Scriven, 1991). 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 (Gornitzka and Sverdrup, 2011; Smith, 2013). Also, very little attention has been given to the Commission's evaluation system, which has been developing gradually over the last 30 years (Stern, 2009). 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 (Furubo et al., 2002). This warrants a focus on the justification and implementation of its own evaluation practices (Hoerner and Stephenson, 2012).

The article concludes that policy learning and accountability are not necessarily opposed in an evaluation 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.

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 Leeuw and Furubo (2008), four elements 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 (Commission and Jacobsen, 2007) 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) (Commission, 2007a: 17).

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 dichotomous and opposed relationship between the two is well-described in several research areas, such as evaluation studies (Scriven, 1991; Albæk, 1995); audit studies (Power, 2005; Boven, 2005; Dubnick, 2005) and studies on knowledge use and governance in the Commission (Boswell, 2008; Böhling, 2013; Mendez and Bachtler, 2011). 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 (Boven, 2005). In fact, the research areas mentioned above generally assume a contradiction between learning and accountability understood as outcomes.

This article focuses on policy learning happening in the political organization or bureaucracy itself (Schofield, 2004: 288; Bennett and Howlett, 1992). Policy learning is defined as ‘a process of updating beliefs about policy based on lived or witnessed experiences, analysis or social interaction’ (Radaelli and Dunlop, 2013: 923). 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 (1997: 3) defines evaluation as ‘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’.

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’ (Boven, 2005). 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. The legal framework of the Commission relates to the Treaty, the Financial Regulation as well as other community legislation (Harlow, 2002).
3. **Financial** accountability is related to budgetary execution and financial controls that an organization can be subject to. One example is audits (Laffan, 2003).
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 (Wille, 2010).

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.

Methodology and data

This article is based on a historical case study of the Commission's evaluation system (for similar approach to studies of the Commission see for example Wonka, 2008; Schout, 2009; Smith, 2013; Böhling, 2013; Burns, 2012). 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 (Bazeley, 2013) and coded according to the interviewees' perceptions of accountability and policy learning (Schreier, 2012; Mayring, 2004; Kohlbacher, 2006).

Analysis

The analysis is presented chronologically and broken down into four periods. Each period emerged from the data and was marked by one or more events related to the evaluation system's focus on policy learning and/or accountability. Table 8-1 sums up the periods.

Table 8-1 Periods of implementing evaluation in the Commission

	1980–94	1995–99	2000–06	2007–13
Key events	<ul style="list-style-type: none"> - Increased budget - NPM paradigm starts in Europe 	<ul style="list-style-type: none"> - SEM2000 - Evaluation linked to policy-cycle - Evaluation units - MEANS guidelines 	<ul style="list-style-type: none"> - Kinnock reform - Better Regulation - Impact Assessment - Evaluation linked to control standards - SG takes over evaluation from DG Budget 	<ul style="list-style-type: none"> - Evidence-based policy making - EP and CoA policy evaluation capacity - DGs develop regulatory evaluation guidelines - Fitness check

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 (Summa and Toulemonde, 2002: 409). 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 (Christiansen, 1997; Ellinas and Suleiman,

2008). 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 (Summa and Toulemonde, 2002: 412; Laffan, 1997b). 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 (Eser and Nussmueller, 2006). 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.' Several interviewees refer to the New Public Management (NPM) reforms in Europe that stressed a focus on control of public spending, accountability, and effectiveness and efficiency of public interventions (Pollitt and Bouckaert, 2004; Ellinas and Suleiman, 2008). 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.

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 (Commission, 1996). 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 (Commission, 1996: 1).

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 important. 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.' (Council, 1995). 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 (Commission, 1996; Laffan, 1997b). 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 (Commission, 1996).

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. 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... evaluation 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 (evaluations)' (Commission, 1996). 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 (Summa and Toulemonde, 2002: 420).

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 (Eser and Nussmueller, 2006). 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 (see also Højlund, 2014a).² 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

² 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.

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 (Commission, 1999). It concluded that more needed to be done in order to 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 (see Laffan, 1997a). 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 (Toulemonde et al., 2005: 82). 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.

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

In 1999, the Santer Commission resigned due to alleged corruption (Spence, 2000). 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 (Commission, 2000b; Kassim, 2008; Schön-Quinlivan, 2008). Focus on financial accountability increased and financial management, control and audit came to play an important role in Commission thinking (Bauer, 2008; Mendez and Bachtler, 2011).

To implement the Kinnock reform in relation to learning from evaluation, the Commission issued the Communication 'Focus on Results: Strengthening Evaluation of Commission Activities' (Commission, 2000a). 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. (2005), 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 (Commission, 2007a; Commission, 2007b). 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 the implementation was reinforced by making evaluation subject to audits and thus adding an extra level of accountability.

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 (Toulemonde et al., 2005: 78). 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 (Ellinas and Suleiman, 2008; Mendez and Bachtler, 2011). 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 (Toulemonde, 2000; Batterbury, 2006; Eser and Nussmueller, 2006). 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 (Williams et al., 2002: Annex, p. 61). 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' (see executive summary in Laat, 2005). 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’ (Ecorys and COWI, 2008: 13-14). 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 as an important part of the Better Regulation agenda of 2000 (Kassim, 2008; OECD, 2002; Francesco et al., 2011). 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 (Radaelli and Meuwese, 2010). 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 impact assessment. With all the resources going into impact assessment 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. Hence, legal accountability was the focal point in this period in order to 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).

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 due 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. (Commission, 2010: 4).

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.' (Commission, 2007a). Particularly in relation to policy learning, the SG issued the communication '*Reinforcing the Use of Evaluation*' (Commission, 2007a), 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.

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 Info, 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 (Commission, 2007a) 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 (Laat, 2005; Williams et al., 2002; Auditors, 2005). 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 accountability (Commission, 2013). 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 (for a similar findings related to audit see Mendez and Bachtler, 2011). However, as systematic regulatory evaluation is implemented with the new guidelines, democratic accountability is reinforced. As we learned previously, programme evaluation do 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.

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.

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

Table 4 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

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.

References

- Albæk E. (1995) Between Knowledge and Power: Utilization of Social Science in Public Policy Making. *Policy Sciences* 28: 79-100.
- Auditors ECo. (2005) Presidential Letter.
- Batterbury SCE. (2006) Principles and purposes of European Union policy evaluation. *Regional Studies* 40: 179-188.
- Bauer M. (2008) Introduction: Organizational change, management reform and EU policy-making. *Journal of European Public Policy* 15.
- Bazeley P. (2013) *Qualitative Data Analysis with NVIVO*, London: Sage Publications Ltd.
- Bennett CJ and Howlett M. (1992) The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences* 25: 275-294.
- Böhling K. (2013) Sidelined Member States: Commission learning from Experts in the Face of Comitology. *Journal of European Integration*.
- 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.
- Boven M. (2005) Public Accountability. In: Ferlie E, Lynn LE and Pollitt C (eds) *The Oxford Handbook of Public Management*. Oxford: Oxford University Press.
- Burns C. (2012) How and When Did We Get Here? An Historical Institutional Analysis of EU Biotechnology Policy. *Journal of European Integration* 34: 341-357.
- Christiansen T. (1997) Tensions of European governance: politicized bureaucracy and multiple accountability in the European Commission. *Journal of European Public Policy* 4: 73-90.
- Commission E. (1996) Concrete steps towards best practice across the Commission. In: Secretariat-General (ed). Brussels.
- Commission E. (1999) Spending more wisely: Implementation of the Commission's evaluation policy. In: Secretariat-General (ed). Brussels.
- Commission E. (2000a) Focus on results: Strengthening evaluation of Commission activities. In: Commission E (ed) SEC(2000)1051. Brussels.
- Commission E. (2000b) *Reforming the Commission: A White Paper*. Brussels.

- Commission E. (2007a) Evalueringsaktiviteter & -resultater på tværs af Europa-Kommissions tjenestegrene. In: Jacobsen S (ed) Dansk Evalueringsselskabs konference. Kolding, Denmark.
- Commission E. (2007b) Responding to Strategic Needs: Reinforcing the use of evaluation. In: Secretariat-General (ed). Brussels.
- Commission E. (2007c) Revision of the Internal Control Standards and Underlying Framework - Strengthening Control Effectiveness. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2010) Smart Regulation in the European Union. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2013) Public Consultation on Commission Guidelines for Evaluation. In: Secretariat-General (ed). Brussels: European Commission.
- Council E. (1995) Financial Regulation Amendment. In: Union TCotE (ed).
- Dubnick M. (2005) Accountability and the promise of performance - In search of the Mechanisms. *Public Performance & Management Review* 28: 376-417.
- Ecorys and COWI. (2008) Study on the state of implementation of Activity Based Management in the European Commission. Rotterdam.
- Ellinas A and Suleiman E. (2008) Reforming the Commission: between modernization and bureaucratization. *Journal of European Public Policy* 15: 708-725.
- Eser TW and 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.
- Francesco FD, Radaelli CM and Troeger VE. (2011) Implementing regulatory innovations in Europe: the case of impact assessment. *Journal of European Public Policy* 19: 491-511.
- Furubo J-E, Rist RC and Sandahl R. (2002) *International Atlas of Evaluation*, London: Transaction Publishers.
- Gornitzka Å and Sverdrup U. (2011) Access of experts: information and EU decision-making. *West European Politics* 34: 1452-1474.
- Harlow C. (2002) *Accountability in the European Union*, Oxford: Oxford University Press.

- Hoerner J and Stephenson P. (2012) Theoretical Perspectives on Approaches to Policy Evaluation in the EU: The Case of Cohesion Policy. *Public Administration* 90: 699-715.
- Højlund S. (Forthcoming) Evaluation Systems and their Implications on Evaluation Use – The Case of the European Union. *Evaluation* Forthcoming.
- Kassim H. (2008) 'Mission impossible', but mission accomplished: the Kinnock reforms and the European Commission. *Journal of European Public Policy* 15: 648-668.
- Kohlbacher F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research* 7.
- Laat Bd. (2005) Study on the Use of Evaluation Results in the Commission. In: Laat Bd (ed). Paris: Technopolis.
- Laffan B. (1997a) The Finances of the European Union, Basingstoke: Macmillan.
- Laffan B. (1997b) From policy entrepreneur to policy manager: the challenge facing the European Commission. *Journal of European Public Policy* 4: 422-438.
- Laffan B. (2003) Auditing and accountability in the European Union. *Journal of European Public Policy* 10: 762-777.
- Leeuw FL and Furubo J-E. (2008) Evaluation Systems : What Are They and Why Study Them? *Evaluation* 14: 157-169.
- Mayring P. (2004) Qualitative Content Analysis. In: Flick U, Von Kardorff E and Steinke I (eds) *A Companion to Qualitative Research*. London: SAGE, 266-269.
- Mendez C and Bachtler J. (2011) Administrative reform and unintended consequences: an assessment of the EU Cohesion policy 'audit explosion'. *Journal of European Public Policy* 18: 746-765.
- OECD. (2002) Regulatory policies in OECD Countires: From Interventionism to Regulatory Governance. Paris: OECD.
- Pollitt C and Bouckaert G. (2004) *Public Management Reform: A Comparative Analysis*, Oxford University Press.
- Power M. (2005) The Theory of Audit Explosion. In: Ferlie E, Lynn LE and Pollitt C (eds) *Oxford Handbook of Public Management*. Oxford: Oxford University Press.

- Radaelli CM and Dunlop CA. (2013) Learning in the European Union: theoretical lenses and meta-theory. *Journal of European Public Policy* 20: 923-940.
- Radaelli CM and Meuwese ACM. (2010) Hard Questions, Hard Solutions: Preceduralism through Impact Assessment in the EU. *West European Politics* 20: 923-940.
- Schofield J. (2004) A model of learned implementation. *Public Administration* 82: 283-308.
- Schön-Quinlivan E. (2008) Implementing organizational change - the case of the Kinnock reforms. *Journal of European Public Policy* 15: 726-742.
- Schout A. (2009) Organizational learning in the EU's multilevel-governance system. *Journal of European Public Policy* 16: 1124-1144.
- Schreier M. (2012) *Qualitative Content Analysis in Practice*, Thousand Oaks: SAGE Publications Ltd.
- Scriven M. (1991) Beyond Formative and Summative Evaluation. In: McLaughlin M and Philips DC (eds) *Evaluation and Education: At Quarter Century*. 3 ed. Chicago: University of Chicago Press.
- Smith AS. (2013) How the European Commission's Policies Are Made: Problematisation, Instrumentation and Legitimation. *Journal of European Integration*.
- Spence D. (2000) Plus ça change, plus c'est la même chose? Attempting to reform the European Commission. *Journal of European Public Policy* 7: 1-25.
- Stern E. (2009) Evaluation policy in the European Union and its institutions.
- Summa H and Toulemonde J. (2002) Evaluation in the European Union: addressing complexity and ambiguity. In: Furubo J-E, Sandahl R and Rist R (eds) *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.
- Toulemonde J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *Vierteljahrshefte zur Wirtschaftsforschung* 69: 350-357.
- Toulemonde J, Summa H and Usher N. (2005) Assessing EU Evaluations. In: Schwartz R and Mayne J (eds) *Quality matters: seeking confidence in evaluation, auditing and performance reporting*. New Brunswick: Transaction Publishers.

- Vedung E. (1997) *Public Policy and Program Evaluation*, New Brunswick: NJ: Transaction Publishers.
- Wille A. (2010) Political–Bureaucratic Accountability in the EU Commission: Modernising the Executive. *West European Politics* 33: 1093-1116.
- Williams K, Laatz B and Stern E. (2002) *The Use of Evaluation in the Commission Services*. Paris: Technopolis.
- Wonka A. (2008) Decision-making dynamics in the European Commission: partisan, national or sectoral? *Journal of European Public Policy* 15: 1145-1163.

8.3 ARTICLE 3 – EVALUATION USE IN EVALUATION SYSTEMS – THE CASE OF THE EUROPEAN COMMISSION

Reference:

Højlund, S. (2014) *Evaluation use in evaluation systems – the case of the European Commission*. Evaluation 20(4), 428-446.

<http://evi.sagepub.com/content/20/1/26.abstract>

*(To avoid any copyright issues in relation to the final print of this thesis, this article is conveyed in its **‘pre-script’ version**. A pre-script version is a near-published version of the article as it was after peer-review and language check by the author as well as after the final acceptance of the article by the journal editor. The pre-script version does not include the final editing and formatting of the article by the journal in which the article was published. The author of this thesis recommends readers of this thesis published by the Library of Copenhagen Business School to seek and purchase this article in its original version at the journal in which it was published.)*

EVALUATION USE IN EVALUATION SYSTEMS – THE CASE OF THE EUROPEAN COMMISSION

This article investigates the European Union's evaluation system and its conduciveness to evaluation use. Taking the European Commission's LIFE programme as its case, the article makes an empirical contribution to an emerging focus in the literature on the importance of organization and institutions when analyzing evaluation use. By focusing on the European Union's evaluation system the article finds that evaluation use mainly takes place in the European Commission and less so in the European Parliament and the European Council. The main explanatory factors enabling evaluation use relate to the system's formalization of evaluation implementation and use securing quality, timeliness and evaluation capacity in the Commission. At the same time, however, the system's formalization also impedes evaluation use, reducing the direct influence of evaluations on policy-making and effectively 'de-politicizing' programme evaluations and largely limiting their use to the level of programme management.

Keywords

European Commission, LIFE programme, programme management, evaluation systems, evaluation use.

Introduction

In the last thirty years, evaluation practices have spread and become common practice in most OECD countries. In parallel with the spread of evaluation practices, national and international organizations have to a large extent institutionalized and ritualized evaluation practices into what has been termed 'evaluation systems'. Prior research has hypothesized about the implications of evaluation systems on evaluation use but the phenomenon still needs more empirical investigation (Leeuw and Furubo, 2008; Furubo, 2006; Rist and Stame, 2006).

This article investigates evaluation use in the European Union's (EU) evaluation system where evaluation practices have been institutionalized over several decades, particularly in the European Commission. Thereby the article aims to contribute empirically to the emerging focus in the evaluation literature on contextual organisational factors explaining evaluation use (Højlund, 2014b). It does so by focusing on the evaluation system understood as the institutionalization of evaluation practices in the EU organizational bodies – in particular the Commission. The evaluation system thus becomes the systemic setting and context in which evaluation use is analyzed. The underlying assumption is therefore that the attributes of the evaluation system can explain the way evaluations are used in this particular system. Thereby the article relies on newer theoretical contributions on evaluation systems (e.g. Leeuw and Furubo, 2008) as well as a more general introduction of organizational theory into the theoretical landscape of evaluation use (Højlund, 2014b).

The main contribution of the article is to improve our understanding of the implications that an evaluation system has on evaluation use. Formal and informal organizational practices both impede and enable the use of evaluation. Despite some evaluation use by policy-makers, the article finds that most use takes place on the programme management level in the Commission. Thus evaluation use at the programme level tends to be instrumental, strategic, legitimizing and informational, whereas policy-makers use evaluations strategically and to get information, albeit to an overall much lesser extent.

This article reports five overall findings: first, the strong formalization of evaluation practices in the system enables findings use but impedes process use.

Evaluations are thus typically used after their completion and not during their implementation, due to the Commission's stress on the independence of the evaluator. Second, significant findings use typically takes place at 'decision points' every seventh year in the programming phase. Other uses do take place at the programme management level, but instrumental use that affects the programme or other policies are typically indirect as evaluations feed into impact assessments and ex-ante evaluations of the new programme. Third, evaluations have little overall relevance for policy-makers and programme management alike. In particular, evaluations are not relevant for policy-makers outside the Commission due to competing information and their technical nature. Fourth and for the above reasons, programme evaluations are 'de-politized' and generally they are not something policy-makers participate in nor have any use for. Fifth and finally, the 'de-politization' represents a paradox since it is the European Parliament and Member States in the European Council that required the Commission to setup the evaluation system and also demand evaluations to be a part of the legal basis of programmes such as LIFE. But this article shows that policy-makers rarely use the evaluations, while at the same time the Commission is burdened by the evaluations and numerous other internal and external assessments and audits.

The article is organized as follows: in the first sections, evaluation systems and evaluation use are defined and discussed. Then a section follows describing the methodology used in the analysis as well as the analysis itself. Finally, a conclusion is followed by reflections on potential extensions of the research on evaluation systems and evaluation use.

Evaluation systems

The discussion on evaluation systems took a leap forward with the book *From Studies to Streams* edited by Ray C. Rist and Nicoletta Stame in 2006 (Rist and Stame, 2006). Several subsequent studies picked up the baton (Imam et al., 2007; Leeuw and Furubo, 2008; Williams and Imam, 2007) improving our conceptual understanding of the phenomenon. The literature on evaluative systems relate to a broader focus in the evaluation literature on evaluation as a phenomena understood as contingent on complex societal contexts such as institutions, norms and power (Dahler-Larsen, 2012; Van der Knaap, 1995). Particularly Peter Dahler-Larsen has used institutional organizational theory to explain the phenomena of evaluation and adaptation of evaluative practices by public

organizations. Only recently has the same theoretical framework been used explicitly to explain the phenomena of evaluation use (Højlund, 2014b).

