

Turning innovation into evidence-based policies: lessons learned from free commune experiments

Jostein Askim*, Ulf Hjelmar** and Lene Holm Pedersen***

* Department of Political Science, University of Oslo; ** The Danish Centre of Applied Social Science (VIVE); *** Department of Business and Politics, Copenhagen Business School and The Danish Centre of Applied Social Science (VIVE)

Address for correspondence:

Jostein Askim, Department of Political Science, University of Oslo, P.O. Box 1097 Blindern, 0317 Oslo, Norway. E-mail: jostein.askim@stv.uio.no

Abstract

There is a great need for innovation in the public sector and for improved understanding of how innovation can be turned into evidence-based public policy. This article discusses potential contributions from 30 years of experience in the Nordic countries with a reform programme called Free Commune Experiments (FCEs). Here, local authorities are granted waivers from legislation to experiment with normally 'illegal' practices. The article asks if the FCEs have managed to synthesise local governments' potential to generate novel practices with central government's potential to evaluate, verify and disseminate effective innovations. The conclusion is that the FCEs have been effective generators of policy ideas, but that limited knowledge has been produced about the effects of innovative practices and whether or not new policies should be replicated beyond 'test sites'. With some changes, it is argued that the programme could nonetheless be a strategy worth emulating beyond the Nordic countries.

Keywords: Policy innovation, policy experiments, evaluation, pilot studies

Introduction

Many governments have launched national programmes aimed at enhancing innovation in public services and regulation (National Audit Office 2000; Lengrand et al. 2006; Hartley et al. 2013; Hansson et al. 2014; Gieske et al. 2016), and international organisations, such as the Organisation for Economic Co-operation and Development (OECD), have called for more public sector innovation (OECD 2015). Public sector innovation involves the generation, practical adoption and dissemination of new and creative ideas – ideas that can lead to a qualitative change in the public sector (Sørensen and Torfing 2011: 849). A large body of academic research has investigated the antecedents and consequences of innovation in the public sector (Mele et al. 2014; Maroulis and Wilensky 2015; Kiefer et al. 2015; Hartley et al. 2013; Fishenden and Thompson 2013; Walker 2006). Several noteworthy contributions have studied diffusion – how and why innovative practices

spread from site to site (Berry and Berry 1991, 1999; Rogers 2003; DiMaggio and Powell 1991; Teodoro 2009; Mizuchi and Fein 1999). However, diffusion is beneficial only if the policy ideas and practices in question are effective (Dolowitz and Marsh 2000). Therefore, in addition to facilitating diffusion, effective public sector innovation programmes need to incorporate evaluating, testing and verifying the effects of the practices in question (Arnold 2004; Lengrand et al. 2006; Word et al. 2011).

At this point, public sector innovation is encountering two barriers. One is the lack of connection between innovation and evaluation (Arnold 2004; Lengrand et al. 2006; Word et al. 2011; European Commission 2012; OECD 2015). ‘Traditional evaluation methods’, says Burt Perrin, ‘punish innovation and those who explore the unknown . . . and reward mediocrity’ (Perrin 2002: 13). ‘Evaluation is good at pinpointing successes and shortfalls and identifying where changes must be made’ says Carol Weiss, ‘but it is not always the best source of fresh ideas for tomorrow’ (Weiss 1998: 28). A second barrier to public sector innovation is particularly acute in systems where the responsibility for public service delivery has been delegated to the local government level. In such systems, say Hill and Hupe (2014: 147), one needs to think of policies as co-formed, and not only co-produced by central and local levels of government. Ministries and state agencies are institutionally responsible for national policies and have, due to reporting and inspection arrangements, a good overview of existing local policy practices. They tend, however, to lack proximity to street-level reality, which is necessary to generate practices that are more effective than the existing ones.

Local governments, in contrast, are skilled at situated knowledge production and generating new policy practices. Furthermore, they tend to draw lessons from context-sensitive, vivid depictions of experiences from other local governments seen as successful and modern; however, such second-order learning is error-prone. Although municipalities are cognizant of what goes on in nearby municipalities (Bhatti et al. 2011), the most effective and relevant policy solutions are not necessarily practiced there. In short, municipalities are good at generating ideas and less effective at verifying and documenting their effects. For central government, it is to some extent the other way around; its oversight, authority and resources enable it to evaluate policies and ensure that effective ones are widely implemented. Detachment from the street level means, however, that central government decision makers cannot easily generate policy ideas themselves.

This article studies a reform programme that has aimed to overcome these barriers to innovation in the public sector: the Free Commune Experiment (FCE). This reform programme has so far largely gone under the radar of international scholarship on public sector reform and policy innovation (but see Baldersheim and Stava 1996; Rose 1994; Loughlin and Martin 2005: 20). This is surprising, since the FCE programme has been implemented six times in Sweden, Denmark, Norway and Finland since the mid-1980s and has features that should be replicable in systems beyond the Nordic ones. The FCE is a joint central–local government reform programme that starts from the premise that the generation, testing, and diffusion of policy innovation takes place at the interface between central steering and local autonomy. Importantly, and contrary to most trial-and-error and piloting programmes (Sanderson 2002), the FCE includes legal procedures for granting temporary waivers from laws that regulate the conduct of local affairs. This provides municipalities with the space necessary for experimenting with new, normally ‘illegal’ practices. Furthermore, the FCE consciously aims to connect innovation and evaluation, with the latter an integral part of the programme’s logic. Local experiments are monitored and evaluated to provide a basis for the

further development of policies and reforms of central–local relations (Rose 1990; Baldersheim and Ståhlberg 1994; Baldersheim and Stava 1996).

The general purpose of this article is to increase scholarly knowledge about joint central–local government reform programmes’ ability to facilitate public sector innovation. Specifically, we ask whether the FCEs, as a practically and analytically interesting exemplar of such reforms, have managed to synthesise local governments’ (presumed) strengths to generate novel practices with central government’s (presumed) ability and resources to evaluate, verify and disseminate effective innovations. A necessary requirement for such synthetisation, we suggest, is that local policy experiments are evaluated in a way that is consistent with the nature of innovation (Perrin 2002; Patton 2011), while facilitating, in later stages, the dissemination of verified effective policies. Hence, this article pays particular attention to the evaluation element in the FCEs.

In a sense, the case under study is a most likely one for such synthetisation. The FCE programme, which is tailored for collaborative innovation, has been implemented in the Nordic countries – a setting where a logic of trial-and-error is integrated into the political-administrative culture (Olsen 1996; Olsen and Lindøe 2004), and where policymaking processes are characterised by strong emphasis on rationality and the inclusion of expertise (Pettersen 2016). Moreover, there has been scope for experience-based improvement in the design of FCEs: the programme has been implemented six times over three decades in countries that are very similar in terms of political context and political culture, and which have a long history of learning from each other. This, we would argue, makes possible a meaningful comparison of the cases covering four different decades (1970s–2010s).

We take as our units of observation the six national FCE programmes (and not the several hundred municipality-level FCE experiences). To facilitate a comparison across the cases and to situate the FCEs vis-à-vis existing research on evaluation, evidence-based policymaking and innovation, in the next section, we conceptualise two approaches to evaluating innovation — impact-focused and systemic. Next, we revisit the five FCEs carried out since the 1980s and the latest, still ongoing one in Denmark (2016–2020). In the final sections of the article, we return to the question of how joint central–local government reform programmes can overcome barriers to public sector innovation. We also discuss why, in light of its mixed legacy and results, the FCE reform programme is a recurring feature in Nordic central–local government relations.

Impact-focused and systemic approaches to evaluating policy innovation

By approaches to evaluating policy innovation, we mean basic theoretical understandings about how to produce valid knowledge from a complex and ever-changing reality. We build on existing research and conceptualise two archetypical approaches to evaluating policy innovations — impact-focused and systemic. The impact-focused approach involves knowledge production through rigorous evaluation methods and experiments. This approach focuses on the goal (or purpose) of the innovation process. The claim is that we have to know what the goal is before we can measure goal achievement and determine whether the innovation works (Halpern 1998). To evaluate the impact of an intervention, we must have some form of theory of why the intended impact will occur.

Logic models (programme theory) derived from realistic evaluations can help us to identify what forms of impact are appropriate to look for at the early and later stages in a project cycle

(Chen 1996; Pawson and Tilley 1997). Logic models outline inputs and activities and their anticipated outputs and outcomes. Thus, the logic model helps identify evaluation measures and indicators. Findings from this identification process can subsequently be fed back to the innovation process in order to adapt and refine the intervention itself (Keller et al. 2009).

The problem with this exercise is that the intended ways in which outputs lead to outcomes are blurred in innovative processes. The evaluator, however, can help. Evaluators can work with participants in the innovative project to strengthen the programme theory, make elements of the programme measurable, and conduct research during the innovative process to test the linkages between outputs and outcomes (Halpern 1998).

The systemic approach to evaluation takes its departure in the challenges faced when dealing with the open-ended nature of innovation and the goal complexity of innovation processes.ⁱ As Halpern has stated: ‘In the absence of pre-set goals, there is no point to evaluation’ (Halpern 1998: 6). Thus, evaluation must acknowledge the complex, iterative and non-linear nature of innovation processes. Evaluations of innovation processes cannot be meticulously planned but must be (re)shaped as the intervention develops. Evaluation activities need to evolve as events unfold and understandings of the intervention develop (Olsen and Lindøe 2004; Boyd et al. 2007; Patton 2011, Hjelmar and Pedersen 2015).

The systemic approach has intellectual roots in communities of practice literature and the associated situated learning theory, which asserts that ‘learning is not only a cognitive activity, but is also a social, situated process realized in the participation in a given practice around which a community (of practitioners) emerges’ (Contu 2008: 210). Studies have shown that municipalities are able to learn from each other without centralised instruction on whom (entities) and what (practices) to draw lessons from (Dahl and Hansen 2006; Bhatti et al. 2011). Such learning can vary from quick imitation to rigorous analytical efforts to decode the experiences of others and recode or translate them into one’s own practices (Baldersheim and Rose 2014; see also Rose 1991; Røvik 2007).

The systemic approach is not experimental but heuristic: knowledge and evidence emanate from practical problem solving, learning and discovery. Knowledge is seen as ‘a living, often tacit process’; it is not ‘reduced to information – to a “thing” that is abstract, formalized, and can be stored in databases’ (Contu 2008: 212). While the impact-focused approach sees knowledge as decoupled from context, the systemic approach sees knowledge as situated in context. Criteria such as internal and external validity are not central in the systemic approach. Knowledge need not be generalisable to a larger population to be valid; knowledge is valid as it resides in the sites, e.g. organisations, which have undertaken the experience.