This article builds on the fundamental idea that institutions and organization determine evaluation use. The focus in the article is on evaluation systems because an evaluation system is composed of several organizational entities that to some degree share formal and informal evaluative practices and norms i.e. a shared evaluation institution. Leeuw and Furubo (2008) stress the following four elements constituting an evaluation system:

1. Participants in the evaluation system share a common understanding of the objectives of evaluation and the means by which the objectives are attained.
2. The evaluation system is institutionalized formally in at least one organizational structure, in which it is separated from the operational structure of this organization. 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 following-up on evaluations.
3. Evaluation systems are permanent in the sense that their setup has no time-limitation. Moreover, evaluations are undertaken continuously and systematically and in relation to previous and future evaluations.
4. In the evaluation system, evaluations are organized and planned so that they relate to the cycle of activities of the organization or the evaluand (e.g. budget or policy cycle).

Based on the four elements above and other contributions, a definition of an evaluation system could be summarized as follows: "an evaluation system is permanent and systematic formal and informal evaluation practices taking place and institutionalized in several interdependent organizational entities with the purpose of informing decision-making and securing oversight."

In relation to evaluation use, evaluation systems are generally assumed to have a negating effect on information- and knowledge use in policy-making (Leeuw and Furubo, 2008; Pollitt et al., 1999). Previous studies suggest that evaluative knowledge tends to be made relevant primarily for administrators and not for policy-makers and that use in administrations will be linked to procedural assurance and legitimization of the organization rather than used to inform policy-making (see also Furubo, 2006; Langley, 1998). The purpose of this article is to

continue the research on evaluation systems' effect on evaluation use and provide empirical evidence where presently there is little.

The EU evaluation system constitutes a very good example (Stern, 2009). Evaluation is an integral part of the activity-based management and budgeting system of the Commission and thus formally related to decision-making regarding EU budgetary allocations. The system's core consists of the European Commission (the Commission), the European Parliament (EP) and the European Council (the Council). As the EU executive body, the Commission is also responsible for commissioning, implementing and disseminating evaluations of EU programmes and policies. The Commission has a legal obligation to evaluate programmes and policies as stipulated in the Commission's management policies as well as the legal basis of the programmes and policies. For this reason, the Commission has institutionalized evaluation practices over the last thirty years in each Directorate General (DG) through evaluation policies, guidelines and standards. In the DGs, designated evaluation units supervise and guide evaluation activity with support from the Secretariat-General. The evaluation units are subject to internal audits as they are described in the Internal Control Standards of the Commission services. It is the Commission that undertakes most evaluations in the system, but the EP and Member States do also carry out or commission evaluations, usually subject to EU evaluation standards and supervised by the Commission (in the case of Member States). About 80% of all evaluations in the Commission are externalized to consultants or groups of experts (Commission and Jacobsen, 2007) and the consultancies are thus also a part of the evaluation system.

Evaluation use

In the 60s, scholars started to question whether knowledge is used to inform policy-makers in order to improve policy (Lazarsfeld et al., 1967; Weiss and Bucuvalas, 1980). The answer to this question was partly negative and the situation was referred to by some scholars as a 'utilization crisis' (Patton, 1997; Floden and Weiner, 1978). In the cases where evaluation information was actually used, evaluation research conceptualized use-categories, which have not changed significantly over the years (Leviton, 2003). Four main types of evaluation use emerged: instrumental-, conceptual-, process- and symbolic use. These four categories are still used as the basis for most research, though newer and more

elaborate conceptual frameworks have been suggested (Kirkhart, 2000; Henry and Mark, 2003; Weiss, 1998).

In the wake of the disenchantment associated with the scarce evidence of use from evaluations, the literature instead asked why evaluations were used or not used (Leviton and Hughes, 1981; Cousins and Leithwood, 1986). Studies focused on factors related to the attributes of the evaluation (e.g. methodology, quality, relevance of findings etc.) or the immediate contextual factors pertaining to the organization in which the evaluation is implemented (e.g. political climate, timing of the evaluation relative to decision-making etc.). These categories were empirically informed from the late 70s and onwards (see for example Leviton and Hughes, 1981). This article leans on the broad definition of evaluation use provided by Johnson et al. (2009b: 378): "any application of evaluation processes, products, or findings to produce an effect." This definition captures the variety of use types applied in this article (see Research question and design).

In relation to the interest in evaluation systems, Furubo (2006: 151) suggests that the literature could still benefit from a better understanding of the effects that evaluation systems have on evaluation use. In general, it seems that most evidence on evaluation use is still linked to single ad hoc evaluation studies rather than systematic evaluation information and does not specifically address the evaluation system. It is on this topic that this article makes its contribution. Similarly, only very few studies in the evaluation literature take into account organizational explanations when analyzing evaluation use (Højlund, 2014b).

Research question and design

This article investigates whether *evaluation systems are conducive to evaluation use*. In order to properly answer this question, three sub-questions are proposed: 1) *How* are evaluations used in evaluation systems? 2) *who* uses evaluation findings in evaluation systems? 3) *why* do – or do not – evaluation systems support the utilization of evaluation findings?

Consequently, the *dependent factors* are evaluation uses. Considering evaluation use, the article distinguishes between ten different types of evaluation use organized under two headings: 'findings uses' (instrumental, conceptual, legitimizing, information and strategic) and 'process uses' (instrumental,

conceptual, symbolic, information and strategic) (Alkin and Taut, 2003; Weiss, 1998; Leviton and Hughes, 1981; Leviton, 2003). Table 1 below gives an overview of the ten types of evaluation use in the analysis.

Table 1 The ten evaluation use types.

Process use <i>(evaluation use during the process of evaluation; typically use of preliminary results etc.)</i>	Findings use <i>(evaluation use after the evaluation process has ended; typically use of the findings and recommendations of a report)</i>
<ul style="list-style-type: none"> - Instrumental: The evaluation findings are used to change the evaluand or the conditions that it is working under. - Conceptual: The evaluation is used to gain new conceptual knowledge. - Symbolic: The evaluation is used to legitimize the organisation that is responsible for the evaluand. - Information: The evaluation is used to acquire information. - Strategic: The evaluation is used for advocacy. 	<ul style="list-style-type: none"> - Instrumental: The evaluation findings are used to change the evaluand or the conditions that it is working under. - Conceptual: The evaluation is used to gain new conceptual knowledge. - Legitimizing: The evaluation is used to legitimize the evaluand. - Information: The evaluation is used to acquire information. - Strategic: The evaluation is used for advocacy.

The ten use categories are informed by existing literature on evaluation use. Hence, Alkin and Taut (2003) proposed the conceptual division between findings use and process use as they recognized that process use (use of the evaluation *during* the evaluation process) was not a type of use in itself as it could both be instrumental, conceptual and legitimizing (e.g. evaluation is legitimizing the organization).³ In addition to instrumental-, conceptual-, legitimizing- and symbolic uses, the evaluation literature has also proposed two other categories of uses that relate to the use of evaluation understood as simply a source of information – a type of use that often precedes other use forms (Alkin and Stecher,

³ At this point it should be noted that these well-known categories are all ex-post to evaluation implementation and therefore do not include the effects and evaluation that exist ex-ante as a consequence of evaluation anticipation (for example redressing or window-dressing before the evaluator starts working). This use type was not included in the analysis because data on such uses is difficult to collect up to one decade after the evaluation was finalized. Moreover, the data collection allowed for accounts of ex-ante uses by asking several open questions, but no examples were given by the interviewees.

1983; Finne et al., 1995). An instance of information use would be to use evaluation information in a presentation or simply reading the evaluation to acquire knowledge. 'Information use' can take place both before and after the completion of the evaluation and is thus both related to 'findings use' and 'process use'.

Finally, scholars have pointed to a fifth type of use often referred to as 'strategic use'. Strategic use is distinguished from symbolic and legitimizing use types as it is not related to securing organizational or programme legitimacy, but rather to advocacy in relation to decision- or policy-making (Weiss, 1992; Pröpper, 1987: cited in Van der Knaap, 1995: 211). Strategic use needs to be included because legitimizing use originally proposed by Rich (1977), does not appropriately cover the strategic and political use of arguments found in evaluations and used to justify political arguments and decisions. Legitimizing use is the evaluating organization justifying the programme or policy that is evaluated. However, in an evaluation system there are more actors involved, who have an interest in using the evaluation as a source of legitimacy to back their positions and political arguments. This type of use I call 'strategic use' as it does not necessarily have to be related to legitimizing the programme (legitimizing use) or the justification of the evaluating organization (symbolic use). Instead it is related to other issues, such as when facts from the evaluation are used to back a certain position in the renegotiation of a new programme.

The overall *independent factor* is the context of the EU evaluation system. However, to better understand the processes in play, the analysis contains intermediate explanatory factors providing for a more detailed understanding of barriers and enablers of evaluation use in the evaluation system. Here, the article relies on the conceptual framework of Cousins and Leithwood (1986) and Johnson et al. (2009a). They refer to twelve specific factors that can influence evaluation use. These factors are divided into two categories. The first category is 'evaluation implementation' (a. evaluation quality, b. evaluation credibility, c. evaluation relevance, d. communication quality, e. evaluation findings, f. timeliness), and the second one is 'decision or policy setting' (a. information needs, b. decision characteristics, c. political climate, d. competing information, e. personal characteristics, f. receptiveness to evaluation). The first category relates to traits about the evaluation in question. The second relates to factors linked to the

organizational decision-making and other contextual factors not directly linked with the evaluation.

Table 2 Explanatory factors.

Explanatory factors	Explanation
<i>Decision- and policy setting</i>	
Commitment and receptiveness	Staff commitment, receptiveness, responsiveness etc. to evaluations, evaluation procedures and practices.
Competing information	The influence of other reports, studies and prior knowledge.
Decision characteristics	The influence of the procedures and practices of decision-making including the barriers and enabling factors that are related to decision-making (e.g. timing, legal framework etc.)
Information needs	The influence of new information on the performance and of the organization.
Personal characteristics	The influence of the involved person's personalities and experience.
Political climate	The influence of the saliency of an issue, political or public focus.
<i>Evaluation implementation</i>	
Timeliness	Timeliness of reports and other deliverables in the evaluation implementation.
Credibility	Perceived credibility of the evaluation overall as well as findings.
Evaluation quality	Overall perception of evaluation quality, soundness of methods and methodology.
Findings	Saliency of findings, conclusions and recommendations.
Relevance	Overall relevance of the evaluation including its methodology, methods and evaluation questions.
Communication quality	Quality of communication in the evaluation deliverables.

Data and methodology

This article analyzes the use of four evaluations of the Commission's Programme for the Environment and Climate Action (LIFE) over a ten-year period (2003-2013). The case is thus the EU's LIFE programme. Case studies like this one are common to the evaluation literature and mirror the fact that interventions and their evaluations are often uniquely tied to a particular organizational or systemic context as is the case here (Easterby-Smith et al., 2000). The EU evaluation system is a well-constituted evaluation system that matches the definition of an evaluation system as described earlier.

The choice of the LIFE programme as the case was made because of data availability and because the LIFE programme has experienced a full Commission evaluation cycle (ex-ante, midterm, final and ex-post) and therefore it represents a complete picture of evaluation use over an entire policy cycle as well as an entire evaluation cycle. Further, evaluation use in the Commission has been given little attention by researchers so far (see Bienias and Iwona, 2009; Zito and Schout, 2009) with the exception of two Commission-sponsored reports (Williams et al., 2002; Laat, 2005). This is unfortunate, because the Commission is important in terms of spreading evaluation practices in Europe (Toulemonde, 2000; Furubo et al., 2002).

The analysis is based on sixteen semi-structured in-depth interviews and eight follow-up interviews. The informants were sampled purposefully according to relevance and availability and consisted primarily of staff from the Directorate General for the Environment (DG ENV), consultants that performed the evaluations, representatives of Members of the EP's Committee for the environment (ENVI-Committee) and Council members (Ritchie et al., 2003). In addition, thirty six background interviews were conducted with Commission staff working in other DGs on other EU programmes to qualify the information and understand the evaluation system. The analysis also included relevant documents such as the four retrospective evaluations of the LIFE programme (Midterm, 2003; ex-post, 2009; midterm, 2010; final, 2013) and several other documents including DG ENV presentations to the Committee of Regions and the EP, internal Commission documents, the combined ex-ante and impact assessment (IA) along with explanatory policy fiches and Commission position papers for the new LIFE programme 2014-2020.

The methodology applied in the article was based on the principles of qualitative content analysis (Schreier, 2012; Mayring, 2000) and the actual coding and analysis of data was carried out using the NVIVO software package (Bazeley, 2013). The first sixteen semi-structured interviews were analyzed with a view to existing conceptual frameworks developed in the evaluation literature and described earlier. The eight follow-up interviews were conducted to check for saturation. The semi-structured interview guides gave the interviewees flexibility to elaborate on evaluation use and explanatory factors in relation to the evaluation in question and to the extent the interviewee found it relevant. Coder reliability

was sought by using the existing conceptual frameworks and subsequently running three rounds of coding on the interview data (Kohlbacher, 2006; Mayring, 2004). Further, the credibility of the findings was strengthened by a prolonged engagement in the field, conducting interviews in four consecutive waves over a period of one year. Findings were triangulated and validated with document data and follow-up interviews and interpretations were checked against interview data. Interviewees were debriefed and had the opportunity to comment on the findings of the article, and peers with comprehensive knowledge on the subject gave important comments on the draft article before submission. Finally, the researcher has several years of experience with evaluation of EU programmes including work as an evaluator on the ex-post evaluation of LIFE.

Analysis

The analysis is divided into three sections in answer to the three research questions. The first section is dedicated to the use of four LIFE evaluations produced between 2003 and 2013. The second section treats the explanations for evaluation use and the final section in the analysis summarizes who the users of the LIFE evaluations are. In each section the findings are recapped at the end of the section.

Uses of LIFE evaluations

This section contains an analysis of the process uses as well as the findings uses of the LIFE evaluations. Table 3 summarizes the distribution of the qualitative codes on three groups of interviewees and provides examples of interviewee quotes. Most notable is the lack of process use, but there are several interesting patterns in the findings uses as well that are described and summarized below.

	Total	Evaluators*	Programme managers**	Policy-makers***	Quotes
Findings use					
Information use	Count: 57 Nominal %: 100% (Relative %: 100%)	17 30% (44%)	20 35% (26%)	20 35% (31%)	- "...of course you read ...what other solutions [were] proposed by the evaluations" (EP policy-maker) - "Every time we do an evaluation we present results to the Life Committee." (Programme manager)
Conceptual use	Count: 22 Nominal %: 100% (Relative %: 100%)	6 27% (43%)	16 73% (57%)	0 0% (0%)	- "But all these discussions [about the effects of the programme] come regularly when we are evaluating, because you get to pose such questions." (Programme manager)
Instrumental use	Count: 59 Nominal %: 100% (Relative %: 100%)	11 19% (31%)	46 78% (65%)	2 3% (3%)	"...we did it to say what we politically wanted in the seventh [Environmental Programme], and yes that is sometimes based on the things said in an evaluation.... we always try to turn it into a forward-looking perspective." (EP policy-maker) "So the recommendation was to do more for communication with the water unit. So, then we did that." (Programme manager) "What we are after in evaluations is for [them] to support policy-change." (Key informant in the Commission)
Legitimizing use	Count: 29 Nominal %: 100% (Relative %: 100%)	3 10% (18%)	20 69% (60%)	6 21% (22%)	- "... it was useful ...that [the contractor] - when talking to all the units - raised the profile of the LIFE programme." (Programme manager) - "...our evaluation reports and results are used and when we have them, it is something that people are happy about, because they prove that we in most cases are doing a good job." (Key informant in the Commission) - "...if you want to survive in this system, then things have to look good. And if it does not look good, then you better have some good reasons, so then you get a consultant..." (Programme manager) - "No, but I think in general they are much more used, as argumentation for your political message, ...when we write a question to a Commissioner, we can say: 'in the evaluation it says that-... How come you still have not done anything? Why have you not done anything? We want you to do something.' Now, that's what we use it for." (EP policy-maker)
Strategic use	Count: 36 Nominal %: 100% (Relative %: 100%)	1 3% (5%)	22 61% (56%)	13 36% (39%)	- "...[programme beneficiaries] generally say that they think it is a problem that when they make an innovative project proposal they have to wait a year to receive the money at which time the project is not that innovative any more. That's fair enough and I do not need an evaluation to know that. But the evaluation can help us to communicate that." (Programme manager) - "[The evaluation] is used on different levels, because even within the Commission there are several people that needs to be convinced to work in one or the other direction." (Programme manager)

Process uses					
Information use	Count: 1 Nominal %: 100% (Relative %: 100%)	0 0% (0%)	1 100% (100%)	0 0% (0%)	-
Conceptual use	Count: 8 Nominal %: 100% (Relative %: 100%)	3 38% (55%)	5 63% (45%)	0 0% (0%)	-
Instrumental use	Count: 1 Nominal %: 100% (Relative %: 100%)	0 0% (0%)	1 100% (100%)	0 0% (0%)	-
Strategic use	Count: 5 Nominal %: 100% (Relative %: 100%)	4 80% (87%)	0 0% (0%)	1 20% (13%)	- "...so you put in your report, [that] it was suggested [by stakeholders that] the way of solving problem X was to do this. And that sort of gets into your conclusions. Say, you know, you can't see any conclusions, where it has come from, you have to sort of dig back in the report and say there is this problem, this will solve it, and then that becomes, you know, a sort of rubber stamp.... to their policy idea." (Evaluator)
Symbolic use	Count: 13 Nominal %: 100% (Relative %: 100%)	9 69% (82%)	4 31% (18%)	0 0% (0%)	- "...the legitimacy [from evaluations] is by now almost automatic...." (Programme manager)

Table 3. References to evaluation use by interviewees.

The table displays the number of coding instances in each use category and on three interviewee categories. The count represents an interviewee's mention of an instance of evaluation use. In bold:

The nominal row percentage. In parenthesis: The relative row percentage calculated as a function of the count and the number of interviewees in each category.

* Consultants. ** DG ENV Staff. *** EP- and Council-representatives.

Process use. The interview data on process use rarely contains references to process use. The few accounts of process use in the data concerns mainly strategic use. In the evaluation process, key stakeholders have the opportunity during the evaluation implementation to influence evaluation findings by coordinating answers to interview questions or raising particular issues of concern in interviews. The impact from this can be directed towards short-term decision-making within the Commission as well as towards programme change. Additionally, the evaluation system makes a strong link between evaluations inducing evaluators to build on previous findings. Evaluations are therefore also used strategically in the long run as issues raised in consecutive evaluations gain prominence in decision-making. Also, symbolic use of evaluations were found to play a role in the evaluation process as DG ENV as well as other DGs are concerned about reducing negative findings about their organizations in the evaluation.