Given the differing views of these approaches on the processes of discovery and the criteria for valid knowledge, their views on the functional requirements for successful knowledge production also differ. In the impact-focused approach, successful knowledge production requires considerable control over the experiment. In the systemic approach, successful knowledge production is dependent upon identifying and nurturing innovative and entrepreneurial actors – individuals, groups or organisations (Contu 2008). A systemic approach to evaluating innovation must be a real-time evaluation following the project in detail throughout its operation. Such studies offer ‘thick’ narratives where attempts at verifying knowledge are delimited to context-sensitive analyses.

Summing up, the impact-focused approach has its main strength in testing and verifying the effects of local innovative processes. Through testing and verification, experiments produce valid knowledge about the effects of interventions. The main strength of the systemic approach, by contrast, is in discovering and generating new ideas (see Table 1).

Table 1: Two basic approaches to evaluate public sector innovation

	Impact-focused approach	Systemic approach
Purpose	Solid knowledge about effects and causality	Generate new practices; nurture innovation and learning
Participants	Controlled selection, few in number, randomisation, strategic sampling, control group	Self-selection, many participants, no control group
Contents	Narrow scope and accommodation	Wide scope; emergent, liberal accommodation
Evaluation	Comprehensive, external party, effect-oriented	Inspiring depictions, context-sensitive narratives, process-oriented

Methods and data

Four FCEs were carried out in the 1980s and 1990s, with Sweden initiating its project first (starting in 1984), followed by Denmark (1985), Norway (1986) and Finland (1988). All four were extended at least once, with each FCE lasting six years or more. The final two exemplars are the second (2012–2015) and third Danish FCEs (2016–2020) (see Table 2).

Table 2: Free commune experiments: Overview

	Coordinator	Participating municipalities (of total)	Comprehensive evaluation
Sweden 1984–1992	Ministry of Public Administration	39 (of 284)	Yes
Denmark 1985–1993	Ministry of the Interior	63 (of 275)	No
Norway 1986–1992	Ministry of Local Government	26 (of 443)	Yes
Finland 1988–1996	Ministry of the Interior	56 (of 461)	No
Denmark 2012–2015	Ministry of Economic Affairs and the Interior	9 (of 98)	No
Denmark 2016–2020	Ministry of Economic Affairs and the Interior	44 (of 98)	No

This article takes each national FCE programme experience as a unit of observation and analyses general aspects of the evaluation practice. Referring to the evaluation theorist Evert Vedung (1997), we focus on three issues:

- (1) The general problem question: What is the purpose of evaluation?
- (2) The organisation question: Who performs the evaluation?
- (3) The impact question: How are cause-effect attribution issues handled?

These issues are, in reality, mixed up with each other. The answer to the general problem question is intertwined with the answer to the organisation question. For some purposes, external and independent evaluation is key, while others may necessitate local embeddedness and heavy stakeholder involvement. The approaches also differ in their views on how best to control participation in and the contents of policy experiments. We consider the programmes' handling of participation and content issues as their answer to the impact question.

In our analysis of FCEs from the 1980s and 1990s, we extensively reviewed existing research – sources that rely on elite and participant interviews, document studies and surveys. We also reviewed official documents (including and beyond ones referenced in Rose's (1989) bibliography on FCEs). In our analysis of the Danish FCE carried out from 2012 to 2015, we studied official documents, FCE project applications, participant observations conducted at a ministerial FCE conference in January 2012 (Ministry of Economic Affairs and the Interior 2012) and results from mid-term and final evaluation reports (Rambøll Consulting 2014, 2016). In analysing the ongoing Danish FCE (2016–2020), we used official documents (e.g. the Free Commune Evaluation Act), FCE project applications and descriptions of FCE evaluation designs.

An analysis with local FCE experiments as the units of analysis would require access to documentation that does not exist today: several hundred evaluation reports that have been produced locally since 1984 (since evaluation has been largely delegated to the municipalities themselves; see below). Moreover, we have not analysed whether policies originating from local FCEs have been implemented in other municipalities, nor have we studied the short- and long-term effects of diffused policies. We provide illustrative examples, but a systematic analysis would require in-depth and extremely complex evaluations far beyond the scope of this article.

The Nordic free commune experiments revisited

Sweden 1984–1992

The Swedish Free Commune Act was passed by *Riksdagen* in 1984 as part of a reform programme responding to societal criticism about excessive bureaucratisation and state regulation (Swedish Government 1991; Lindgren and Strömberg 1994; Rose 1990; Baldersheim and Ståhlberg 1994). Municipalities were invited to experiment with new ways of organising political committees, and to apply for exemptions from almost any rule or regulation and to insert any intervention in its place. Special priority was given, however, to proposals with projected results that could be widely adopted. No control group was established.

In 1984, 12 municipalities were admitted to the programme. These municipalities submitted 280 applications. Following that, 12 more municipalities were admitted in 1988, and another 15 in 1990. Participating municipalities were not obliged to evaluate their own experiments. In 1986, the FCE secretariat and line ministries agreed to entrust evaluation of the FCE to the University of Gothenburg (Strömberg 1986), and in 1990, the government appointed its own evaluator. Overall, external observers judged the evaluation of the Swedish FCE as patchy. The design of the programme reduced its evaluability; it proved difficult to isolate the impact of experiments carried out in a programme spanning eight years. During this period, a range of events and changes took place. As a result, distinguishing 'signal from noise' proved to be nearly impossible (Lindgren and Strömberg 1994).

As a result of the FCEs, the government was reassured that it was a good idea to increase municipalities' discretion on how to organise local political committees. Rules that restricted committee flexibility were therefore permanently lifted in 1992. Additionally, several further changes to Swedish central–local relations were inspired or informed by the FCE experiences. For example, devolution of responsibilities in the educational sector in the 1990s was informed by positive FCE experiences from granting municipalities more freedom in running schools. In addition, the FCE was an important input in the process of preparing Sweden's new Local Government Act in 1991. However, later political decisions did not draw heavily or explicitly on evaluations of the FCE, and observers have struggled to find direct linkages between the FCE and subsequent changes (Lindgren and Strömberg 1994).

Denmark 1985–1993

The Danish Free Commune Act was passed by *Folketinget* in 1985 and the FCE programme was administered by the Ministry of the Interior. The stated purposes were increased self-government, fostering local entrepreneurship and innovation, improved services and increased efficiency (Albæk 1994; Schou 1988). In addition, de-bureaucratisation had been on the agenda in the preceding years. Contrary to the Swedish case, no interventions were pre-specified; municipalities were invited to apply for waivers from almost any rule or regulation.

Twenty-seven municipalities were selected in 1987, and another 36 municipalities were selected in 1989. The selection of the participants was aimed at a geographical, structural, size-related and political spread. Selected waivers were made valid for all municipalities and not just for FCE participants. By 1992, 866 applications had been submitted. Among these, 390 applications were approved and 476 were rejected. Many municipalities found line ministries to be overly resistant to change and to display a lack of understanding of the logic of innovation (Albæk 1994; Schou 1988; Ministry of the Interior, Denmark 1988).

No comprehensive evaluation was carried out. The opposition in parliament called for an evaluation when the Free Commune Act was passed, but the Minister of the Interior refused due to scepticism that any knowledge produced would be useful for future policymaking. In addition, there were concerns that an evaluation would be costly and could produce politically skewed results (Albæk 1994). Instead, the Ministry of the Interior commissioned an evaluation of four specific FCE projects deemed particularly policy-relevant. In addition, participating municipalities were required to evaluate the experiments themselves.

The lack of a comprehensive evaluation did not prevent the diffusion of some results. Anecdotes of FCE experiences were used as arguments for changes in the laws in parliamentary discussions. In addition, line ministries followed up the FCE by giving municipalities general rights to experiment within their areas of authority (Albæk 1994: 65).

Norway 1986–1992

The Norwegian Free Commune Act was passed by *Stortinget* in 1986 and administered by the Ministry of Local Government. The purposes were to advance flexibility at the local level and coordination between the national level and local governments. Both purposes concerned governance, not the quality or efficiency of service production. Participants were invited to apply

for waivers from almost any rule or regulation (Ministry of Local Government 1986a; Norwegian Parliament 1988; Baldersheim and Fimreite 1994).

The selection of the participants for the Norwegian FCE differed somewhat from the selection in the Swedish and Danish cases. The selection process was relatively faithful to a pre-set number of participants (26 admitted, not 24 as planned). The participating municipalities submitted 260 applications for waivers, spanning 80 national acts (Ministry of Local Government 1986b). As in Sweden and Denmark, the coordinating ministry worked to broker agreements between municipalities and line ministries. Seventy experiments were eventually approved and waivers from nearly 100 legal acts were granted (Ministry of Local Government 2005: 141; Norwegian Parliament 1991).

What sets the Norwegian FCE apart from the other cases is its approach to evaluation. A comprehensive evaluation plan was launched from the start and entrusted to a consortium of research institutions led by the University of Bergen. Thirty reports and four books covering the whole programme were produced, and four projects were evaluated in depth (Rose 1989). A 1991 governmental White Paper on the FCE drew heavily on findings from this evaluation (Ministry of Local Government 1991).

In 1993, two major pieces of legislation were introduced as a direct result of the FCE – a new local government act and a new act on experimentation in public governance (*Forsøksloven*). Furthermore, experiences from the FCE projects were diffused broadly among municipalities. The FCE also had unanticipated long-term effects. For example, a major administrative reform in 2006 – one that merged the national apparatuses for employment pensions and municipal social services – had its precedent in a 1980s FCE project in the municipality of Grimstad (Fimreite 2008).

Finland 1988–1996

The Finnish *Riksdag* passed the Free Commune Act in 1988, and the FCE programme was administered by the Ministry of the Interior. Stated purposes were increased self-government and popular participation, improved services and increased efficiency (Ministry of the Interior, Finland 1988). Finland had gone through a severe economic crisis in the 1980s, and empowering municipalities to make cutbacks in public expenditures was a motivating factor for launching the FCE (Ståhlberg 1994). As in the Danish and Norwegian cases, Finnish municipalities were invited to apply for waivers from almost any rule or regulation.

Fifty-six municipalities were selected to participate in the FCE. Certain waivers were extended to all 461 municipalities in Finland, thus giving the programme features akin to a national administrative reform (Ståhlberg 1994).

No comprehensive evaluation was conducted, but reports were prepared that analysed the experiences of selected target groups and projects. Some FCE experiences were disseminated nonetheless. Among the most tangible results were a new Local Government Act, permanent deregulations in the education sector and a follow-up experimental policy programme on citizen participation. Inspired by the results of the FCE projects, many line ministries, which initially resisted experimentation within their areas of responsibility, later changed legislation (Ståhlberg 1994).

Denmark II, 2012–2015

The second Danish FCE started in 2012 and ended in 2015, born in a political context of ‘cutting red tape’ (Rambøll 2016). The FCE purported to enhance entrepreneurship among municipalities. Twenty-one municipalities applied to obtain status as free communes. Among these, nine were selected as participants, although the original plan was to admit only five in the FCE (Rambøll Consulting 2014). Of the 400 applications for specific interventions, 75% were accepted (Ministry of Economic Affairs and the Interior 2014). One observation from reading 2012’s first batch of 215 applications in the employment sector is that many applications sought to reduce excessive bureaucratisation. A second observation is that not all suggested interventions required the implementation of waivers; a third is that applications were more specific about problems (rules that needed lifting) than about solutions (practices to implement and/or interventions).