Apart from these few instances of evaluation process use, there were no accounts of process uses in the data. There are two main reasons: first, the evaluation process is carried out mainly by the external evaluator in relative seclusion from potential users in the Commission or in other parts of the evaluation system. The evaluation process is typically managed by one desk officer in DG ENV, who is the liaison between the evaluator and the Steering Committee that oversees the evaluation at regular intervals (5-7 meetings during the evaluation process). This process is standard in the Commission and is meant to secure the independence of the evaluation as well as the proper and efficient evaluation execution. However, it also limits use in the evaluation process, because the potential users are rarely directly involved in the evaluation activity. Second, the evaluation findings can rarely be put directly to use during the evaluation process in DG ENV. Whether process findings are instrumental or symbolic, the use would normally require the evaluation to be finalized in order for them to be used for **instrumental use** and symbolically as well. DG ENV is expected to evaluate as it is stipulated in the LIFE Regulation. Flagging evaluation activity as a symbolic act during the evaluation process to gain external legitimacy in the system is therefore not necessary, as evaluation is expected by the organizational environment. One interviewee from the Commission put it this way: "*...the legitimacy [from*

evaluations] is by now almost automatic..." However, failing to evaluate or delaying evaluation activity would be perceived negatively by other actors in the EU evaluation system.

The LIFE evaluations analyzed in this article took place every 2-3 years. Therefore the **conceptual use** of LIFE evaluations related to methods and programme-related concepts was limited, as evaluation practice and findings were repetitive. The staff implementing the evaluation were also very knowledgeable about the LIFE programme and therefore there is also little **information use** during the evaluation implementation.

Findings use. LIFE evaluations are mainly used **conceptually** in the Commission, whereas policy-makers do not mention conceptual use in the interviews. Evaluations are used conceptually to inform discussions in DG ENV about evaluation methodology, programme indicators and the purpose of evaluation and programme impacts. The Commission's own ideas and opinions about the programme are clarified and sharpened by evaluations. This includes findings that might go against the conventional wisdom of the programme management such as the feasibility of indicator systems across the heterogeneous project portfolio of the LIFE programme.

Instrumental use is linked intrinsically to programme management in the two LIFE units in DG ENV. Instrumental use takes place to a large extent based on the evaluation's recommendations through follow-ups and most recommendations are addressed after evaluation implementation. Examples of instrumental use could be work optimisation or improved internal communication between the LIFE units and the thematic units in DG ENV. However, fundamental changes to the programme administration are rare because they require changes to the LIFE Regulation or the Commission's Financial Regulation. Also, the programme has been optimised over twenty years and thus several interviewees argue that most workable options have already been tried out.

On the other hand, the data suggests that policy-makers did not use LIFE evaluations instrumentally. That is, evaluations did not directly inform policy-making on the policy-level through, for example, the evaluation's recommendations. Policy-makers do not in general read the LIFE evaluations, so evaluations feed into the policy cycle through instrumental use on the programme

management level, where evaluations are the knowledge base for policy development in, for example, ex-ante evaluations and IAs. Thus, the effect of evaluations on policy is indirect. Additionally, evaluations are also used to update the evaluation system based on knowledge and experience linked to evaluation practices. This information is collected by the Commission's Secretariat General. As a result of changes to evaluation practice, a final evaluation is no longer a requirement in the new LIFE Regulation (2014-2020).

Information use was referred to by all interviewees. NGOs and potential beneficiaries as well as people new to the programme use evaluations to get an overview and understanding of the way the programme works. MEPs use the Commission's presentations but also the evaluation documents needed to get up to speed with the LIFE programme. When a decision has to be made regarding LIFE, they read the executive summary and browse or search keywords in the document. Also, MEPs and particularly the LIFE units in DG ENV use evaluations for presentations and to communicate about LIFE. Evaluations typically contain aggregated information and facts about the programme as well as graphs and figures, which can readily be used for presentation purposes. However, evaluations seldom contain information that is completely new to the programme management. Finally, evaluations are systematically referred to in evaluation tender material and so evaluation findings and recommendations are used by other evaluators in subsequent evaluations. Information use is linked closely to legitimizing and strategic uses. Using evaluations to legitimize the programme is a common practice, mentioned both by programme management and policy-makers.

Legitimizing use is most common in relation to informational practices such as when DG ENV is required to report evaluation findings or make presentations to the EP, the Council or the LIFE Committee (consisting of Member State representatives). The extent to which these activities are related to legitimizing use depends very much on the timing of the evaluation. Flagging that the Commission is doing a good job is important. One key informant in DG ENV described it in the following way: *"our evaluation reports and results are used and when we have them, it is something that people are happy about, because they prove that we in most cases are doing a good job."* Legitimizing use, however, is not the most common use type because many of the recommendations are programme-specific and not orientated towards legitimizing the programme on a political level.

Interviewees specified that the evaluation system carries legitimacy by default and therefore one LIFE evaluation does not add significantly to the legitimacy of the programme. Rather, it is the EU evaluation system that gives legitimacy to the LIFE programme and DG ENV. Finally, the LIFE programme is very popular among stakeholders and not very contested. Evaluations of larger EU expenditure programmes as well as policies may be considerably more contested resulting in more legitimizing evaluation use.

Strategic use of the LIFE evaluations coincided with legitimizing use, particularly when DG ENV presents evaluation results to the EP and the Committee of Regions in relation to the new LIFE programme proposal. In addition to this, DG ENV uses the evaluations as a reference document to raise issues and problems about the programme that they are aware of, but which are easier for them to communicate with an independent evaluation. This is done externally to the EP and Council but also internally to other units in the DG. Internally, other units in DG ENV needed to be convinced about certain directions of the programme. The Members of Parliament (MEPs) and Member States used the LIFE evaluations as a basis for arguments in the negotiation for the new LIFE programme and in questions to the Commission. The MEPs can base opinions on evaluation data or findings. Occasionally, lobbyists raise issues to MEPs stemming from an evaluation that the MEP might in turn take up with the Commission.

Summary of uses of LIFE evaluations. The main finding of the analysis of evaluation use is that process use almost never takes place. On the other hand, several instances of findings use were discovered. The data suggests that after the LIFE evaluations were implemented, instrumental- and information uses were the most common, followed by strategic use. Legitimizing and conceptual uses were mentioned less often by interviewees. On the level of programme managers, instrumental use happens to adjust the programme within the limitations of the LIFE Regulation and the Commission's Financial Regulation. Also, instances of legitimizing- and strategic uses were found in the data related to programme managers.

Regarding policy-makers, evaluations were used to acquire information (information use) and to advocate certain policy issues or positions (strategic use). Data suggests that programme evaluations are not used in a direct instrumental

way by policy-makers for the purpose of making significant programme changes. Rather, they are used indirectly to support IAs and policy fiches in the negotiations for the new programme at the end of each programme cycle. Evaluators, who have an informed outsider's look at things, mainly refer to information use and instrumental use of the evaluations.

Explaining evaluation use of LIFE evaluations

This section contains the analysis of factors explaining evaluation use observed from the LIFE evaluations. Table 4 gives an overview of the factors that explain evaluation use. The table summarizes the distribution of the qualitative codes for three groups of interviewees and provides examples of interviewee quotes. The findings from the data are described and summarized below.

Table 4. Explanatory factors.

		Total	Evaluator *	Program me managers **	Policy- makers***	Quotes
Decision and policy setting	Enabling	Count: 34 Row: 100%	13 38% (62%)	16 47% (48%)	5 15% (25%)	- "Yeah, I think, virtually all of them [in DG Environment] were happy and interested [in the evaluation]." (Evaluator)
	Impeding	Count: 40 Row: 100%	8 20% (38%)	17 43% (52%)	15 38% (75%)	- "...the interest is only with the LIFE units. The thematic units have no interest in the evaluation and they do not use evaluations." (Programme manager)
	Total		Count: 21 Column: (100%)	Count: 23 Column: (100%)	Count: 20 Column: (100%)	- "...I think the political interest is often not there. There has to be a political reason to, you know, to [use the evaluation]." (Policy-maker)
			Count: 21 Row: 100%	Count: 21 Column: (100%)	Count: 21 Column: (100%)	
Competing information	Enabling	Count: 21 Row: 100%	0 0% (0%)	12 57% (26%)	9 43% (38%)	- "...when the new LIFE programme was proposed, then of course they [the Commission] would present results [from evaluations]." (Policy-maker)
	Impeding	Count: 64 Row: 100%	14 22% (100%)	35 55% (74%)	15 23% (63%)	- "...it is frustrating that [the evaluator] gets a lot of good ideas that you already had previously and already have tried out." (Programme manager)
	Total		Count: 14 Column: (100%)	Count: 47 Column: (100%)	Count: 24 Column: (100%)	- "We only had [a few] years of extra stuff to look at compared to all the other [evaluations]. So... it was just the same story again, really." (Evaluator)
			Count: 14 Column: (100%)	Count: 47 Column: (100%)	Count: 24 Column: (100%)	- "...I believe that we have been evaluated so often that [the programme] cannot be done much different now." (Programme manager)
			Count: 14 Column: (100%)	Count: 47 Column: (100%)	Count: 24 Column: (100%)	- "There is a lot of information out there and you try to take in and digest as much as possible" (Policy-maker)
			Count: 14 Column: (100%)	Count: 47 Column: (100%)	Count: 24 Column: (100%)	- "[first we look at] ...the presentations that the Commission does and the fiches that we receive, then you go one level deeper and you look at the impact assessments yourself. And at the next level, you look at the evaluations." (Policy-maker)

	Total	Evaluator *	Program me managers **	Policy- makers***	Quotes
Decision characteristic s	Enablin g	Count: 34 Row: 100%	4 17% (36%)	3 13% (75%)	- "...if you see the impact assessment, in many cases we put ...the recommendations from the midterm, the ex-post or the ex-ante... so that you can see where this knowledge comes from [and so that you can see] that it is not really something that we have invented [and that] there is some background research behind it." (Programme manager)
	Impedi ng	Count: 40 Row: 100%	7 18% (64%)	1 3% (25%)	- "...the frustration, when you get a lot of new good ideas [from the evaluation], but at the same time you are bound by the reality of things, which makes it impossible to implement the ideas in practice. It is very frustrating in relation to e.g. Communication... [The recommendation was to] give it some more money to get much more out of it. But at the same time, they remove one of our staff." (Programme manager) - "...[The Commission is] going to say, 'Yes there is a problem with the payment speed. But the reason for that is because of the administrative procedures within the Commission... and there is nothing that we as DG environment can do about this', which is true..." (Evaluator) - "There were some other recommendations that couldn't be taken on board in LIFE+ because they needed an amendment of the Regulation." (Programme manager) - "We did not even have time to really do a follow-up [of the evaluation] because we are just really doing the next evaluations and ... [then] we had to start drafting the new Regulation." (Programme manager) - "I think [the evaluation] gets to that level [Director's level] mostly when they are discussing the new Regulation" (Programme manager)
	Total		Count: 11 Column: (100%)	Count: 4 Column: (100%)	
	Enablin g	Count: 31 Row: 100%	0 0% (0%)	9 29% (32%)	- "We always need knowledge. Believe me, we need knowledge." (Programme manager) - "Typically, the European Parliament would like to see reports almost every other day." (Programme manager)
Information needs	Impedi ng	Count: 38 Row: 100%	9 24% (100%)	19 50% (68%)	"...they say, ...you must take great heed of what it says in the previous evaluations because there is a lot of analysis and a lot of consultancy time being spent on working these things out." (Evaluator) - "...external evaluations... they confirm what you know." (Programme manager)
	Total		Count: 9 Column: (100%)	Count: 28 Column: (100%)	
	Enablin g	Count: 5 Row: 100%	1 20% (50%)	4 80% (80%)	- "...experienced desk officers know that this [the evaluation] is a way of airing their problems and what they want..." (Evaluator)
	Impedi ng	Count: 6 Row: 100%	1 16% (27%)	1 16% (27%)	"...they don't show up [to presentations of evaluations in the EP]... I look at myself and it happens too."
Personal characteristic s	Enablin g	Count: 5 Row: 100%	1 20% (50%)	4 80% (80%)	- "...experienced desk officers know that this [the evaluation] is a way of airing their problems and what they want..." (Evaluator)
	Impedi ng	Count: 6 Row: 100%	1 16% (27%)	1 16% (27%)	"...they don't show up [to presentations of evaluations in the EP]... I look at myself and it happens too."
	Total		Count: 9 Column: (100%)	Count: 28 Column: (100%)	
	Enablin g	Count: 31 Row: 100%	0 0% (0%)	9 29% (32%)	- "We always need knowledge. Believe me, we need knowledge." (Programme manager) - "Typically, the European Parliament would like to see reports almost every other day." (Programme manager)

	Total	Evaluator *	Program me managers **	Policy- maker***	Quotes
	ng	Row: 100%	17% (20%)	67% (100%)	(Policy-maker)
	Total	Count: 2 Column: (100%)	Count: 5 Column: (100%)	Count: 4 Column: (100%)	-
Political climate	Enablin g	Count: 6 Row: 100%	1 17% (17%)	0 0% (0%)	-
	Impedi ng	Count: 13 Row: 100%	5 46% (55%)	2 15% (100%)	- "It is not policy. Let us say it in this way; it is a programme. It is well-managed and therefore [there is only little interest on the political level]. (Programme manager)
	Total	Count: 11 Column: (100%)	Count: 6 Column: (100%)	Count: 2 Column: (100%)	-
Evaluation implementation					
Timeliness	Enablin g	Count: 1 Row: 100%	1 100% (100%)	0 0% (0%)	- "The consultants were also very timely in the execution of the evaluation." (Programme manager)
	Impedi ng	Count: 1 Row: 100%	0 0% (0%)	1 100% (100%)	-
	Total	Count: 0 Column: (0%)	Count: 1 Column: (100%)	Count: 1 Column: (100%)	
Credibility	Enablin g	Count: 13 Row: 100%	9 69% (56%)	4 31% (100%)	- "I expect them [evaluations] to be the most sound and the most neutral source of information." (Policy-maker)
	Impedi ng	Count: 7 Row: 100%	7 100% (44%)	0 0% (0%)	- "Some say that the dog does not bite the feeding hand" (Said in relation to the relationship between the Commission and the Consultants) (Key Commission informant)
	Total	Count: 0 Column: (0%)	Count: 16 Column: (100%)	Count: 4 Column: (100%)	

		Total	Evaluator *	Program me managers **	Policy- makers***	Quotes
Evaluation quality	Enablin g	Count: 19 Row: 100%	2 11% (67%)	16 84% (89%)	1 5% (100%)	"It was a solid piece of work." (Programme manager)
	Impedi ng	Count: 3 Row: 100%	1 33% (33%)	2 67% (11%)	0 0% (0%)	"...they [the consultants] provided good examples, but we needed more down to earth data and that was a bit missing in the report..." (Programme manager)
	Total		Count: 3 Column: (100%)	Count: 18 Column: (100%)	Count: 1 Column: (100%)	
Findings	Enablin g	Count: 9 Row: 100%	4 44% (50%)	0 0% (0%)	5 56% (56%)	"[The Commission] had not done [an analysis] like that before." (Evaluator)
	Impedi ng	Count: 15 Row: 100%	4 27% (50%)	7 47% (100%)	4 27% (44%)	"...[the evaluation] is just sort of confirming stuff which was in [previous evaluations] really." (Evaluator)
	Total		Count: 8 Column: (100%)	Count: 7 Column: (100%)	Count: 9 Column: (100%)	
Relevance	Enablin g	Count: 6 Row: 100%	0 0% (0%)	6 100% (22%)	0 0% (0%)	"I think [the evaluation] gets to that level [level of directors] mostly when they are discussing the new regulation." (Programme manager)
	Impedi ng	Count: 36 Row: 100%	3 8% (100%)	21 58% (78%)	12 33% (100%)	- "[Compared to evaluations], it's much better to have technical studies that you do when things are ripe." (Programme manager) - "The Commission reports when they have reporting obligations ...and we get these reports and ...it's not always necessarily interesting for us [in the EP] because it can be very technical, very routine things about procedural things and how things are being sort of processed and taken forward." (Policy-maker) - "I don't think we can gain much from it [the evaluation]. I don't think it's much used, but it is from a requirement so... And even from the accountability point of view, I don't think it's the best too, but it is there because of that and we have to do it." (Programme manager) - "We did not really see the added value of having a final evaluation just because it is a requirement of the Regulation." (Programme manager)
	Total		Count: 3	Count: 27	Count: 12	

		Total	Evaluator *	Program me managers **	Policy- maker***	Quotes
			Column: (100%)	Column: (100%)	Column: (100%)	
Communicati on quality	Enablin g	Count: 3 Row: 100%	0 0% (0%)	3 100% (60%)	0 0% (0%)	-
	Impedi ng	Count: 2 Row: 100%	0 0% (0%)	2 100% (40%)	0 0% (0%)	-
	Total		Count: 0 Column: (0%)	Count: 5 Column: (100%)	Count: 0 Column: (0%)	-

The table displays the number of coding instances in each explanatory category and in three interviewee categories. The count represents an interviewee's mention of an explanatory factor. In bold: the row percentage. In parenthesis: the column percentage. * Consultants. ** DG ENF Staff. *** EP- and Council- representatives.

Decision and policy setting. Regarding the explanatory factors relating to *Decision and policy setting*, the EU evaluation system has institutionalized a high level of **receptiveness and commitment** to evaluation in the Commission and also in DG ENV. Evaluation practice is considered part of the administrative practice in DG ENV including obligations towards the other organizations in the system, which again creates commitment. Overall, the interviews and desk research point to an organizational willingness to use evaluations. However, receptiveness in relation to LIFE programme evaluations is largely limited to the desk officers involved in programme management within the Commission. Interest and knowledge of LIFE evaluations in the EP and Council is very low, mainly due to low relevance of evaluation information to the work of policy-makers.

There is a lot of **competing information** regarding LIFE evaluations. This is one factor that impedes evaluation use both for programme managers and policy-makers. Competing information primarily includes previous evaluations of LIFE (either produced by the Commission or, for example, by the European Court of Auditors). But the experience that programme managers have managing the LIFE programme should also be considered as competing information. It is simply very difficult for consultants to bring about new information, new findings or produce knowledge that the experienced programme staff do not already have. Finally, evaluations are relatively broad in scope, covering the entire programme. In terms of use, specific studies about parts of the programme are more focused on key contested areas, which according to interviewees make them more likely to be used than the broadly-scoped evaluations. At the same time, however, interviewees consider LIFE evaluations to be the best general source of assessment information about the LIFE programme.

According to many interviewees, the **decision characteristics** of the EU institutions do to a large extent impede direct evaluation use, particularly on the level of programme management. The seven-year budget period in the EU makes substantial and incremental programme changes difficult. Also, political decisions about the programme are complex and depend on several political actors (EP and Council) potentially with very diverse political interests and priorities.