The Government’s design for the evaluation of the FCE was to have each participating municipality evaluate its own experiment. Local evaluations received state-subsidised support from KORA – the Danish Institute of Governmental Research (Ministry of Economic Affairs and the Interior 2014).

KORA was contracted to make two in-depth evaluations of specific experiments (Hjelmar and Christiansen 2016) and a consultancy firm, Rambøll Consulting, was contracted to make a mid-term and a final evaluation of the FCE (Rambøll Consulting 2014, 2016). The final evaluation used evaluations made by the municipalities and data from a survey to professionals in charge of the local experiments. Only very few among the several hundred experiments were evaluated in detail by either Rambøll Consulting or KORA. Still, the evaluators claimed that 20% of the experiments strongly influenced new legislation, most of which were in employment policy and basic education policy (Rambøll Consulting 2016).

Denmark III, 2016–2020

The third and ongoing Danish FCE started in 2016. This FCE was inspired by experiences from the preceding Danish FCEs but the focus on enhancing innovative practices on a wider scale was greater than before. The 2016 FCE is organised around networks of municipalities instead of individual municipalities. Each network has applied to conduct policy experiments within only one relatively narrow and delimited policy issue, for example, ‘cross-sectoral cooperation for elderly persons with chronic diseases’. Within such an issue, there can be several experiments, but they are all supposed to be inter-related and to support the same overall goal.

Forty-three municipalities applied to obtain status as a free commune. Eight networks were formed with all the original applicants plus one additional municipality as participants. A first set of waivers was passed in parliament in mid-2017, and a second set of waivers was passed in late 2017. Seventy-one applications for specific interventions were submitted, of which 79% were accepted. It is the responsibility of each network to make its own evaluations. The networks receive state-subsidised evaluation support from KORA. KORA will also independently evaluate selected experiments and will publish a final evaluation report when the FCE ends in 2020.

Key similarities and differences across the six cases

There are a number of key characteristics of the FCEs across the cases. First, governments have launched the FCEs with a mix of three purposes: eliminating unnecessary administrative burdens

(deregulation), enhancing innovation, and improving the quality and efficiency of public services. The outliers are Norway’s FCE, which purported to improve governance only (not the quality or efficiency of service production), and Denmark’s second FCE, which had a distinct de-bureaucratisation emphasis.

Second, the FCEs have typically included two rounds of applications. First, municipalities are invited to apply to be included in the programme. Getting exemptions from national rules and regulations is not the only motivation for municipalities to be included. It also matters that they are perceived as modernising frontrunners; the FCE status provides leverage for local leaders to push reform agendas locally. As expressed by a top executive at a mid-term conference in the Danish FCE II: ‘We need to get rid of some corrosion. It is not just central regulations that cause us trouble’.

Central government – typically a coordinating secretariat located in a ministry – assesses the proposals and selects participants. The resulting samples have tended not to be representative of the population, but rather to over-represent municipalities with a record of implementing and adapting to change.

In the second round of applications, FCE participants apply for waivers from specific national rules and describe new practices they wish to implement. Central government then assesses these proposals and approves or dismisses them according to set criteria. Each waiver tends to be made valid for all municipalities in the FCE programme, and not just for the municipality that applied for it.

The FCEs have typically stated the desire to evaluate programme experiences with the aim of providing a basis for future policymaking (Rose 1990). However, actual evaluation ambitions and practices have been limited. FCEs are characterised by their strong decentralisation of evaluation to FCE participants themselves (particularly FCEs II and III in Denmark). The Swedish FCE and the Danish FCE I are examples of how evaluation can become politicised. FCE coordinators and line ministries fought for power over evaluation (Sweden), and comprehensive external evaluation was avoided because it could produce politically unwanted results (Denmark).

Table 3 illustrates the key characteristics in the six different FCEs. As shown, there are quite substantial differences between the FCEs in terms of evaluation practices:

Table 3: Cases summarized in relation to Vedung’s basic evaluation practices

	Sweden	Denmark I	Norway	Finland	Denmark II	Denmark III
Purpose (which focus)	Input to legislation	Innovation	Improved governance	Broad	Debureaucratization	Innovation
Organization (who evaluates)	Centralized	Centrally and locally	Centralized	Centralized	Centrally and locally	Centrally and locally
Impact (which focus)	Not comprehensive	Not comprehensive	Impact-focus comprehensive	Not comprehensive	Systemic-focus comprehensive	Systemic-focus Not comprehensive

Discussion

Evaluation and innovation

Decentralising responsibility for evaluation to the local governments themselves – the most common response among FCE cases to Vedung’s (1997) ‘organisation question’ – can be

interpreted as a recipe for descriptive accounts with limited attention to external validity. Summary reports written by external parties (like in Denmark II and III) can add crosscutting observations, but as long as they use locally produced evaluations as raw material, independence is superficial. In short, local embeddedness and stakeholder co-ownership of evaluations (Amba 2004: 995) was prioritised over external and independent evaluation.