Nevertheless, evaluation use is an integrated part of the evaluation system every seventh year when the new programme is being prepared. An IA must draw on available information, including evaluations, and the importance of the budget decision draws attention to all available information on LIFE. Thus the EU policy cycle – which the evaluation system is designed to feed information - both enables and impedes evaluation use.

In relation to evaluations of the LIFE programme, the programme management also has **information needs**, despite the fact that the programme management is by far the most knowledgeable in relation to the programme. Nevertheless, some specific parts of the programme might be unknown even to the programme management, in which case an evaluation can be used to shed light on the issue. In the EP and the Council, the need for evaluation information is much lower, mainly because policy-makers need information that communicates key information about issues related to decision-making. This is rarely considered to be the case for programme evaluations of the LIFE programme.

Interviewees did not often mention **personal characteristics** in relation to evaluation use. DG ENV has substantial experience with evaluation over several decades. As the practice of evaluation is highly institutionalized, the LIFE evaluations were carried out and supervised according to the Commission's general evaluation policy and standards. Both the Commission and the external evaluators had sufficient capacity to manage and carry out the evaluations. The data suggests that the institutionalized practice of the evaluation system has a positive effect on evaluation use because staff are trained and used to working with evaluations, which in turn influences the quality of the evaluations positively.

Few interviewees mentioned the **political climate** as an explanatory factor in relation to evaluation use. Evaluations are more likely to be used to justify positions, if the programme or an issue is contested politically (Cousins and Leithwood, 1986). But as the LIFE programme is an overall popular and well-run programme, the political climate is not important for the use of evaluation findings in this case. As long as evaluations are not very negative, evaluations are used mainly as information in the IAs. Hence, evaluations are not brought into policy-making directly but only as secondary information after IAs and fiches in the programming of the new LIFE programme.

Evaluation Implementation. Considering the explanatory factors linked to *Evaluation implementation*, the interviewees mentioned these factors considerably less than the decision and policy setting. Very few interviewees mentioned **timeliness** as an important factor. Similar to the personal characteristics, the evaluation system produces evaluations with a high degree of professionalism and timeliness. Evaluation management in the Commission is supervised by the evaluation unit and the capacity to manage evaluations is highly developed in the organization. At the same time, evaluators are carefully selected through tendering. Due to the institutionalization of evaluation practice in the evaluation system, timeliness is less of an issue. However, timeliness is important for the evaluation system because of the policy cycle, and therefore securing timely delivery of evaluations has a positive impact on evaluation use.

The **credibility** of evaluations is considered crucial for the use of evaluations in the Commission. To several of the interviewees in DG ENV, the credibility of Commission evaluations is related to the credibility of the Commission itself as well as the high quality of evaluation work and independence of evaluators. The high credibility of Commission evaluations enables the justificatory uses of the LIFE evaluations towards the EP and Council, particularly in the renegotiation of a new programme. This is induced by the way the evaluation system is constructed, with a stress on external evaluation and thus a relative independence in the evaluation implementation.

The situation is similar in relation to **evaluation quality**. The evaluation quality is perceived by interviewees to be high and is also considered important for evaluation use. Contrary to the enabling factor played by the evaluation quality in the evaluation system, interviewees see evaluation **findings** as constrained by the Commission's evaluation guidelines and standards. The Commission's evaluation tenders include evaluation questions as well as a relatively rigid format for the evaluation process including a specific number of meetings with the Steering Committee. Some interviewees argue that this procedural format leaves little room for innovation during the evaluation process as well as in evaluation findings because both the process and the findings are scoped and framed relatively narrowly by the tender material which is more or less standardized in the common evaluation guidelines. In turn, this impedes evaluation use as evaluators are constrained in the process and findings seldom are surprising or completely

innovative to the Commission. The lack of innovative recommendations is also due to knowledge-asymmetry between the evaluator and programme management. This is linked to the high complexity of the LIFE programme and the decision characteristics mentioned earlier, which often render innovative recommendations useless due to legal constraints on the Commission that can render innovative recommendations useless.

The **relevance** of evaluations is considered by interviewees to be very important for evaluation use of the LIFE programme. In the EU evaluation system, the relevance of evaluations is closely linked to the timing of the evaluation relative to decision-situations in which the evaluation can be used. That is crucial both on the political level as well as the administrative level in the Commission. If the timing of the evaluation does not match the programme cycle, then it can have several implications. First, if the evaluation is too early, there is little new data available and the evaluation will be less usable compared to the previous evaluation (partly the case in the 2010 midterm evaluation). If the evaluation is too late, then the evaluation findings are of little or no use (the case in the 2012 final evaluation). Also, an evaluation can be timed too close to other evaluations and thus leave no time for follow-up before the next evaluation starts. Mistiming of evaluations relative to decisions or other evaluations is an impeding factor relative to evaluation use.

LIFE evaluations were not perceived to be relevant directly for policy-making as the information contained in evaluations is too general and relates exclusively to programme implementation and less to the overall political rationales behind the programme. Also, evaluations are considered too technical and detailed by policy-makers. Evaluations are therefore not directly relevant for policy-makers, except when they serve as information feeding into the Commission's preparations for a new programme at the end of every programme cycle.

Finally, the data suggests that the standard evaluation implementation process of the Commission also secures a high **communication quality** of evaluations often through an iterative work process involving several parallel quality checks in the final phases of the evaluation implementation.

Summary of factors explaining evaluation use in LIFE evaluations. The results of the data analysis show that decision and policy settings are far more important

than factors related to evaluation implementation. In relation to the decision and policy setting, the most prevalent impeding explanatory factors on evaluation use are competing information and decision characteristics. Particularly for policy-makers, there is a lot of information and very little time to digest it. For programme management, both are major obstacles to the use of evaluation findings and recommendations and substantial changes to the programme are difficult to make. Also the political climate is impeding evaluation in this case, as the LIFE programme is a small and not very contested, hence the Commission's need to prove accountable and the EP's urge to check the Commission are reduced relative to other programmes.

At the same time, explanatory factors related to decision and policy settings also enable evaluation use. The personal characteristics are enabling use, as evaluation training as well as procurement procedures make the right people available for the job both internally and externally. Evaluation use is also enabled by high receptiveness and commitment to evaluation in the Commission. The advantages of the system's institutionalized practice is mainly relevant for the Commission and less so for other organizations in the system such as the EP and Council. However, all actors in the system share the need for information, which is a major enabling factor both in the Commission as well as outside the Commission.

Table 5 provides a summary of the explanatory factors' effect on use as they are reported by interviewees.

Table 5 *Overview of the effect of the explanatory factors.*

Explanatory factors	Enabling/impeding
Decision and policy settings	
Commitment and receptiveness	Enabling
Competing information	Impeding
Decision characteristics	Impeding
Information needs	Enabling
Personal characteristics	Enabling
Political climate	Impeding
Evaluation implementation	
Timeliness	Enabling
Credibility	Enabling
Evaluation quality	Enabling
Findings	Impeding
Relevance	Impeding
Communication quality	Enabling

In relation to evaluation implementation, the main findings of the analysis are that the evaluation system reduces the most negative effects of several of the explanatory factors linked to evaluation implementation. Hence, timeliness, credibility, evaluation quality and communication quality are all factors that are supported by routinized and systematized institutions of practice in the Commission. Also, these advantages of the system's institutionalized practices are mainly relevant for the Commission and less so for other organizations in the EU system such as the EP and Council.

As just explained, the institutionalized practices of the evaluation system enable evaluation use. However, the rigidity of the system's practices also impedes evaluation use by controlling and thereby reducing the innovation and possible outcome of evaluation findings. The relevance of evaluations is limited to policy-makers in terms of content and to programme managers because of evaluation timing, which is not always optimal for evaluation use. The relevance of evaluations is limited to policy-makers in terms of content and to programme managers because of the timing. Evaluation findings and relevance are the impeding factors related to the evaluation implementation.

Users of LIFE evaluations

The main users of LIFE programme evaluations are the programme management in DG ENV. Evaluation use by other actors depends mainly on the relevance (timing relative to a decision) and the demand for information. Generally, evaluations are not used directly for policy making outside the Commission; i.e. the programme management will use evaluations to prepare the new LIFE programme and adjust its implementation. Thus LIFE programme evaluations are supporting documents in the programme proposals that decision-makers outside the Commission read and use every seventh year in the policy cycle. Therefore evaluation use for policy making is mainly indirect outside the Commission. Rather, evaluations outside the Commission are used for information purposes and sometimes for advocacy (strategic use) by for example NGOs.

Table 6 summarizes the findings from the above sections including the explanatory factors and use types that were mentioned most by interviewees. Explanatory factors not included in the table were those perceived to be less significant by interviewees. These include evaluation quality, timeliness, credibility, personal characteristics etc.

Table 6 *Summary of most mentioned evaluation use types and explanatory factors per user group.*

	Programme management	Policy-makers
Most relevant explanatory factors		
Enabling	<ul style="list-style-type: none"> - <i>Commitment and receptiveness</i> - <i>Information need</i> 	<ul style="list-style-type: none"> - <i>Information need</i>
Impeding	<ul style="list-style-type: none"> - <i>Decision characteristics</i> - <i>Findings</i> - <i>Relevance</i> 	<ul style="list-style-type: none"> - <i>Competing information</i> - <i>Commitment and receptiveness</i> - <i>Relevance</i>
Most common use types		
Process use	<i>NA</i>	<i>NA</i>
Findings use	<ul style="list-style-type: none"> - <i>Instrumental</i> - <i>Legitimizing use</i> - <i>Strategic use</i> 	<ul style="list-style-type: none"> - <i>Information use</i> - <i>Strategic use</i>

Commitment and receptiveness to programme evaluations is an enabling factor mostly relevant to the programme management, as evaluations are too detailed and not directly relevant for policy-makers and other stakeholders. However, most stakeholders to the LIFE programme express a need for knowledge. Information need is thus an enabling factor for most evaluation users.

Regarding factors impeding evaluation use, interviewees mentioned mostly decision characteristics, relevance and findings and competing information. For programme management it is the decision characteristics of the programme cycle and the EU decision-making structure that impede evaluation use. Given that programme managers are the most knowledgeable with regard to the programme, institutionalized evaluation practices limit the relevance and innovation of

findings, which could induce evaluation use. For policy-makers overload of information, lack of receptiveness to evaluation and the low relevance of evaluations are the main obstacles to evaluation use.

Conclusion

The aim of this article is to find answers to how, who and why – or why not – an evaluation system affects the utilization of evaluations. Throughout the analysis it has been demonstrated that in the case of the LIFE programme, the EU evaluation system is conducive to evaluation use, while at the same time also impeding use in several ways. The article finds that the evaluation system is conducive to instrumental-, strategic- and legitimizing types of evaluation use on the level of programme management. On the level of policy-makers, the evaluation system is conducive to strategic- and information types of uses. Regarding the users of evaluation findings, the instrumental use of LIFE evaluations is linked intrinsically with programme management in the two LIFE units in DG ENV while being far less used by policy-makers in the EP and the Council. While the system does enable uses, it also impedes process use of evaluations during the evaluation implementation.

With that in mind, the decision characteristics are key in order to understand the use of the LIFE evaluations and in particular the absence of process use. Decision characteristics are contextual relating to the legal- and organizational setup as well as the formal and informal practices of the evaluation system. The Commission decision characteristics largely determine the effect that other explanatory factors have on evaluation use. In particular, the alignment of programme evaluation to the Commission's budget- and policy cycle has several consequences for evaluation use. These consequences of Commission decision characteristics are elaborated in the five main findings below.

Findings use over process use.

The EU evaluation system is primarily designed to feed information into the EU decision-making procedure every seventh year before the beginning of a new policy cycle. Commission staff working with programme evaluation are managing the evaluations with the main objective of satisfying the evaluation obligation by securing the timeliness, quality and independence of the final output. To that end, the evaluation standards including the guidelines and terms of reference specify in

great detail how the evaluation process should be executed in such a way that process use - as envisaged by for example Michael Patton (1997) - is not enabled. As will be elaborated below, the limited process use is a consequence of choices made deliberately to secure and improve findings use for decision-making, in particular 'decision points'. The loss of process use is thus a direct consequence of policy-making practices in the EU political system as a whole as well as the work practices decided for and by the Commission.

Findings use in 'decision-points'.

The policy-cycle's decision-points enable findings use because the evaluation system is designed to deliver feedback into a particular decision-point at the time for programme renegotiating. In the case of the LIFE programme, this decision-point enables findings uses including instrumental-, legitimizing and strategic uses. These types of findings use are enabled because the decision point allows for potential programme change and because the Commission needs programme legitimacy and overall accountability when the focus of policy-makers in the EP and Council is on the programme. It is at these points in time that LIFE evaluations will be used indirectly (through for example IAs) to change the programme. The evaluation system is designed to deliver independent quality evaluations that are timely, well communicated, independent and credible in order to secure the legitimacy of the programme and the strategic position of the Commission in such negotiations. The institutionalization of evaluation practices including the highly standardized evaluation process, staff training and guidelines almost completely negates process uses as explained previously. However, they also enable a smoother execution of evaluation processes and ultimately deliver a more standardized evaluation product with minimal 'risks', as potential negative influences from personal characteristics of evaluators and staff, organizational deficiencies etc. are largely avoided.

Low evaluation relevance regarding significant programme change.

Due to mistiming and competing information, LIFE evaluations are seldom relevant to potential users in decision-making. The relevance of evaluations is affected negatively and directly by decision characteristics. Sometimes, evaluations are mistimed in relation to decision events such as the preparation of an IA or the proposal for a new programme. Further, the evaluation system

generates competing information, including first and foremost many evaluations on the LIFE programme as well as other studies and audits. Altogether, this reduces the relevance of evaluations and impedes the instrumental use of them. Also, the legal structures that govern the Commission's work (Financial Regulation etc.) reduce the possibilities of innovative and surprising recommendations because consultants' recommendations are steered first of all by the evaluation questions in the tender material that is typically in accordance with what the Commission can influence within the scope of its legal competences. For example would potential improvements to the LIFE Regulation be downplayed relative to recommendations that could readily be applied by the programme units in DG ENV.

The 'de-politicization' of programme evaluations.

Following the three previous points, instrumental evaluation use is mainly limited to managers of the LIFE programme. Also, evaluations are rarely used in evaluation implementation and only rarely used directly for policy-making (for a similar finding see Laats, 2005). Due to mistiming of evaluations, other competing information and non-innovative findings, evaluations' relevance for decision-making is very limited. These two findings imply a de facto de-politicization of programme evaluations in the EU evaluation system, where evaluation information conforms to the administrative context of programme management in the Commission instead of the political context of policy-makers.

The paradoxical evaluation system.

The 'de-politization' represents a paradox since it is the EP and the EU Member States that compelled the Commission to develop its evaluation practices in the first place. But this article shows that policy-makers rarely use the evaluations while at the same time the Commission also does not maximize utility of evaluations. Also, the Commission allocates considerable resources to evaluations and numerous other internal and external assessments and audits. This paradox is probably explained best by turning to the complex nature of the inter-institutional context of the EU evaluation system and EU evaluations (see for example Dahler-Larsen, 2012; Højlund, 2014b). To do that would be too much to cover at this point and more evidence is also needed to support such an analysis. But hopefully this article will spawn more interest in other aspects of the EU's evaluation system,

such as the birth of systematic policy-evaluations (Stern, 2009). These often-contested and highly political evaluations have other attributes than the Commission's programme evaluations and therefore they are also likely to be used in different ways than we have seen is the case with the Commission's programme evaluations.

References

- Alkin M and Taut S. (2003) Unbundling Evaluation Use. *Studies in Educational Evaluation* 29: 1-12.
- Alkin MC and Stecher B. (1983) Evaluation in Context: Information Use in Elementary School Decision Making. *Studies in Educational Evaluation* 9: 23-32.
- Bazeley P. (2013) *Qualitative Data Analysis with NVIVO*, London: Sage Publications Ltd.
- Bienias S and Iwona L. (2009) Evaluation Systems in the Visegrad Member States. Warsaw: Polish Ministry of Regional Development.
- Commission E. (2007) Evalueringsaktiviteter & -resultater på tværs af Europa-Kommissionens tjenestegrene. In: Jacobsen S (ed) *Dansk Evalueringsselskabs konference*. Kolding, Denmark.
- Cousins JB and Leithwood KA. (1986) Current empirical research on evaluation utilization. *Review of Educational Research* 56: 331-364.
- Dahler-Larsen P. (2012) *The Evaluation Society*, Stanford, Ca.: Stanford University Press.
- Easterby-Smith M, Crossan M and Nicolini D. (2000) Organizational Learning: Debates Past, Present and Future. *Journal of Management Studies* 37: 783-796.
- Finne H, Levine M and Nilssen T. (1995) Trailing research: A model for useful program evaluation. *Evaluation* 1: 11-31.
- Floden RE and Weiner SS. (1978) Rationality to Ritual: The Multiple Roles of Evaluation in Governmental Processes. *Policy Sciences* 9: 9-18.

- Furubo J-E. (2006) Why Evaluations Sometimes Can't be Used - and Why They Shouldn't. In: Rist R and Stame N (eds) *From Studies to Streams*. New Jersey: Transaction Publishers, 147-165.
- Furubo J-E, Rist RC and Sandahl R. (2002) *International Atlas of Evaluation*, London: Transaction Publishers.
- Henry GT and Mark MM. (2003) Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation* 24: 293.
- Højlund S. (2014) Evaluation use in the organisational context – Changing focus to improve theory. *Evaluation* 20: 26-43.
- Imam I, LaGoy A and Williams B. (2007) Introduction. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes: EdgePress of Inverness.
- Johnson K, Greenseid LO, Toal SA, et al. (2009) Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation* 30: 377-410.
- Kirkhart KE. (2000) Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation* 88: 5-23.
- Kohlbacher F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research* 7.
- Laat Bd. (2005) Study on the Use of Evaluation Results in the Commission. In: Laat Bd (ed). Paris: Technopolis.
- Langley A. (1998) In Search of Rationality - The Purpose Behind the Use of Formal Analysis in Organizations. In: Maanen JV (ed) *Qualitative Studies of Organizations*. Thousand Oaks: SAGE Publications, 51-90.
- Lazarsfeld PF, Sewell WH and Wilensky HL. (1967) *The uses of Sociology*, New York: Basic Books.
- Leeuw FL and Furubo J-E. (2008) Evaluation Systems : What Are They and Why Study Them? *Evaluation* 14: 157-169.
- Leviton LC. (2003) Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation* 24: 525-535.
- Leviton LC and Hughes EFX. (1981) Research on the utilization of evaluations: A review and synthesis. *Evaluation Review* 5: 525-548.
- Mayring P. (2000) Qualitative Content Analysis. *Forum: Qualitative Social Research* 1.

- Mayring P. (2004) Qualitative Content Analysis. In: Flick U, Von Kardorff E and Steinke I (eds) *A Companion to Qualitative Research*. London: SAGE, 266-269.
- Patton MQ. (1997) *Utilization-focused evaluation: The new century text*, Thousand Oaks: Sage.
- Pollitt C, Girre X, Lonsdale J, et al. (1999) *Performance Audit and Public Management in Five Countries* Oxford and New York: Oxford University Press.
- Pröpper IMAM. (1987) Beleidsevaluatie als argumentatie. *Beleidswetenschap* 2.
- Rich RF. (1977) Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding. In: Weiss CH (ed) *using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Rist R and Stame N. (2006) *From Studies to Streams: Managing Evaluative Systems*, London: Transaction Publishers.
- Ritchie J, Lewis J and Elam G. (2003) Designing and Selecting Samples. In: Ritchie J and Lewis J (eds) *Qualitative Research Practice - A Guide for Social Science Students and Researchers*. London: SAGE.
- Schreier M. (2012) *Qualitative Content Analysis in Practice*, Thousand Oaks: SAGE Publications Ltd.
- Stern E. (2009) *Evaluation policy in the European Union and its institutions*.
- 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.
- Weiss CH. (1992) *Organizations for Policy Analysis - Helping Government Think*, Newbury Park, CA: SAGE.
- Weiss CH. (1998) Have We Learned Anything New About the Use of Evaluation. *American Journal of Evaluation* 19: 21-33.
- Weiss CH and Bucuvalas M. (1980) *Social Science Research and Decision Making*, New York: Columbia University Press.
- Williams B and Imam I. (2008) Evaluation Systems: What Are They and Why Study Them? *Evaluation* 14: 157-169.

- Williams K, Laet B and Stern E. (2002) *The Use of Evaluation in the Commission Services*. Paris: Technopolis.
- Zito A and Schout A. (2009) Learning theory reconsidered: EU integration theories and learning. *Journal of European Public Policy* 16: 1103-1123.

8.4 ARTICLE 4 – EVALUATION AND POLICY LEARNING – THE LEARNERS’ PERSPECTIVE

Reference:

Borrás, S. and Højlund, S. (2014) *Evaluation and Policy Learning – The Learners’ Perspective*. European Journal of Political Research. 53(5)

<http://onlinelibrary.wiley.com/doi/10.1111/1475-6765.12076/abstract>

(To avoid any copyright issues in relation to the final print of this thesis, this article is conveyed in its ‘pre-script’ version. A pre-script version is a near-published version of the article as it was after peer-review and language check by the author as well as after the final acceptance of the article by the journal editor. The pre-script version does not include the final editing and formatting of the article by the journal in which the article was published. The author of this thesis recommends readers of this thesis published by the Library of Copenhagen Business School to seek and purchase this article in its original version at the journal in which it was published.)

EVALUATION AND POLICY LEARNING – THE LEARNERS’ PERSPECTIVE

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 what different types of learners actually learn within the context of the evaluation framework (the set of administrative structures defining the evaluation goals and process). Taking the empirical case of three most-similar 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 (program overview, small-scale programme adjustments, policy change, and evaluation methods), and that different learners are in control of different aspects of the evaluation (learning objectives and processes) according to the evaluation framework established by the European Commission.

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: 3). This paper looks at the second dimension of evaluation, examining how it induces policy learning⁴. '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 this 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 learners and learning. Even if some recent literature makes approaches to different types of learning, 'learners' are typically considered in a single category, without distinguishing among different types of learners and how these learners might be related to different types of learning. In this paper, learning is defined as the adaptation of beliefs and views that result from the learners' sense-making of past experience

⁴ This paper uses the term "policy learning". See section 2 for the definition and analytical levels of policy 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. This paper approach on different types of learners aims to contribute to the literature on policy learning, establishing an unexplored link between the types of learners and the types of policy learning. Furthermore, 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, it is important to note that the evaluation framework is far from automatically determining the actual learning (of 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 learning in different ways. This means that the way in which the learners learn in a context of evaluation, and the types of policy learning that emanate from this is an open question for both the evaluation literature as well as the policy learning literature.

Empirically, the paper takes the case of the European Commission and studies what is learned from three midterm evaluations of three most-similar EU expenditure programs. With a learners’ perspective, the paper ascertains what learners actually learn from the evaluation. 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. The study also finds that these different forms of learning is related to the fact that different learners are in control of different aspects of the evaluation (learning objectives and processes) according to the evaluation framework established by the European Commission.

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.

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, 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 embeddedness and actors' interactions in the learning process provides the background for this paper, which builds from the few evaluation studies that look into learning, and from the few policy learning studies that look into learners. From the literature on evaluation, the approach of this paper is based on two general assumptions: 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: 209) 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: 212). The second general assumption from the evaluation literature that this paper follows 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: 250). 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.

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: 3), 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 according to Bennett and Howlett, 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. In their recent seminar work systematising the literature, Dunlop and Radaelli identify four generic types of policy learning based on two crucial dimensions, namely, the tractability of the problem, and the degree of certification of the actors involved in the learning (Dunlop and Radaelli, 2013). In the formalized context of evaluation that interests us here, the certification of the actors involved is very high, because the evaluation of governmental interventions usually takes place within a pre-defined evaluation framework that identifies formally the actors that have the authority to act. Likewise, the tractability of the problem tends to be high as well. The typical problem that evaluations of governmental interventions aim to address is to improve the governmental intervention in question, which is a well-defined and tractable problem. From this perspective, the policy learning that results from the evaluation of governmental intervention (our current object of study) falls under the generic category of "shadow of hierarchy" in the typology of Dunlop and Radaelli. Taking a step further, these authors identify some more specific types of learning within this generic 'shadow of hierarchy' learning. This specific variation is based on the work of Mocker and Spear (1982), and depends on the different degrees of control that learners exercise over the objectives, and over the means and process. In other words, even in contexts where the problems are highly tractable and the actors are highly certified, there might be some important differences in the learning

outcomes, according to the role that learners assume in the definition of the content and the process. More concretely, Dunlop and Radaelli suggest the following specific types of learning in contexts of “shadow of hierarchy”:

Figure 1: Specific typology of learning under conditions of high tractability of problems and high certification of actors (learning under the shadow of hierarchy)

		Learners' control over learning objectives	
Learners' control over learning means		High	Low
	High	Autonomy	Instrumental
	Low	Delegation	Hetero-directed

Source: Dunlop and Radaelli (2013: 612)

This specific typology will serve to locate analytically the empirical cases of this paper (read below). Whereas the hetero-directed is a learning type that is traditionally associated to a steering mode of learning, most of the learning in highly actor certified and high problem tractability contexts is far more flexible, possibly attaining a large degree of autonomy in situations where learners determine the goals and processes of learning. This remains, however, an empirical question to be examined in this particular context of the policy learning that emanates from policy evaluation.

The reader must keep in mind that in this paper we argue further that the possible differences of learning types are related to who the learners are, because in evaluation exercises there is a number of quite different actors taking part in defining goals and processes, and hence quite different learners according to their role in the evaluation. Likewise, this paper argues that these differences are also related to how these actors interpret the evaluation framework (the formal set of rules) within which they operate. The argumentative-interpretative approach used here does not assume any form of automatism in the forms of learning just because the context is related to a well-defined problem/high actor certification in evaluation exercises. The point is that there might be a substantial difference in the learning that emanates from evaluations according to other relevant dimensions mentioned above, not least the differentiated nature of the actors involved in

evaluations. The next section sets up a more specific analytical framework for the empirical study that will follow in the subsequent sections of this paper, when answering the question, who learns what and how from evaluations?

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 first type is the "Program Unit", which is the administrative unit responsible for the evaluated governmental intervention. In a sense, the program unit can be seen as the "evaluated". A second 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. The third type of actor is the broader set of social and political actors that we define here as the "External Stakeholders". They do not necessarily know about the intervention in detail or the program unit, but 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. However, at times, they might become part of the evaluating organization during the evaluation process through contractual measures (e.g. consultants) or as part of the Steering Committee. Last but not least, the fourth type of actor is the "External Evaluators", which are the 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

2 may potentially learn. Identifying who actually does learn is an open empirical question.

Figure 2: 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 the learning typology 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. Thus, the evaluation framework sets the organizational, analytical and legal boundaries of the evaluation. 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, 2014b; 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, 2014a; 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 have based their specific typology upon (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. This is to say that, 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. 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.

Designing our Empirical Study

Our empirical study is based on the selection of three retrospective program evaluations carried out by the European Commission at approximately the same time. All three evaluation cases have similar types of actors involved (see Figure 2). These three case evaluations were all produced by external evaluators, they all

used the same evaluation framework, they were of similar budget sizes for the program cycle 2007-2013, and they work in similar ways through project support. Hence, the cases were chosen on the basis of their similarity, rather than being extreme, deviant or diverse cases.

This case study selection technique was chosen for two main reasons. Firstly, because it provides a broad empirical evidence, with more possible generalizability across similar cases. Most similar cases provide detailed account on relevant issues, and a broader empirical material, which is one good basis for replicability. Secondly, the aim of our paper is not to determine the differences across cases (which have similar budget, similar time-line, similar evaluation framework, etc), but across the different types of learners. All three cases have the same typology of learners. For this reason we treat the material from these three most-similar cases in a way that allows having a large basis for looking into how different types of actors learn differently. Using three most-similar rather than just one single case study gives us a broader material to define if, how and what different actors learn differently.

Regarding the choice of these particular three cases, we have chosen mid-term evaluations of the European Commission because the evaluation framework and learning from evaluations in the Commission have not been empirically studied before (see for example Boswell, 2008; Zito and Schout, 2009; Radaelli and Dunlop, 2013; Torres and Preskill, 2001). Moreover these three Commission cases are interesting because the European Commission has played an instrumental role in proliferating evaluation practices and institutionalizing program evaluation practices in Europe over the last three decades (Furubo et al., 2002; Toulemonde, 2000).

The first case is the midterm evaluation of the Program 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 Program (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 the European Parliament), 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.

When studying learning phenomena, the choice of methodology usually is either observation of learning outcomes (e.g. through tests of comprehension) or inquiring the learning subject. We chose the latter option because of obvious feasibility problems related to the former. However, inquiring the learning subject has also several drawbacks. For example it is likely that learners exaggerate learning in hindsight because learning often is intangible and because non-learning is shameful. We have tried to mitigate this possible bias by including various types of stakeholders in the sample and thereby triangulating from different interviewee sources as well as documents.

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 program 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 qualitative content analysis approach in Table 2 to illustrate the patterns in the data of the three cases.

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, 2014a). 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 1 Overview of the respondents' Indications of learning and no learning

What is being learned	# Interviewees	# Reference	Examples of coded quotes	Typical learner
Learning as 'overview'	15	28	<ul style="list-style-type: none"> - "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	<ul style="list-style-type: none"> - "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	<ul style="list-style-type: none"> - 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 	Program unit

			<i>designed, so that we can have a better program in the next seven years?" (Deputy Head of program Unit)</i>	
Learning about 'methodology, and evaluation	8	11	<ul style="list-style-type: none"> - "[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	<ul style="list-style-type: none"> - "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 seem to relate learning to an 'overview' of the program, when asked the question about what they did not have before the evaluation. Both program units and external 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, external 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 between the midterm evaluations and the ex-ante and impact assessment (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.

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 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 application 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.

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.

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 Head of Program Unit ...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, 2014a). 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, 2014a).

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.

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 (Mocker and Spears, 1982; Dunlop and Radaelli, 2013) and our focus on the evaluation framework where actors are highly certified and the problems are highly tractable, we have focused on the extent to which the learners were in control of 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 over 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. Turning back to the typology shown above in figure 1 of this paper, our empirical findings show that external evaluators are learners who learn instrumentally, because they have a very low control over objectives, but high control over the processes and means.

“Instrumental” refers to the instrumentalization of the learning. In other words: “There is pressure to learn but the logic of discovery is relatively free in relation to the means” (Dunlop and Radaelli, 2013: 613). Likewise, our findings show that program units are learning through delegation, with high control over objectives, but very little on the processes. This corresponds to a ‘delegation’ type of policy learning. “Here learners set their learning objectives autonomously, but are constrained as to the contents and means of learning. Here the shadow can be a resource used to organize the production and dissemination of knowledge” (Dunlop and Radaelli, 2013:613).

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. Furthermore, when asked the learners indicate different things about learning. Our data show 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 other questions related to learning and evaluation. First, it is unclear whether these findings hold in other empirical cases of evaluation such as prospective ex-ante evaluation or impact assessment or retrospective regulatory evaluation, and if these findings from the particular evaluation framework of the Commission are also relevant in other organizations. 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

- Albæk E. (1995) Between Knowledge and Power: Utilization of Social Science in Public Policy Making. *Policy Sciences* 28: 79-100.
- Alkin M and Christie CA. (2004) An evaluation theory tree revisited. In: Alkin M (ed) *Evaluation roots*. Thousand Oaks: Sage.
- Alkin M and Taut S. (2003) Unbundling Evaluation Use. *Studies in Educational Evaluation* 29: 1-12.
- Alkin MC and Stecher B. (1983) Evaluation in Context: Information Use in Elementary School Decision Making. *Studies in Educational Evaluation* 9: 23-32.
- Auditors ECo. (2005) Presidential Letter.
- Balthasar A and Rieder S. (2000) Learning from Evaluations: Effects of the Evaluation of the Swiss Energy 2000 Programme. *Evaluation* 6: 245-260.
- Barnes M, Matka E and Sullivan H. (2003) Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems. *Evaluation* 9: 265-284.
- Batterbury SCE. (2006) Principles and purposes of European Union policy evaluation. *Regional Studies* 40: 179-188.
- Bauer M. (2008) Introduction: Organizational change, management reform and EU policy-making. *Journal of European Public Policy* 15.
- Bazeley P. (2013) *Qualitative Data Analysis with NVIVO*, London: Sage Publications Ltd.
- Benjamin L. (2012) Nonprofit Organizations and Outcome Measurement: From Tracking Program Activities to Focusing on Frontline Work. *American Journal of Evaluation* 33.
- Bennett CJ and Howlett M. (1992) The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences* 25: 275-294.
- Bernard HR. (2013) *Social Research Methods*, Thousand Oaks: Sage.
- Bernard HR and Ryan GW. (2010) *Analyzing Qualitative Data: Systemic Approaches*, Thousand Oaks: Sage.
- Beyer JM and Trice HM. (1982) The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly* 27: 591-622.
- Bienias S and Iwona L. (2009) Evaluation Systems in the Visegrad Member States. Warsaw: Polish Ministry of Regional Development.
- Borrás S. (2011) Policy learning and organizational capacities in innovation policies. *Science and Public Policy* 38: 725-734.
- Borrás S and Højlund S. (2014) Evaluation and Policy Learning - The Learners' Perspective. *European Journal of Political Research* 53.
- Borum F and Hansen HF. (2000) The Local Construction and Enactment of Standards for Research Evaluation - The Case of the Copenhagen Business School. *Evaluation* 6: 281-299.

- 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.
- Boven M. (2005) Public Accountability. In: Ferlie E, Lynn LE and Pollitt C (eds) *The Oxford Handbook of Public Management*. Oxford: Oxford University Press.
- Bryman A. (2012) *Social Research Methods*, Oxford: Oxford University Press.
- Burns C. (2012) How and When Did We Get Here? An Historical Institutional Analysis of EU Biotechnology Policy. *Journal of European Integration* 34: 341-357.
- Buray J, Alkin M and Ruskus J. (1985) Organizing evaluations for use as a management tool. *Studies in Educational Evaluation* 11: 131-157.
- Böhling K. (2013) Sidelined Member States: Commission learning from Experts in the Face of Comitology. *Journal of European Integration*.
- Cassell C and Symon G. (1994) *Qualitative Methods in Organizational Research*, London: Sage.
- Christiansen T. (1997) Tensions of European governance: politicized bureaucracy and multiple accountability in the European Commission. *Journal of European Public Policy* 4: 73-90.
- Christie CA. (2007) Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation* 28: 8-25.
- Cohen MD, March JG and Olsen J. (1972) A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly* 17.
- Commission E. (1996) Concrete steps towards best practice across the Commission. In: Secretariat-General (ed). Brussels.
- Commission E. (1999) Spending more wisely: Implementation of the Commission's evaluation policy. In: Secretariat-General (ed). Brussels.
- Commission E. (2000a) Focus on results: Strengthening evaluation of Commission activities. In: Commission E (ed) *SEC(2000)1051*. Brussels.
- Commission E. (2000b) Reforming the Commission: A White Paper. Brussels.
- Commission E. (2004) Evaluating EU Activities - A practical guide for the Commission services. Brussels: European Commission.
- Commission E. (2007a) Responding to Strategic Needs: Reinforcing the use of evaluation. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2007b) Revision of the Internal Control Standards and Underlying Framework - Strengthening Control Effectiveness. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2010) Smart Regulation in the European Union. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E. (2013) Public Consultation on Commission Guidelines for Evaluation. In: Secretariat-General (ed). Brussels: European Commission.
- Commission E and Jacobsen S. (2007) Evalueringsaktiviteter & -resultater på tværs af Europa-

- Kommissions tjenestegrene. *Dansk Evalueringsselskabs konference*. Kolding, Denmark.
- Cook TD and Pollard WE. (1977) Guidelines: how to recognize and avoid some common problems of mis-utilization of evaluation research findings. *Evaluation* 4: 161-164.
- Council E. (1995) Financial Regulation Amendment. In: Union TCotE (ed).
- Cousins B. (2004) Commentary: Minimizing Evaluation Misuse as Principled Practice. *American Journal of Evaluation* 25: 391-397.
- Cousins JB and Earl LM. (1995a) *Participatory evaluation in education: Studies in evaluation use and organizational learning*, London: Falmer Press.
- Cousins JB and Earl LM. (1995b) Participatory evaluation: Enhancing evaluation use and organizational learning capacity. *The Evaluation Exchange* 1: 2-3.
- Cousins JB, Goh SC, Clark S, et al. (2004) Integrating evaluative inquiry into the organisational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation* 19: 99-141.
- Cousins JB and Leithwood KA. (1986) Current empirical research on evaluation utilization. *Review of Educational Research* 56: 331-364.
- Cronbach LJ and Suppes P. (1969) *Research for tomorrow's schools: Disciplined inquiry for education*, New York: Macmillan.
- Dahler-Larsen P. (2006) Organizing Knowledge: Evidence and the Construction of Evaluative Information Systems. In: Rist R and Stame N (eds) *From Studies to Streams*. New Brunswick: Transaction Publishers.
- Dahler-Larsen P. (2012) *The Evaluation Society*, Stanford, Ca.: Stanford University Press.
- Dahler-Larsen P and Krogstrup HK. (1998) *Den Rituelle Refleksion – om evaluering i organisationer*, Odense: Odense Universitetsforlag.
- Dahler-Larsen P and Krogstrup HK. (2006) Evaluering og institutionelle standarder: Nyinstitutionelle betragtninger af evaluering som vor tids ritual. *Tidsskrift.dk*: 283-299.
- Deutsch KW. (1966) *The Nerves of Government*, New York: The Free Press.
- Dey I. (1993) *Qualitative Data Analysis: A User-Friendly Guide for Social Scientists*, London: Routledge.
- DiMaggio PJ and Powell W. (1991) Introduction. In: Powell WW and DiMaggio PJ (eds) *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press, 1-38.
- Dobbin F, Simmons B and Garrett G. (2007) The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning. *Annual Review of Sociology* 33: 449-472.
- Dolowitz D and Marsh D. (1996) Who learns what from whom: A review of the policy transfer literature. *Political Studies* 44: 343-357.
- Dubnick M. (2005) Accountability and the promise of performance - In search of the Mechanisms. *Public Performance & Management Review* 28: 376-417.
- Dunlop CA and Radaelli CM. (2013) Systematising Policy Learning: From Monolith to Dimensions.