Moving on to Vedung's (1997) 'impact question', we have seen that in cases where evaluation was decentralised, the lack of causal identification in underlying reports will have constrained the handling of cause-effect attribution issues in summary reports. Moreover, national coordinators consistently allowed municipalities to self-select into the programmes, opened the programme to a high number of participants (with the exception of Denmark II), and had no concept of a control group. National coordinators also exercised limited control over which policy experiments, i.e. interventions, the programmes should cover. Coordinators essentially took a broker role and left decisions about programme contents to the municipalities themselves and to line ministries that decided which waivers to grant. The resulting broad portfolios of themes and policy interventions further reduced the 'evaluability' of the programmes – at least seen from an impact-focused evaluation approach. The chances of identifying causal effects of any single policy experiment were also reduced by allowing many simultaneous policy experiments within each participating municipality (Denmark III was a partial exception).

The FCE response to Vedung's (1997) 'general problem question' – what purpose evaluation serves – is overall more consistent with the system approach to evaluation than to the impact-focused approach. Evaluation was not used to generate solid knowledge about the causal effects of the new practices tried out by municipalities, but rather as a support structure for the generation of new practices among municipalities.

We argued in the introduction that avoiding 'punishing . . . those who explore the unknown' (Perrin 2002: 13) is not sufficient, in terms of evaluation, for synthesising local governments' strengths to generate novel practices with central government's ability and resources to evaluate, verify and disseminate effective innovations. To support innovation, evaluation must also facilitate, in later stages, the dissemination of verified effective policies. To satisfy this requirement, a joint central–local government innovation programme would need to incorporate more elements of impact-focused evaluation than the FCEs have done (we give some suggestions below).

The problem with the FCE programme, in the context of understanding policy innovation as more than just the discovery of new practices, is that claims about local successes have only a low external validity. Low external validity is particularly problematic at the national level, because it means that whatever knowledge exists about new practices offers little guidance for replication. Successful policy transfer requires some level of de-contextualisation of the domain site's experience. The policy's generic factors must be identified and understood to avoid 'uninformed policy transfer' (Dolowitz and Marsh 2000). Furthermore, the policy's isolated effects and its interaction with contextual factors at the domain site (Rose 1991) must be understood to avoid an 'incomplete' transfer (Dolowitz and Marsh 2000). Without such de-contextualisation, potential target sites will not be able to judge whether and how they can replicate the domain site's success in adopting or adapting a policy. Such explication or unpacking is difficult for the domain site, since local practitioners are situated in their own experience. Policy transfer can therefore mean that inappropriate and ineffective interventions are diffused alongside effective ones (Dolowitz and Marsh 2000).

Sustained enthusiasm despite mixed experiences

It can appear paradoxically that the FCE programme has been implemented numerous times in the Nordic countries, and especially in Denmark, despite the programme's limited ability to support policy innovation, as conceptualised in this article. To explain why FCEs were launched in the 1980s, 1990s and 2010s, Baldersheim and Rose (2014) point to the FCEs' political usefulness. The programme's experimentation and evaluation aspects correspond to criticisms that public policymaking is insufficiently evidence-based (Pawson and Tilley 1997; Andersen et al. 2011; Bhatti et al. 2006; Kongsted and Konnerup 2011) and to criticisms that the public sector has not been sufficiently innovative (Campbell and Pedersen 2014). Moreover, the deregulation aspect of the programme speaks to anti-bureaucratic sentiments and red-tape research that criticise state regulations for being excessively controlling, thereby harming employee motivation and effective service provision (Bozeman and Feeney 2011; Frey and Jegen 2001; Andersen et al. 2015). Launching an FCE programme enables governments to respond simultaneously to different criticisms of the public sector. All can find something they like in the FCE: conservative parties who want to cut bureaucracy, leftist parties who want to improve the quality of public services, municipalities who want more room for innovation at the local level, and state officials who demand more effective practices at the local level.

The FCE can also be a useful pawn in a game of blame-avoidance. In this game, municipalities cannot easily blame central government for increasingly detailed central regulation, and central government can in return challenge municipalities to cite specific examples of rules that should be changed (which may prove to be a more complex matter than municipalities anticipate). However, delivering on a multitude of potentially contradictory promises is difficult. The actual results of FCEs can hardly please all interested parties. Therefore, perhaps unsurprisingly, evaluation has been – and is still – downplayed, decentralised, and subjected to power plays.

How to design better programmes

Despite mixed experiences, it is worthwhile to build on the FCE when designing joint central–local government innovation programmes. The FCE has been dubbed a 'particular Nordic approach to modernising local government' (Ministry of Local Government, Norway 2005: 140; see also Baldersheim and Ståhlberg 1994: 4). There is probably some truth to presenting the FCE as a method characterised by '... a [non-confrontational] partnership relationship between central and local government' (op. cit.). However, we believe the underlying approach, with methodological adaptations, can fit different national and situational contexts.

A clear break with the systemic approach to evaluation does not appear to be a good choice. It is reasonable to assume that retaining elements of systemic evaluation is necessary to unleash the enthusiasm at the municipality level that have given energy and local co-ownership in the case of FCEs. Still, stronger elements of impact-focused evaluation should be incorporated. For example, future programmes could have a clearer *sequential* nature, where new policy practices are generated in a first step, followed by a second step where practices are treated as interventions whose effects are tested and verified with carefully selected participants and narrowly defined scopes for problems and solutions. Discovery and verification can also be integrated *simultaneously*. One tier of the programme could be open-ended for problems and solutions and focus on innovative

processes sensitive to local contexts. A second tier could be narrower in scope, emphasising an impact-focused evaluation approach. The challenge is to integrate these two tiers continually so they can inform and enrich each other.