Political Studies 61: 599-619.

- Easterby-Smith M, Crossan M and Nicolini D. (2000) Organizational Learning: Debates Past, Present and Future. *Journal of Management Studies* 37: 783-796.
- Easton D. (1965) *A Framework for Political Analysis*, Englewood Cliffs: Prentice-Hall.
- Eckerd A and Moulton S. (2011) Heterogeneous Roles and Heterogeneous Practices: Understanding the Adoption and Uses of Nonprofit Performance Evaluations. *American Journal of Evaluation* 32: 98-117.
- Ecorys and COWI. (2008) Study on the state of implementation of Activity Based Management in the European Commission. Rotterdam.
- Egan M. (2009) Governance and learning in the post-Maastricht era? *Journal of European Public Policy* 16: 1244-1253.
- Ellinas A and Suleiman E. (2008) Reforming the Commission: between modernization and bureaucratization. *Journal of European Public Policy* 15: 708-725.
- Eser TW and 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.
- Eureval-C3E. (2006) Study on the Use of Cost-effectiveness Analysis in EC's Evaluations. Lyon: Eureval-C3E.
- Ferry M and Olejniczak K. (2008) The use of evaluation in the management of EU programmes in Poland. Warsaw: Ernst & Young.
- Fetterman DM. (1994) Steps of empowerment evaluation: From California to Cape Town. *Evaluation and Program Planning* 17: 305-313.
- Finne H, Levine M and Nilssen T. (1995) Trailing research: A model for useful program evaluation. *Evaluation* 1: 11-31.
- Floden RE and Weiner SS. (1978) Rationality to Ritual: The Multiple Roles of Evaluation in Governmental Processes. *Policy Sciences* 9: 9-18.
- Flyvbjerg B. (2006) Five Misunderstandings About Case-Study Research. *Qualitative Inquiry* 12: 219-245.
- Francesco FD, Radaelli CM and Troeger VE. (2011) Implementing regulatory innovations in Europe: the case of impact assessment. *Journal of European Public Policy* 19: 491-511.
- Freeman R. (2006) Learning in public policy. In: Moran M, Rein M and Goodin RE (eds) *Oxford Handbook of Public Policy*. Oxford: Oxford University Press, 367-389.
- Furubo J-E. (2006) Why Evaluations Sometimes Can't be Used - and Why They Shouldn't. In: Rist R and Stame N (eds) *From Studies to Streams*. New Jersey: Transaction Publishers, 147-165.
- Furubo J-E, Rist RC and Sandahl R. (2002) *International Atlas of Evaluation*, London: Transaction Publishers.
- Gornitzka Å and Sverdrup U. (2011) Access of experts: information and EU decision-making. *West European Politics* 34: 1452-1474.

- Grin J and Loeber A. (2007) Theories of policy learning: agency, structure, and change. In: Fischer F, Miller GJ and Sidney MS (eds) *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. Boca Raton: CRC Press, 201-219.
- Grøn C. (2009) Same Procedure as Last Year? An analysis of constellations of trust and control in management in the European Commission. *Department of Political Science*. Copenhagen: University of Copenhagen.
- Hanberger A and Schild I. (2004) Strategies to Evaluate a University–Industry Knowledge-exchange Programme. *Evaluation* 10: 475-492.
- Hansen HF and Borum F. (1999) The Construction and Standardization of Evaluation. The Case of the Danish University Sector. *Evaluation* 5: 303-329.
- Hansen M, Alkin MC and Wallace TL. (2012) Depicting the logic of three evaluation theories. *Evaluation and Program Planning*.
- Harlow C. (2002) *Accountability in the European Union*, Oxford: Oxford University Press.
- Hartley J. (2004) Case studies in organizational research. In: Cassell C and Symon G (eds) *Qualitative methods in organizational research, a practical guide*. London: Sage, 208-229.
- Heclo H. (1974) *Modern Social Politics in Britain and Sweden*, New Haven: Yale University Press.
- Henry GT. (2004) Why not use? In: Caracelli VJ and Preskill H (eds) *The expanding scope of evaluation use. New Directions for Evaluation*. San Francisco, CA: Jossey-Bass, 85–98.
- Henry GT and Mark MM. (2003) Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation* 24: 293.
- Hoerner J and Stephenson P. (2012) Theoretical Perspectives on Approaches to Policy Evaluation in the EU: The Case of Cohesion Policy. *Public Administration* 90: 699-715.
- Hofstetter C and Alkin M. (2003) Evaluation use revisited. In: Kellaghan T, Stufflebeam D and Wingate L (eds) *International Handbook of Educational Evaluation*. London: Kluwer Academic.
- Hood C. (1991) A public management for all seasons? *Public Administration* 69.
- Højlund S. (2014a) Evaluation use in evaluation systems - the case of the European Commission. *Evaluation* 20: 428-446.
- Højlund S. (2014b) Evaluation use in the organisational context – Changing focus to improve theory. *Evaluation* 20: 26-43.
- Højlund S. (Forthcoming) Between Accountability and Learning - The Case of the European Commission's Evaluation System.
- Imam I, LaGoy A and Williams B. (2007) Introduction. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes: EdgePress of Inverness.
- Jacob S. (2005) *Institutionnaliser l'évaluation des politiques publiques. Étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*, Bruxelles: PIE-Peter Lang.
- James O and Lodge M. (2003) The limitations of "policy transfer" and "lesson drawing" for public research. *Political Studies Review* 1: 179-193.

- Johnson K, Greenseid LO, Toal SA, et al. (2009a) Research on Evaluation Use – A review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation* 30: 377-410.
- Johnson K, Greenseid LO, Toal SA, et al. (2009b) Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation* 30: 377-410.
- Kassim H. (2008) 'Mission impossible', but mission accomplished: the Kinnock reforms and the European Commission. *Journal of European Public Policy* 15: 648-668.
- King JA and Thompson B. (1983) Research on school use of program evaluation: A literature review and research agenda. *Studies in Educational Evaluation* 9: 5-21.
- Kirkhart KE. (2000) Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation* 88: 5-23.
- Kohlbacher F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research* 7.
- Laffan B. (1997a) *The Finances of the European Union*, Basingstoke: Macmillan.
- Laffan B. (1997b) From policy entrepreneur to policy manager: the challenge facing the European Commission. *Journal of European Public Policy* 4: 422-438.
- Laffan B. (2003) Auditing and accountability in the European Union. *Journal of European Public Policy* 10: 762-777.
- Langley A. (1998) In Search of Rationality - The Purpose Behind the Use of Formal Analysis in Organizations. In: Maanen JV (ed) *Qualitative Studies of Organizations*. Thousand Oaks: SAGE Publications, 51-90.
- Lazarsfeld PF, Sewell WH and Wilensky HL. (1967) *The uses of Sociology*, New York: Basic Books.
- Leeuw FL and Furubo J-E. (2008) Evaluation Systems : What Are They and Why Study Them? *Evaluation* 14: 157-169.
- Legard R, Keegan J and Ward K. (2003) In-depth Interviews. In: Ritchie J and Lewis J (eds) *Qualitative Research Practice*. London: Sage Publications.
- Levin B. (1987) The uses of research: A case study in research and policy. *The Canadian Journal of Program Evaluation* 2: 43-55.
- Leviton LC. (2003) Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation* 24: 525-535.
- Leviton LC and Hughes EFX. (1981) Research on the utilization of evaluations: A review and synthesis. *Evaluation Review* 5: 525-548.
- Laat Bd. (2005) Study on the Use of Evaluation Results in the Commission. In: Laat Bd (ed). Paris: Technopolis.
- March J. (2010) *The ambiguities of experience* Ithaca, N.Y.: Cornell University Press.
- Mark MM and Henry GT. (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation* 10: 35-57.
- Mayring P. (2000) Qualitative Content Analysis. *Forum: Qualitative Social Research* 1.

- Mayring P. (2002) Qualitative content analysis - research instrument or mode of interpretation. The Role of the REsearcher in Qualitative Psychology. In: Kieglmann M (ed) *Nexus Qualitative Psychology*. Tübingen: Verlag Ingeborg Huber, 139-148.
- Mayring P. (2003) *Qualitative Inhaltsanalyse, Grundlage und Techniken*, Weinheim: Beltz, UTB.
- Mayring P. (2004) Qualitative Content Analysis. In: Flick U, Von Kardorff E and Steinke I (eds) *A Companion to Qualitative Research*. London: SAGE, 266-269.
- Mendez C and Bachtler J. (2011) Administrative reform and unintended consequences: an assessment of the EU Cohesion policy 'audit explosion'. *Journal of European Public Policy* 18: 746-765.
- Meyer JW and Rowan B. (1977) Institutionalized organizations: Formal structure as Myth and Ceremony. *American Sociological Review* 83: 340-363.
- Meyer JW and Rowan B. (1991) Institutionalized Organizations: Formal Structure as Myth and Ceremony. In: Powell WW and DiMaggio PJ (eds) *The New Institutionalism in Organizational Analysis*. Chicago: The University Chicago Press, 41-62.
- Miles MB and Huberman AM. (1994) *Qualitative Data Analysis*, Thousand Oaks, CA: Sage Publications.
- Mocker DW and Spear GE. (1982) *Lifelong Learning: Formal, Nonformal, Informal, Informal, and Self-Directed.*, Columbus OH: Eric.
- OECD. (2002) Regulatory policies in OECD Countries: From Interventionism to Regulatory Governance. Paris: OECD.
- Olejniczak K. (2013) Mechanisms Shaping an Evaluation System - A Case Study of Poland 1999-2010. *Europe-Asia Studies* 65: 1642-1666.
- Owen JM. (2002) Linking evaluation use to the research utilisation literature. *Annual Meeting of the American Evaluation Association*. Washington DC.
- Owen JM and Lambert Fc. (1995) Roles for Evaluation in Learning Organizations. *Evaluation* 1: 237-250.
- Owen JM and Rogers P. (1999) *Program evaluation: Forms and approaches*, London: Sage.
- Owens S, Rayner T and Bina O. (2004) New agendas for appraisal: reflections on theory, practice, and research. *Environement and Planning* 36: 1943-1959.
- Patton MQ. (1997) *Utilization-focused evaluation: The new century text*, Thousand Oaks: Sage.
- Pawson R and Tilley N. (1997) *Realistic Evaluation*, London: SAGE Publication Ltd.
- Pelz DC. (1978) Some expanded perspectives on use of social science in public policy. In: Yinger JM and Cutler SJ (eds) *Major Social Issues: A Multidisciplinary View*. New York: Macmillan, 346-357.
- Pollitt C and Bouckaert G. (2004) *Public Management Reform: A Comparative Analysis*, Oxford University Press.
- Pollitt C, Girre X, Lonsdale J, et al. (1999) *Performance Audit and Public Management in Five Countries* Oxford and New York: Oxford University Press.

- Porter TM. (1995) *Trust in Numbers: the pursuit of objectivity in science and public life*, UK: Princeton University Press.
- Power M. (1997) *The Audit Society - Rituals of Verification*, Oxford: Oxford University Press.
- Power M. (2005) The Theory of Audit Explosion. In: Ferlie E, Lynn LE and Pollitt C (eds) *Oxford Handbook of Public Management*. Oxford: Oxford University Press.
- Preskill H. (2008) Evaluation's Second Act: A Spotlight on Learning. *American Journal of Evaluation* 29: 127.
- Preskill H and Torres RT. (1999) Building Capacity for Organizational Learning Through Evaluative Inquiry. *Evaluation* 5: 42-60.
- Preskill H, Zuckerman B and Matthews B. (2003) An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation* 24.
- Pröpper IMAM. (1987) Beleidsevaluatie als argumentatie. *Beleidswetenschap* 2.
- Radaelli CM. (2009) Measuring policy learning: regulatory impact assessment in Europe. *Journal of European Public Policy* 16: 1145-1164.
- Radaelli CM and Dunlop CA. (2013) Learning in the European Union: theoretical lenses and meta-theory. *Journal of European Public Policy* 20: 923-940.
- Radaelli CM and Meuwese ACM. (2010) Hard Questions, Hard Solutions: Proceduralism through Impact Assessment in the EU. *West European Politics* 20: 923-940.
- Remenyi D, Money A, Price D, et al. (2002) The creation of knowledge through case study research. *Irish Journal of Management* 23: 1-17.
- Rich RF. (1977) Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding. In: Weiss CH (ed) *using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Rist R and Stame N. (2006) *From Studies to Streams: Managing Evaluative Systems*, London: Transaction Publishers.
- Ritchie J, Lewis J and Elam G. (2003) Designing and Selecting Samples. In: Ritchie J and Lewis J (eds) *Qualitative Research Practice - A Guide for Social Science Students and Researchers*. London: SAGE.
- Ryan GW and Bernard HR. (2000) Data management and analysis methods. In: Denzin N and Lincoln YS (eds) *Handbook for qualitative research*. Thousand Oaks: Sage.
- Sager F and 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.
- Schofield J. (2004) A model of learned implementation. *Public Administration* 82: 283-308.
- Schout A. (2009) Organizational learning in the EU's multilevel-governance system. *Journal of European Public Policy* 16: 1124-1144.
- Schreier M. (2012) *Qualitative Content Analysis in Practice*, Thousand Oaks: SAGE Publications Ltd.

- Schwandt TA. (1997) Evaluation as Practical Hermeneutics. *Evaluation* 3: 69–83.
- Schwandt TA. (2009) Globalizing influences on the Western evaluation imaginary. In: Ryan KE and Cousins JB (eds) *Sage international handbook on educational evaluation*. Los Angeles: Sage, 119-138.
- Schön-Quinlivan E. (2008) Implementing organizational change - the case of the Kinnock reforms. *Journal of European Public Policy* 15: 726-742.
- Scott RW. (1995) *Institutions and organizations*, Thousand Oaks: SAGE.
- Scott RW and Meyer JW. (1994) *Institutional environments and organizations: Structural complexity and individualism*, Thousand Oaks: SAGE Publications.
- Scriven M. (1967) The methodology of evaluation. In: Stake RE (ed) *Curriculum evaluation*. Chicago: Rand McNally.
- Scriven M. (1991) Beyond Formative and Summative Evaluation. In: McLaughlin M and Philips DC (eds) *Evaluation and Education: At Quarter Century*. 3 ed. Chicago: University of Chicago Press.
- Selznik P. (1949) *TVA and the Grass Roots*, Berkeley: University of California Press.
- Selznik P. (1984) *Leadership in administration: A sociological interpretation*, Berkeley: University of California Press.
- Shulha LM and Cousins B. (1997) Evaluation Use: Theory, Research and Practice Since 1986. *Evaluation Practice* 18: 195-208.
- Simon H. (1957) *Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization*, New York: Macmillan.
- Smith AS. (2013) How the European Commission's Policies Are Made: Problematization, Instrumentation and Legitimation. *Journal of European Integration*.
- Spence D. (2000) Plus ca change, plus c'est la meme chose? Attempting to reform the European Commission. *Journal of European Public Policy* 7: 1-25.
- Stake RE. (2000) Case studies. In: Denzin N and Lincoln YS (eds) *Handbook of qualitative research*. Thousand Oaks: Sage, 435-453.
- Stake RE. (2006) *Multiple Case Study Analysis*, Thousand Oaks, CA: Sage Publications.
- Stern E. (2006) Contextual challenges for evaluation practice. In: Shaw I, Green J and M. M (eds) *The SAGE Handbook of Evaluation*. London: SAGE.
- Stern E. (2009) *Evaluation policy in the European Union and its institutions*.
- Stewart DW and Kamins MA. (1993) *Secondary research: information sources and methods*, Thousand Oaks, CA: Sage Publications.
- Stufflebeam D. (2001) Evaluation Models. *New Directions for Evaluation* 89: 7-98.
- Summa H and Toulemonde J. (2002) Evaluation in the European Union: addressing complexity and ambiguity. In: Furubo J-E, Sandahl R and Rist R (eds) *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.
- Thompson B and King JA. (1981) Evaluation utilization: A literature review and research agenda.

- Annual Meeting of the American Educational Research Association*. Los Angeles, CA.
- Titscher S, Meyer M, Wodak R, et al. (2000) *Methods of Text and Discourse Analysis*, Thousand Oaks: Sage.
- Torres RT and Preskill H. (2001) Evaluation and Organizational Learning: Past, Present, and Future. *American Journal of Evaluation* 22: 387-395.
- Toulemonde J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *Vierteljahrshefte zur Wirtschaftsforschung* 69: 350-357.
- Toulemonde J, Summa H and Usher N. (2005) Assessing EU Evaluations. In: Schwartz R and Mayne J (eds) *Quality matters: seeking confidence in evaluation, auditing and performance reporting*. New Brunswick: Transaction Publishers.
- Van der Knaap P. (1995) Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation? *Evaluation* 1: 189-216.
- van der Meer F-B and Edelenbos J. (2006) Evaluation in Multi-Actor Policy Process. Accountability, Learning and Co-operation. *Evaluation* 12: 201-218.
- Varone F, Jacob S and Winter LD. (2005) Polity, Politics and Policy Evaluation in Belgium. *Evaluation* 11: 253.
- Vedung E. (1997) *Public Policy and Program Evaluation*, New Brunswick: NJ: Transaction Publishers.
- Vedung E. (2010) Four Waves of Evaluation Diffusion. *Evaluation* 16: 263-277.
- Volden C, Ting MM and Carpenter DP. (2008) A Formal Model of Learning and Policy Diffusion. *American Political Science Review* 102: 319-332.
- Weiss CH. (1978) Improving the linkage between social research and public policy. In: Lynn LE (ed) *Knowledge and Policy: The Uncertain Connection*. Washington DC: National Research Council.
- Weiss CH. (1979) The many meanings of research utilization. *Public Administration* 39: 426-431.
- Weiss CH. (1992) *Organizations for Policy Analysis - Helping Government Think*, Newbury Park, CA: SAGE.
- Weiss CH. (1998) Have We Learned Anything New About the Use of Evaluation. *American Journal of Evaluation* 19: 21-33.
- Weiss CH and Bucuvalas M. (1980) *Social Science Research and Decision Making*, New York: Columbia University Press.
- Wille A. (2010) Political-Bureaucratic Accountability in the EU Commission: Modernising the Executive. *West European Politics* 33: 1093-1116.
- Williams B and Imam I. (2007) *Systems concepts in evaluation - An Expert Anthology*, Point Reyes: EdgePress.
- Williams K, Laatz B and Stern E. (2002) The Use of Evaluation in the Commission Services. Paris: Technopolis.
- Wonka A. (2008) Decision-making dynamics in the European Commission: partisan, national or sectoral? *Journal of European Public Policy* 15: 1145-1163.