Conclusions

FCEs provide participants with the opportunity to experiment with new policy solutions within boundaries that are wider than usual (Rose 1990), and they have unleashed thousands of local policy experiments across the Nordic countries since 1984. We have shown that the FCEs have been better as generators of new policy ideas than as test beds for the effects of interventions. Limited knowledge has been produced about the effects of innovative practices, and about whether new policies should be replicated in other municipalities. Ideas and interventions generated in the experiments have nonetheless spread among municipalities, not only through voluntary horizontal diffusion but also via vertical diffusion; central government agencies have been inspired by FCEs and, in a number of cases, new policies and legislation have been the result. This is not a strategy that strongly connects policymaking to evidence or which leads to evidence-based policymaking.

The limitations of the FCE strategy were clarified by interpreting the FCE vis-à-vis two general approaches to evaluation: systemic and impact-focused. The FCE programme has taken its design principles predominantly from the systemic approach. Sculpting a policy innovation programme on this approach does carry the potential for discovering and developing new practices at the local level, but it hardly enables national-level policymakers to enact rules and regulations that promote policy practices with proven effectiveness. Often, evaluation designs of FCEs have fragmented into process evaluations in individual municipalities. Generating ideas and new practices is a necessary but insufficient condition for public sector innovation when one understands innovation as a process that also involves verification and diffusion.

There is, however, an untapped potential for integrating the discovery, testing and diffusion phases of policy innovation in the FCE. The strengths of the systemic and impact-focused approaches to evaluation can be combined and their weaknesses minimised through various changes discussed in the article. Even if the bottom-up approach, which stresses learning on the local level, has been made the default for the FCE, elements of an impact-focused approach that stresses general learning about the effects could also be incorporated, especially when selecting participants for the experiments and for reducing the number of themes and interventions. Such a synthesis holds the potential to facilitate innovation in the initial phase, and the potential to test possible effects systematically to ensure the production of more evidence-based knowledge and, hence, reduce the chances of diffusing the wrong lessons. These steps could make the FCE a programme worth emulating in and beyond the Nordic countries.

There has been called for an increased methodological rigor in public administration research to ensure effective policies and avoid policy endogeneity (Besley and Case, 2000; Blom-Hansen, Morton & Serritslew, 2015). The FCE are one way to address this. The FCE ensures that policy changes and interventions are made in response to real problems at the local level, and rigorous evaluation of the interventions inform us about real time effects.

References

- Albæk, E. (1994) 'The Danish case: rational or political change' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 41–68. Aldershot: Dartmouth.
- Amba, T. (2004) 'Responsive Evaluation: the Meaning and Special Contribution to Public Administration'. *Public Administration*, 82(4): 993–1012.
- Andersen, L., Kristensen, N., and Pedersen, L. (2015) 'Documentation requirements, intrinsic motivation, and worker absence'. *International Public Management Journal*, 18(4): 483–513.
- Andersen, O., Broegaard, E., and Klejnstrup, N. (2011) 'En effektiv udviklingsbistand: Om evidens, 'randomistas' og relevans' [Effective international aid: On evidence, 'randomistas' and relevance], *Samfundsøkonomen*, 2: 18–23
- Arnold, E. (2004) 'Evaluating research and innovation policy: a systems world needs systems evaluations'. *Research Evaluation*, 13(1): 3–17.
- Baldersheim, H. and Fimreite, A. L. (1994): 'The Norwegian case: an experiment in bottom-up learning' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 69–84. Aldershot: Dartmouth.
- Baldersheim, H. and Rose, L. (2014) *Det kommunale laboratorium: Teoretiske perspektiver på lokal politikk og organisering* [The local government laboratory: Theoretical perspectives on local politics and organization], 3rd ed, Bergen: Fagbokforlaget
- Baldersheim, H. and Stava, P. (1996) 'Free communes, pilots and pathfinders: A new vocabulary of local government reforms in Scandinavian countries – The case of Norway' in *Strategic changes and organizational reorientations in local government*, N. Ben-Elia (ed.), pp. 57–74. Houndmills: Macmillan.
- Baldersheim, H. and Ståhlberg, K (1994) 'From top-down to bottom-up: free communes and the politics of administrative modernization' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 3–20. Aldershot: Dartmouth.
- Berry, F. and Berry, W. (1991) 'Specifying a model of state policy innovation'. *American Political Science Review*, 85(2): 571–579.
- Berry, F. and Berry, W. (1999) 'Innovation and diffusion models in policy research' in *Theories of the policy process*, P. Sabatier (ed.), pp. 169–200. Boulder, CO: Westview Press.
- Besley, T. and A. Case. (2000) 'Unnatural Experiments? Estimating the Incidence of Endogenous Policies.' *Economic Journal* 110(467): F672–F694.
- Blom-Hansen, J., Morton, R., and Serritslew, S. (2015). Experiments in Public Management Research. *International Public Management Journal*, 18(2), 151-170.
- Bhatti, Y., Hansen, H., and Rieper, O. (2006) *Evidensbevægelsens udvikling organisering og arbejdsform: en kortlægningsrapport* [The development, organization and work form of the evidence movement: an overview report], København: AKF forlag.
- Bhatti, Y., Olsen, A. and Pedersen, L. H. (2011) 'Administrative professionals and the diffusion of innovations: the case of citizen service centres'. *Public Administration*, 89(2): 577–594
- Boyd, A., Geerling, T., Gregory, W. J., Kagan, C., Midgley, G., Murray, P., and Walsh, M.P. (2007) 'Systemic Evaluation: A Participative, Multi-Method Approach'. *The Journal of the Operational Research Society*, 58(10): 1306–1320.
- Bozeman, B., and Feeney, M. (2011) *Rules and red tape: A prism for public administration theory and research*. London: M. E. Sharpe