- Yin RK. (2003) *Case study research: Design and methods*, Thousand Oaks, CA: Sage.
- Zald MN and Denton P. (1963) From Evangelism to General Services: the Transformation of the YMCA. *Administrative Science Quarterly* 8: 214-234.
- Zito A and Schout A. (2009) Learning theory reconsidered: EU integration theories and learning. *Journal of European Public Policy* 16: 1103-1123.

Appendix 1 – Interview guide

Introduction questions

Q1. What is your name, your unit, and your job description?

Q2. What was your role in relation to the midterm evaluation?

Learning from the evaluation

Q3. Did the participants learn something from the midterm evaluation?

Q4. Did you learn something?

Q5. Did the midterm evaluation influence some changes of work organization around the program?

Q6. Which participants learned most (and learned the least) from the midterm evaluation, and why?

Q7. Were the findings of the midterm evaluation used afterwards? By whom?

Q8. What could be done in the future to enhance learning from mid-term evaluations like this one?

The objectives of the evaluation

Q9. Which units and departments in the European Commission were involved in the design and planning of the mid-term evaluation before tendering? What were their relative roles? [OBJ CONTROL]

Q10. Who had most influence on the evaluation questions and the formulation of evaluation objectives? [OBJ CONTROL]

Q11: Describe the interaction between the stakeholders during the design-phase. Who was the most active in terms of influencing the evaluation design? [OBJ INTERACT] [OBJ PRO-ACT]

Q12. In your opinion, was the information and knowledge requested in the terms of reference and other objective-defining documents a sufficient starting point for the mid-term evaluation, or were there some missing aspects? [OBJ LEGIT]

Q13. Describe your influence on the formulation of evaluation questions (ToR, inception- and interim report) [OBJ CONTROL] [OBJ PRO-ACT]

Q14. Did the objectives influence the way in which the mid-term evaluation was conducted, why, how?

<i>The process of the evaluation</i>

Q15. Please describe briefly the evaluation process of the midterm evaluation of program XX. (Probe: Conflictual? Consensual? Collaborating?) [PROCESS INTERACT]

Q16. Describe your role in the evaluation implementation process. Were you actively involved in the evaluation implementation? Why? [PROCESS PRO-ACT]

Q17. Did you exercise some influence in the implementation of the evaluation? [PROCESS control]

Q18 Did you exercise some influence in the definition of the methodology and data collection during the implementation phase? [PROCESS METHOD]

Q19. Do you think that the process produced useful valuable findings? [PROCESS METHOD-LEG]

Q20. How would you assess the capacity of the Steering committee and the consultant/expert group to conduct the evaluation? [FRAME-LEG]

Appendix 2 – List of interviewees

Case-specific interviews (25)

Interview number	Case	Interviewee function	Interview date	NVIVO Nodes	NVIVO References
1	FP7	Evaluator	19th June 2013	16	27
2	FP7	Evaluation function	20th 2013	25	65
3	FP7	Desk Officer	25th June 2013	16	27
4	FP7	Evaluator	5th June 2013	20	36
5	FP7	Evaluator	25th Jun 2013	18	40
6	FP7	Evaluator	25th June 2013	1	1
7	FP7	Evaluator	12th June 2013	21	43
8	FP7	Evaluator	18th September 2013	16	21
9	FP7	Evaluation function	20th June 2013	17	26
10	FP7	Evaluator	26th June 2013	20	43
11	IEE	Evaluator	18th June 2013	19	38
12	IEE	Desk Officer	19th June 2013	21	61
13	IEE	Desk Officer	2nd July 2013	18	50
14	IEE	Management	31st July 2013	25	70
15	IEE	Management	31st July 2013	18	44
16	LIFE	Management	20th June 2013	28	115
17	LIFE	Consultant	24th June 2013	25	67

18	LIFE	Evaluation Function	25th June 2013	22	58
19	LIFE	Evaluation manager	8th August 2013	22	63
20	LIFE	Policy Unit	25th June 2013	27	67
21	LIFE	Management	5th September 2013	25	67
22	LIFE	Desk Officer	11th July 2013	16	25
23	LIFE	Evaluator	3rd July 2013	24	51
24	LIFE	Desk Officer	9th September 2013	23	43
25	LIFE	Evaluator	27th June 2013	18	42

Explorative background interviews (35)

Interview number	Organization	Function
1	DG REGIO	Evaluation function
2	Consultancy	Evaluator and Commission evaluation trainer
3	DG DEVCO	Evaluation function
4	OECD (Formerly in a consultancy)	Evaluator and Commission evaluation trainer
5	Consultancy	Consultant
6	DG DEVCO	Evaluation function
7	European Parliament	Member of European Parliament
8	DG REGIO	Head of Unit
9	DG TAXUD	Evaluation function
10	DG INFSO/CONNECT	Evaluation function
11	DG MARKT	Head of Unit
12	Secretariat General	Evaluation function
13	DG INFSO/CONNECT	Evaluation function

14	DG DEVCO	Evaluation function
15	DG BUDGET	Head of Unit
16	DG INFO/CONNECT	Evaluation function
17	Consultancy	Evaluator and Commission evaluation trainer
18	Secretariat General	Head of Unit
19	DG DEVCO	Evaluation function
20	DG ENV	Head of Unit
21	Consultancy	Evaluator
22	DG ENV	Project manager
23	European Parliament	Policy officer
24	European Parliament	Assistant to MEP
25	European Parliament	Assistant to MEP
26	European Parliament	Assistant to MEP
27	European Council	National Representative
28	DG ENV	Head of Unit
29	DG ENV	Head of Unit
30	DG ENV	Evaluation function
31	DG ENV	Evaluation function
32	DG ENV	Desk officer
33	DG ENV	Desk officer
34	Consultancy	Evaluator
35	Consultancy	Evaluator

ⁱ The keyword 'institutionalism' resulted in two articles in 'The American Journal of Evaluation'; eighteen articles in 'Evaluation'; one article (plus two EES-conference papers) in the 'Canadian Journal of Program Evaluation'; two articles in 'Evaluation Review'. The keyword 'neo-institutional*' resulted in one relevant article in 'The American Journal of Evaluation'; one article in 'Evaluation'; the search engine could not make the search in the 'Canadian Journal of Program Evaluation'; no articles in 'Evaluation Review'. All searches were carried out in 'All fields' signifying all searchable elements combined (e.g. author, references, title, abstract etc.).

The four journals were sampled according to the following four criteria: The source should; 1) be peer-reviewed; 2) address the common evaluation community across sectors, methods and geography; 3) be searchable online; 4) be in English.

Reflections on sampling methodology:

I recon that the third sampling criterion might have a bias towards sources from the US and the EU, while the fourth criterion most certainly is biased towards the Anglo-Saxon world. I am unaware of peer-reviewed journals that are not published online, but that is of course not to say that they do not exist or are not relevant to this study. Also, it is likely that sources exist in Asian languages that I would be unable to read. However, I consider it more likely that debates in French and German have not fully been integrated into the 'mainstream' English literature. That is not uncommon in other fields of study such as EU-studies and at the EES Conference in Helsinki in 2012, I overheard at least one German colleague arguing in plenum that this was a problem.

I acknowledge that there is a high probability that literature excluded from this review by the four sampling criteria contains information on the use of organisational theory in relation to evaluation use. However, I argue that the sample is representative of the mainstream literature on evaluation, which has its roots in the Anglo-Saxon world (notably the US). Therefore, previous attempts to integrate an organisational institutionalist framework into explanations of evaluation use should have been reflected in the sample if it had value to the general field of evaluation. Moreover, other reviews of the literature on evaluation use were examined and referred to accordingly in the article's page 6. One could expect that these reviews would have referred to sources containing institutionalist explanations of evaluation use, if these sources were relevant to the general field of evaluation. I am able to read and comprehend German and French, so that has not been an obstacle during the analysis of data in this article. Finally, I consistently refer candidly to the literature that I have come across, where organisational institutional arguments are used. That is particularly the case of Dahler-Larsen (2012).

ⁱⁱ The literature on evaluation use and (expert) knowledge use have run parallel and crossed in many years, which makes sense since evaluation and (expert) knowledge use

are often the same. Therefore, I regard Boswell's (2008) study as a valuable contribution also to the understanding of evaluation use.

TITLER I PH.D.SERIEN:

– *a Field Study of the Rise and Fall of a Bottom-Up Process*

2004

1. Martin Grieger
Internet-based Electronic Marketplaces and Supply Chain Management
2. Thomas Basbøll
*LIKENESS
A Philosophical Investigation*
3. Morten Knudsen
*Beslutningens vaklen
En systemteoretisk analyse af moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000*
4. Lars Bo Jeppesen
*Organizing Consumer Innovation
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers*
5. Barbara Dragsted
*SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*
6. Jeanet Hardis
*Sociale partnerskaber
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet*
7. Henriette Hallberg Thygesen
System Dynamics in Action
8. Carsten Mejer Plath
Strategisk Økonomistyring
9. Annemette Kjærgaard
Knowledge Management as Internal Corporate Venturing
10. Knut Arne Hovdal
*De professionelle i endring
Norsk ph.d., ej til salg gennem Samfundslitteratur*
11. Søren Jeppesen
*Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
– A Critical Realist Approach*
12. Lars Frode Frederiksen
*Industriel forskningsledelse
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder*
13. Martin Jes Iversen
*The Governance of GN Great Nordic
– in an age of strategic and structural transitions 1939-1988*
14. Lars Pynt Andersen
*The Rhetorical Strategies of Danish TV Advertising
A study of the first fifteen years with special emphasis on genre and irony*
15. Jakob Rasmussen
Business Perspectives on E-learning
16. Sof Thrane
*The Social and Economic Dynamics of Networks
– a Weberian Analysis of Three Formalised Horizontal Networks*
17. Lene Nielsen
Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca
18. S.J Valstad
*Organisationsidentitet
Norsk ph.d., ej til salg gennem Samfundslitteratur*

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets
 20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice
 21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation
 22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis
 23. Pernille Schnoor
*Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv*
 24. Sidsel Fabech
*Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser*
 25. Klavs Odgaard Christensen
*Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada*
 26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations
 27. Holger Højlund
*Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003*
 28. Christine Mølgaard Frandsen
*A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten*
 29. Sine Nørholm Just
The Constitution of Meaning – A Meaningful Constitution? Legitimacy, identity, and public opinion in the debate on the future of Europe
- 2005**
1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development
 2. Helle Hedegaard Hein
Mellem konflikt og konsensus – Dialogudvikling på hospitalsklinikker
 3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development
 4. Søren Buhl Pedersen
*Making space
An outline of place branding*
 5. Camilla Funck Ellehave
*Differences that Matter
An analysis of practices of gender and organizing in contemporary work-places*
 6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger
 7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain Benchmarking as a Means to Managing Supply Chains
 8. Caroline Aggestam-Pontoppidan
*From an idea to a standard
The UN and the global governance of accountants' competence*
 9. Norsk ph.d.
 10. Vivienne Heng Ker-ni
An Experimental Field Study on the

- Effectiveness of Grocer Media Advertising*
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales
11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives
12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo
13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003
14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks
15. Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour
16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change
17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa
18. Signe Jarlov
Konstruktioner af offentlig ledelse
19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
- An empirical study employing data elicited from Danish EFL learners*
20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information
21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants
22. Annie Bekke Kjær
Performance management i Process-innovation
– belyst i et social-konstruktivistisk perspektiv
23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gørem i den kunstneriske arbejdspraksis
24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser
25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees
26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions
27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgte fornyende rolle
28. Gitte Thit Nielsen
Remerger
– skabende ledelseskrafter i fusion og opkøb
29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederens meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter
31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur
32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring
33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer
34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals
35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under 1. verdenskrig
36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery
37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management
2. Niels Rom-Poulsen
Essays in Computational Finance
3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production
4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures
5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj
6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China
7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff
8. Cynthia Selin
Volatile Visions: Transactons in Anticipatory Knowledge
9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies
10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?
11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

- A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy*
12. Mikkel Flyverbom
*Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks*
 13. Anette Grønning
*Personen bag
Tilstedevær i e-mail som interaktionsform mellem kunde og medarbejder i dansk forsikringskontekst*
 14. Jørn Helder
*One Company – One Language?
The NN-case*
 15. Lars Bjerregaard Mikkelsen
*Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets*
 16. Lise Granerud
*Exploring Learning
Technological learning within small manufacturers in South Africa*
 17. Esben Rahbek Pedersen
*Between Hopes and Realities:
Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)*
 18. Ramona Samson
*The Cultural Integration Model and European Transformation.
The Case of Romania*
- 2007**
1. Jakob Vestergaard
*Discipline in The Global Economy
Panopticism and the Post-Washington Consensus*
 2. Heidi Lund Hansen
*Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices*
 3. Sudhanshu Rai
*Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"*
 4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur
 5. Serden Ozcan
*EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective*
 6. Kim Sundtoft Hald
*Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships*
 7. Tobias Lindeberg
*Evaluative Technologies
Quality and the Multiplicity of Performance*
 8. Merete Wedell-Wedellsborg
*Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer*
 9. Lars Frederiksen
*Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays*
 10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredsheds-målinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv
12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry
13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sam-eksistens i værdibaseret ledelse!
14. Kristina Birch
Statistical Modelling in Marketing
15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing
16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation
17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets
18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation
19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia
20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analyser af en organisations kontinuerlige arbejde med informations-teknologi
21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis
22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring
23. Christa Breum Amhøj
Det selvskabte medlemskab om managementsstaten, dens styringsteknologier og indbyggere
24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC
25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice
26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective
27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45
28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

- on the WWW – an implementation and evaluation*
29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative
 30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?
 31. Christina Holm-Petersen
*Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion*
 32. Hans Peter Olsen
*Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing*
 33. Lars Bøge Sørensen
Risk Management in the Supply Chain
 34. Peter Aagaard
*Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde*
 35. Yun Mi Antorini
*Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community*
 36. Joachim Lynggaard Boll
*Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives*
 3. Marius Brostrøm Kousgaard
*Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren*
 4. Irene Skovgaard Smith
*Management Consulting in Action
Value creation and ambiguity in client-consultant relations*
 5. Anders Rom
*Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology*
 6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms
 7. Morten Schnack
*Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling*
 8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano
 9. Lise Justesen
*Kunsten at skrive revisionsrapporter.
En beretning om forvaltningsrevisionsens beretninger*
 10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s

2008

1. Frederik Christian Vinten
Essays on Private Equity
2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions
11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives
 13. Anders Bojesen
The Performative Power of Competence – An Inquiry into Subjectivity and Social Technologies at Work
 14. Satu Reijonen
*Green and Fragile
A Study on Markets and the Natural Environment*
 15. Ilduara Busta
*Corporate Governance in Banking
A European Study*
 16. Kristian Anders Hvass
*A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry*
 17. Trine Paludan
*De uvidende og de udviklingsparate
Identitet som mulighed og restriktion
blandt fabriksarbejdere på det aftayloriserede fabriksgulv*
 18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice
 19. Jakob Elming
Syntactic reordering in statistical machine translation
 20. Lars Brømsøe Termansen
*Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics*
 21. Mia Reinholdt
The Motivational Foundations of Knowledge Sharing
 22. Frederikke Krogh-Meibom
*The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times*
 23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS
 24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers
 25. Mads Vangkilde
*Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective*
 26. Hubert Buch-Hansen
*Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective*
- 2009**
1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours
 2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice
 3. Chris Nøkkentved
*Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management*
 4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv
6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee
7. Salla Lutz
Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers
8. Jens Forssbæck
Essays on market discipline in commercial and central banking
9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?
10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry
11. Nicolaas Mouton
On the evolution of social scientific metaphors:
A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.
12. Lars Andreas Knutsen
Mobile Data Services: Shaping of user engagements
13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities
14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser
15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business
16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context
17. Dorte Hermansen
“Living the brand” som en brandorienteret dialogisk praksis:
Om udvikling af medarbejdernes brandorienterede dømmekraft
18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity
19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?
20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector
21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management
22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse
23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

- | | | | |
|-----|--|----|--|
| 24. | Christian Scheuer
<i>Employers meet employees
Essays on sorting and globalization</i> | | <i>End User Participation between Processes of Organizational and Architectural Design</i> |
| 25. | Rasmus Johnsen
<i>The Great Health of Melancholy
A Study of the Pathologies of Performativity</i> | 7. | Rex Degnegaard
<i>Strategic Change Management
Change Management Challenges in the Danish Police Reform</i> |
| 26. | Ha Thi Van Pham
<i>Internationalization, Competitiveness Enhancement and Export Performance of Emerging Market Firms: Evidence from Vietnam</i> | 8. | Ulrik Schultz Brix
<i>Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet</i> |
| 27. | Henriette Balieu
<i>Kontrolbegrebets betydning for kausalalternationen i spansk
En kognitiv-typologisk analyse</i> | 9. | Jan Ole Similå
<i>Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter</i> |
- 2010**
- | | | | |
|----|---|-----|---|
| 1. | Yen Tran
<i>Organizing Innovation in Turbulent Fashion Market
Four papers on how fashion firms create and appropriate innovation value</i> | 10. | Susanne Boch Waldorff
<i>Emerging Organizations: In between local translation, institutional logics and discourse</i> |
| 2. | Anders Raastrup Kristensen
<i>Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management</i> | 11. | Brian Kane
<i>Performance Talk
Next Generation Management of Organizational Performance</i> |
| 3. | Margrét Sigrún Sigurdardóttir
<i>Dependently independent
Co-existence of institutional logics in the recorded music industry</i> | 12. | Lars Ohnemus
<i>Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts</i> |
| 4. | Ásta Dis Óladóttir
<i>Internationalization from a small domestic base:
An empirical analysis of Economics and Management</i> | 13. | Jesper Schlamovitz
<i>Håndtering af usikkerhed i film- og byggeprojekter</i> |
| 5. | Christine Secher
<i>E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf</i> | 14. | Tommy Moesby-Jensen
<i>Det faktiske livs forbindtlighed
Følsomt informeret, ny-aristotelisk
ἦθος-tænkning hos Martin Heidegger</i> |
| 6. | Marianne Stang Våland
<i>What we talk about when we talk about space:</i> | 15. | Christian Fich
<i>Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power</i> |

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillingsforløb i fire virksomheder
17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy
18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere
19. Heidi Boye
Fødevarer og sundhed i sen-modernismen
– En indsigt i hyggefænomenet og de relaterede fødevarerpraksisser
20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival
21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective
22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work
23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen
24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire
25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry
26. Javier Busquets
Orchestrating Network Behavior for Innovation
27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan
28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position
29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study
30. Qingxin Shi
An Empirical Study of Thinking Aloud
Usability Testing from a Cultural Perspective
31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft
32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting
33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity
34. Sabrina Speiermann
Modstandens Politik
Kampagnestyling i Velfærdsstaten.
En diskussion af trafikcampagners styringspotentiale
35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care
37. Athur Mühlen-Schulte
Organising Development Power and Organisational Reform in the United Nations Development Programme
38. Louise Rygaard Jonas
Branding på butiksgulvet Et case-studie af kultur- og identitets-arbejdet i Kvickly
8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland
9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication
10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch A Study of Product Launches in the Swedish Pharmaceutical Industry
2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice
3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin – en analyse af coachingsdiskursens genealogi og governmentality
4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation
5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication
6. Simon S. Torp
Employee Stock Ownership: Effect on Strategic Management and Performance
7. Mie Harder
Internal Antecedents of Management Innovation
11. Moshe Yonatany
The Internationalization Process of Digital Service Providers
12. Anne Vestergaard
Distance and Suffering Humanitarian Discourse in the age of Mediatization
13. Thorsten Mikkelsen
Personlighedens indflydelse på forretningsrelationer
14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen ud-over "the tipping point"? – en empirisk analyse af information og kognitioner om fusioner
15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making
16. Thomas Stengade Sønderskov
Den nye mulighed Social innovation i en forretningsmæssig kontekst
17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries
18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

- integration in the process of product innovation*
19. Throstur Olaf Sigurjonsson
Governance Failure and Iceland's Financial Collapse
 20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets
 21. Heidi Tscherning
Mobile Devices in Social Contexts
 22. Birgitte Gorm Hansen
*Adapting in the Knowledge Economy
Lateral Strategies for Scientists and Those Who Study Them*
 23. Kristina Vaarst Andersen
*Optimal Levels of Embeddedness
The Contingent Value of Networked Collaboration*
 24. Justine Grønbæk Pors
*Noisy Management
A History of Danish School Governing from 1970-2010*
 25. Stefan Linder
*Micro-foundations of Strategic Entrepreneurship
Essays on Autonomous Strategic Action*
 26. Xin Li
*Toward an Integrative Framework of National Competitiveness
An application to China*
 27. Rune Thorbjørn Clausen
*Værdifuld arkitektur
Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse*
 28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett
 29. Christian Wymann
*Tattooing
The Economic and Artistic Constitution of a Social Phenomenon*
 30. Sanne Frandsen
*Productive Incoherence
A Case Study of Branding and Identity Struggles in a Low-Prestige Organization*
 31. Mads Stenbo Nielsen
Essays on Correlation Modelling
 32. Ivan Häuser
*Følelse og sprog
Etablering af en ekspressiv kategori, eksemplificeret på russisk*
 33. Sebastian Schwenen
Security of Supply in Electricity Markets
- 2012**
1. Peter Holm Andreasen
*The Dynamics of Procurement Management
- A Complexity Approach*
 2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment
 3. Line Kirkegaard
*Konsulenten i den anden nat
En undersøgelse af det intense arbejdsliv*
 4. Tonny Stenheim
Decision usefulness of goodwill under IFRS
 5. Morten Lind Larsen
*Produktivitet, vækst og velfærd
Industrirådet og efterkrigstidens Danmark 1945 - 1958*
 6. Petter Berg
Cartel Damages and Cost Asymmetries
 7. Lynn Kahle
*Experiential Discourse in Marketing
A methodical inquiry into practice and theory*
 8. Anne Roelsgaard Obling
*Management of Emotions
in Accelerated Medical Relationships*

9. Thomas Frandsen
Managing Modularity of Service Processes Architecture
10. Carina Christine Skovmøller
*CSR som noget særligt
Et casestudie om styring og menings-skabelse i relation til CSR ud fra en intern optik*
11. Michael Tell
*Fradragsbeskæring af selskabers finansieringsudgifter
En skatteretlig analyse af SEL §§ 11, 11B og 11C*
12. Morten Holm
*Customer Profitability Measurement Models
Their Merits and Sophistication across Contexts*
13. Katja Joo Dyppel
*Beskatning af derivater
En analyse af dansk skatteret*
14. Esben Anton Schultz
*Essays in Labor Economics
Evidence from Danish Micro Data*
15. Carina Risvig Hansen
"Contracts not covered, or not fully covered, by the Public Sector Directive"
16. Anja Svejgaard Pors
*Iværksættelse af kommunikation
- patientfigurer i hospitalets strategiske kommunikation*
17. Frans Bévert
*Making sense of management with logics
An ethnographic study of accountants who become managers*
18. René Kallestrup
The Dynamics of Bank and Sovereign Credit Risk
19. Brett Crawford
*Revisiting the Phenomenon of Interests in Organizational Institutionalism
The Case of U.S. Chambers of Commerce*
20. Mario Daniele Amore
Essays on Empirical Corporate Finance
21. Arne Stjernholm Madsen
The evolution of innovation strategy Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008
22. Jacob Holm Hansen
*Is Social Integration Necessary for Corporate Branding?
A study of corporate branding strategies at Novo Nordisk*
23. Stuart Webber
Corporate Profit Shifting and the Multinational Enterprise
24. Helene Ratner
*Promises of Reflexivity
Managing and Researching Inclusive Schools*
25. Therese Strand
The Owners and the Power: Insights from Annual General Meetings
26. Robert Gavin Strand
In Praise of Corporate Social Responsibility Bureaucracy
27. Nina Sormunen
*Auditor's going-concern reporting
Reporting decision and content of the report*
28. John Bang Mathiasen
*Learning within a product development working practice:
- an understanding anchored in pragmatism*
29. Philip Holst Riis
Understanding Role-Oriented Enterprise Systems: From Vendors to Customers
30. Marie Lisa Dacanay
*Social Enterprises and the Poor
Enhancing Social Entrepreneurship and Stakeholder Theory*

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge
 32. Henrik Barslund Fosse
Empirical Essays in International Trade
 33. Peter Alexander Albrecht
*Foundational hybridity and its reproduction
Security sector reform in Sierra Leone*
 34. Maja Rosenstock
*CSR - hvor svært kan det være?
Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi*
 35. Jeanette Rasmussen
*Tweens, medier og forbrug
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug*
 36. Ib Tunby Gulbrandsen
*'This page is not intended for a US Audience'
A five-act spectacle on online communication, collaboration & organization.*
 37. Kasper Aalling Teilmann
Interactive Approaches to Rural Development
 38. Mette Mogensen
*The Organization(s) of Well-being and Productivity
(Re)assembling work in the Danish Post*
 39. Søren Friis Møller
*From Disinterestedness to Engagement
Towards Relational Leadership In the Cultural Sector*
 40. Nico Peter Berhausen
Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks
 41. Balder Onarheim
*Creativity under Constraints
Creativity as Balancing 'Constrainedness'*
 42. Haoyong Zhou
Essays on Family Firms
 43. Elisabeth Naima Mikkelsen
*Making sense of organisational conflict
An empirical study of enacted sense-making in everyday conflict at work*
- 2013**
1. Jacob Lyngsie
Entrepreneurship in an Organizational Context
 2. Signe Groth-Brodersen
*Fra ledelse til selvet
En socialpsykologisk analyse af forholdet imellem selvedelse, ledelse og stress i det moderne arbejdsliv*
 3. Nis Høyrup Christensen
Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China
 4. Christian Edelvold Berg
*As a matter of size
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS*
 5. Christine D. Isakson
*Coworker Influence and Labor Mobility
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry*
 6. Niels Joseph Jerne Lennon
*Accounting Qualities in Practice
Rhizomatic stories of representational faithfulness, decision making and control*
 7. Shannon O'Donnell
*Making Ensemble Possible
How special groups organize for collaborative creativity in conditions of spatial variability and distance*

8. Robert W. D. Veitch
Access Decisions in a Partly-Digital World
Comparing Digital Piracy and Legal Modes for Film and Music
9. Marie Mathiesen
Making Strategy Work
An Organizational Ethnography
10. Arisa Shollo
The role of business intelligence in organizational decision-making
11. Mia Kaspersen
The construction of social and environmental reporting
12. Marcus Møller Larsen
The organizational design of offshoring
13. Mette Ohm Rørdam
EU Law on Food Naming
The prohibition against misleading names in an internal market context
14. Hans Peter Rasmussen
GIV EN GED!
Kan giver-idealiteter forklare støtte til velgørenhed og understøtte relationsopbygning?
15. Ruben Schachtenhaufen
Fonetisk reduktion i dansk
16. Peter Koerver Schmidt
Dansk CFC-beskatning
I et internationalt og komparativt perspektiv
17. Morten Froholdt
Strategi i den offentlige sektor
En kortlægning af styringsmæssig kontekst, strategisk tilgang, samt anvendte redskaber og teknologier for udvalgte danske statslige styrelser
18. Annette Camilla Sjørup
Cognitive effort in metaphor translation
An eye-tracking and key-logging study
19. Tamara Stucchi
The Internationalization of Emerging Market Firms: A Context-Specific Study
20. Thomas Lopdrup-Hjorth
"Let's Go Outside": The Value of Co-Creation
21. Ana Alačovska
Genre and Autonomy in Cultural Production
The case of travel guidebook production
22. Marius Gudmand-Høyer
Stemningssindssygdommenes historie i det 19. århundrede
Omtydningen af melankolien og manien som bipolære stemningslidelser i dansk sammenhæng under hensyn til dannelsen af det moderne følelseslivs relative autonomi.
En problematiserings- og erfarings-analytisk undersøgelse
23. Lichen Alex Yu
Fabricating an S&OP Process
Circulating References and Matters of Concern
24. Esben Alfort
The Expression of a Need
Understanding search
25. Trine Pallesen
Assembling Markets for Wind Power
An Inquiry into the Making of Market Devices
26. Anders Koed Madsen
Web-Visions
Repurposing digital traces to organize social attention
27. Lærke Højgaard Christiansen
BREWING ORGANIZATIONAL RESPONSES TO INSTITUTIONAL LOGICS
28. Tommy Kjær Lassen
EGENTLIG SELVLEDELSE
En ledelsesfilosofisk afhandling om selvledelsens paradoksale dynamik og eksistentielle engagement

29. Morten Rossing
Local Adaption and Meaning Creation in Performance Appraisal
 30. Søren Obed Madsen
*Lederen som oversætter
Et oversættelsesteoretisk perspektiv på strategisk arbejde*
 31. Thomas Høgenhaven
*Open Government Communities
Does Design Affect Participation?*
 32. Kirstine Zinck Pedersen
*Failsafe Organizing?
A Pragmatic Stance on Patient Safety*
 33. Anne Petersen
*Hverdagslogikker i psykiatrisk arbejde
En institutionsetnografisk undersøgelse af hverdagen i psykiatriske organisationer*
 34. Didde Maria Humle
Fortællinger om arbejde
 35. Mark Holst-Mikkelsen
*Strategieksekvering i praksis
– barrierer og muligheder!*
 36. Malek Maalouf
*Sustaining lean
Strategies for dealing with organizational paradoxes*
 37. Nicolaj Tofte Brenneche
*Systemic Innovation In The Making
The Social Productivity of Cartographic Crisis and Transitions in the Case of SEEIT*
 38. Morten Gylling
*The Structure of Discourse
A Corpus-Based Cross-Linguistic Study*
 39. Binzhang YANG
*Urban Green Spaces for Quality Life
- Case Study: the landscape architecture for people in Copenhagen*
 40. Michael Friis Pedersen
*Finance and Organization:
The Implications for Whole Farm Risk Management*
 41. Even Fallan
Issues on supply and demand for environmental accounting information
 42. Ather Nawaz
*Website user experience
A cross-cultural study of the relation between users' cognitive style, context of use, and information architecture of local websites*
 43. Karin Beukel
The Determinants for Creating Valuable Inventions
 44. Arjan Markus
*External Knowledge Sourcing and Firm Innovation
Essays on the Micro-Foundations of Firms' Search for Innovation*
- 2014**
1. Solon Moreira
Four Essays on Technology Licensing and Firm Innovation
 2. Karin Strzeletz Ivertsen
*Partnership Drift in Innovation Processes
A study of the Think City electric car development*
 3. Kathrine Hoffmann Pii
Responsibility Flows in Patient-centred Prevention
 4. Jane Bjørn Vedel
*Managing Strategic Research
An empirical analysis of science-industry collaboration in a pharmaceutical company*
 5. Martin Gylling
*Processuel strategi i organisationer
Monografi om dobbeltheden i tænkning af strategi, dels som vidensfelt i organisationsteori, dels som kunstnerisk tilgang til at skabe i erhvervsmæssig innovation*

6. Linne Marie Lauesen
*Corporate Social Responsibility in the Water Sector:
How Material Practices and their Symbolic and Physical Meanings Form a Colonising Logic*
7. Maggie Qiuzhu Mei
*LEARNING TO INNOVATE:
The role of ambidexterity, standard, and decision process*
8. Inger Hædt-Rasmussen
*Developing Identity for Lawyers
Towards Sustainable Lawyering*
9. Sebastian Fux
Essays on Return Predictability and Term Structure Modelling
10. Thorbjørn N. M. Lund-Poulsen
Essays on Value Based Management
11. Oana Brindusa Albu
*Transparency in Organizing:
A Performative Approach*
12. Lena Olaison
Entrepreneurship at the limits
13. Hanne Sørum
*DRESSED FOR WEB SUCCESS?
An Empirical Study of Website Quality in the Public Sector*
14. Lasse Folke Henriksen
*Knowing networks
How experts shape transnational governance*
15. Maria Halbinger
*Entrepreneurial Individuals
Empirical Investigations into Entrepreneurial Activities of Hackers and Makers*
16. Robert Spliid
Kapitalfondenes metoder og kompetencer
17. Christiane Stelling
*Public-private partnerships & the need, development and management of trusting
A processual and embedded exploration*
18. Marta Gasparin
Management of design as a translation process
19. Kåre Moberg
*Assessing the Impact of Entrepreneurship Education
From ABC to PhD*
20. Alexander Cole
*Distant neighbors
Collective learning beyond the cluster*
21. Martin Møller Boje Rasmussen
*Is Competitiveness a Question of Being Alike?
How the United Kingdom, Germany and Denmark Came to Compete through their Knowledge Regimes from 1993 to 2007*
22. Anders Ravn Sørensen
*Studies in central bank legitimacy, currency and national identity
Four cases from Danish monetary history*
23. Nina Bellak
*Can Language be Managed in International Business?
Insights into Language Choice from a Case Study of Danish and Austrian Multinational Corporations (MNCs)*
24. Rikke Kristine Nielsen
*Global Mindset as Managerial Meta-competence and Organizational Capability: Boundary-crossing Leadership Cooperation in the MNC
The Case of 'Group Mindset' in Solar A/S.*
25. Rasmus Koss Hartmann
*User Innovation inside government
Towards a critically performative foundation for inquiry*

26. Kristian Gylling Olesen
Flertydig og emergerende ledelse i folkeskolen
Et aktør-netværksteoretisk ledelsesstudie af politiske evalueringsreformers betydning for ledelse i den danske folkeskole
 27. Troels Riis Larsen
Kampen om Danmarks omdømme 1945-2010
Omdømmearbejde og omdømmepolitik
 28. Klaus Majgaard
Jagten på autenticitet i offentlig styring
 29. Ming Hua Li
Institutional Transition and Organizational Diversity: Differentiated internationalization strategies of emerging market state-owned enterprises
 30. Sofie Blinkenberg Federspiel
IT, organisation og digitalisering: Institutionelt arbejde i den kommunale digitaliseringsproces
 31. Elvi Weinreich
Hvilke offentlige ledere er der brug for når velfærdstænkningen flytter sig – er Diplomuddannelsens lederprofil svaret?
 32. Ellen Mølgaard Korsager
Self-conception and image of context in the growth of the firm – A Penrosian History of Fiberline Composites
 33. Else Skjold
The Daily Selection
 34. Marie Louise Conradsen
The Cancer Centre That Never Was
The Organisation of Danish Cancer Research 1949-1992
 35. Virgilio Failla
Three Essays on the Dynamics of Entrepreneurs in the Labor Market
 36. Nicky Nedergaard
Brand-Based Innovation
Relational Perspectives on Brand Logics and Design Innovation Strategies and Implementation
 37. Mads Gjedsted Nielsen
Essays in Real Estate Finance
 38. Kristin Martina Brandl
Process Perspectives on Service Offshoring
 39. Mia Rosa Koss Hartmann
In the gray zone
With police in making space for creativity
 40. Karen Ingerslev
Healthcare Innovation under The Microscope
Framing Boundaries of Wicked Problems
 41. Tim Neerup Thomsen
Risk Management in large Danish public capital investment programmes
- 2015**
1. Jakob Ion Wille
Film som design
Design af levende billeder i film og tv-serier
 2. Christiane Mossin
Interzones of Law and Metaphysics
Hierarchies, Logics and Foundations of Social Order seen through the Prism of EU Social Rights
 3. Thomas Tøth
TRUSTWORTHINESS: ENABLING GLOBAL COLLABORATION
An Ethnographic Study of Trust, Distance, Control, Culture and Boundary Spanning within Offshore Outsourcing of IT Services
 4. Steven Højlund
Evaluation Use in Evaluation Systems – The Case of the European Commission

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
*Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer
ved nordiske virksomhedsetableringer i Kina*

1999

3. Mogens Bjerre
*Key Account Management of Complex
Strategic Relationships
An Empirical Study of the Fast Moving
Consumer Goods Industry*

2000

4. Lotte Darsø
*Innovation in the Making
Interaction Research with heterogeneous
Groups of Knowledge Workers
creating new Knowledge and new
Leads*

2001

5. Peter Hobolt Jensen
*Managing Strategic Design Identities
The case of the Lego Developer
Network*

2002

6. Peter Lohmann
*The Deleuzian Other of Organizational
Change – Moving Perspectives of the
Human*
7. Anne Marie Jess Hansen
*To lead from a distance: The dynamic
interplay between strategy and
strategizing – A case study of the
strategic management process*

2003

8. Lotte Henriksen
*Videndeling
– om organisatoriske og ledelsesmæssige
udfordringer ved videndeling i
praksis*
9. Niels Christian Nickelsen
*Arrangements of Knowing: Coordinating
Procedures Tools and Bodies in
Industrial Production – a case study of
the collective making of new products*

2005

10. Carsten Ørts Hansen
*Konstruktion af ledelsesteknologier og
effektivitet*

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
*Endeavoring to Understand Market
Orientation – and the concomitant
co-mutation of the researched, the
researcher, the research itself and the
truth*

2009

1. Torkild Leo Thellefsen
*Fundamental Signs and Significance
effects
A Semeiotic outline of Fundamental
Signs, Significance-effects, Knowledge
Profiling and their use in Knowledge
Organization and Branding*
2. Daniel Ronzani
*When Bits Learn to Walk Don't Make
Them Trip. Technological Innovation
and the Role of Regulation by Law
in Information Systems Research: the
Case of Radio Frequency Identification
(RFID)*

2010

1. Alexander Carnera
*Magten over livet og livet som magt
Studier i den biopolitiske ambivalens*