- Campbell, J., and Pedersen, O. (2014) *The National Origins of Policy Ideas: Knowledge Regimes in the United States, France, Germany, and Denmark*. Princeton, NJ: Princeton University Press
- Chen, H. T. (1996) 'A Comprehensive Typology for Program Evaluation'. *Evaluation Practice*, 17(2): 121–130.
- Contu, A. (2008) 'Communities of practice' in *International encyclopaedia of organization studies*, S. R. Clegg and J. R. Bailey (eds), pp. 210–213. London: Sage.
- Dahl, P. and Hansen, K. (2006) 'Diffusion of standards: the importance of size, region and external pressures in diffusion processes'. *Public Administration*, 84(2): 441–459.
- DiMaggio, P., and Powell, W. (1991) 'The iron cage revisited: institutional isomorphism and collective rationality in organizational fields', in *The new institutionalism in organizational analysis*, W. Powell and P. DiMaggio (eds), pp. 63–82. Chicago: University of Chicago Press.
- Dolowitz, D.P. and Marsh, D. (2000) 'Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making'. *Governance*, 13(1): 5–21.
- European Commission (2012) *Evaluation of Innovation Activities. Guidance on methods and practices*. Brussels: European Commission, Directorate for Regional Policy.
- Fimreite, A. (2008) *Mission impossible made possible? Tenkning og argumentasjon bak partnerskapet mellom stat og kommune i NAV* [(...) Ideation and argument behind the partnership between the state and local governments in the Norwegian welfare administration]. Working paper 14/2008. Bergen: Stein Rokkan Centre for Social Studies.
- Fishenden, J., and Thompson, M. (2013) 'Digital government, open architecture, and innovation: why public sector IT will never be the same again'. *Journal of Public Administration Research and Theory*, 23(4): 977–1004.
- Frey, B. and Jegen, R. (2001) 'Motivation crowding theory'. *Journal of Economic Surveys*, 15(5): 589–611.
- Gieske, H., van Buuren, A. and Bekkers, V. (2016) 'Conceptualizing public innovative capacity: A framework for assessment'. *Innovation Journal*, 21(1): 1–27.
- Halpern, G. (1998) 'From Hubris to Reality: Evaluating Innovative Programs in Public Institutions'. *The Innovation Journal: The Public Sector Innovation Journal*, 3(3).
- Hansson, F. Norn, M. T., and Vad, T. B. (2014) 'Modernize the public sector through innovation? A challenge for the role of applied social science and evaluation'. *Evaluation*, 20(2): 244–60.
- Hartley, J. (2005) 'Innovation in governance and public services: past and present'. *Public Money and Management*, 25(1): 27–34.
- Hartley, J., Sørensen, E., and Torfing, J. (2013) 'Collaborative innovation: a viable alternative to market competition and organizational entrepreneurship'. *Public Administration Review*, 73(6): 821–830.
- Hill, M. and P. L. Hupe (2014) *Implementing public policy. An introduction to the study of operational governance* (3rd ed.). London: Sage.
- Hjelmar, U. and Pedersen, P. V. (2015) 'Kompleks evaluering – tre metodiske læringspunkter fra en case' [Complex evaluation – three lessons learned from a case study]. *Metode og Forskningsdesign* [Methods and research design], 2: 32-56.
- Hjelmar, U. and Christiansen, A. M. (2016) *Tilkøbsydelser på ældreområdet – en ny mulighed for kommunerne? En evaluering af erfaringerne fra frikommunerne* [Supplementary services in eldercare: a new possibility for the municipalities? Evaluating experiences of free communes]. København: KORA.
- Keller, C. Gäre, K., Edenius, M. and Lindblad, S. (2009) *Designing for Complex Innovations in Health Care: Design Theory and Realist Evaluation Combined*. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (article 3). The Association for Computing Machinery: The ACM Digital Library. doi>10.1145/1555619.1555623

- Kiefer, T., Hartley, J., Conway, N., and Briner, R. (2015) 'Feeling the squeeze: public employees' experiences of cutback- and innovation-related organizational changes following a national announcement of budget reductions'. *Journal of Public Administration Research and Theory*, 25(4): 1279–1305.
- Kongsted, H. and Konnerup, M. (2011) 'At måle en effekt. Ufordringer og løsninger' [Measuring an effect. Challenges and solutions]. *Samfundsøkonomen*, 1: 11–17.
- Lengrand, L. & Associates, PREST, ANRT and Reidev Ltd (2006) *Smart Innovation: A Practical Guide to Evaluating Innovation Programmes*. Brussels-Luxembourg: ECSC-EC-EAEC.
- Lindgren, L. and Strömberg, L. (1994) 'The Swedish case: Political change and administrative resistance' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 21–40. Aldershot: Dartmouth.
- Loughlin, J. P. and S. J. Martin (2005) *Options for Reforming Local Government Funding to Increase Local Streams of Funding: International Comparisons*. Discussion paper prepared for the Lyons Inquiry into Local Government Funding. London: HM Treasury.
- Maroulis, S. and Wilensky, U. (2015) 'Social and task interdependencies in the street-level implementation of innovation'. *Journal of Public Administration Research and Theory*, 25 (3): 721–750.
- Mele, V., Compagni, A. and Cavazza, M. (2014,) 'Governing through evidence: a study of technological innovation in health care'. *Journal of Public Administration Research and Theory*, 24(4): 843–877.
- Ministry of Economic Affairs and the Interior, Denmark [Økonomi- og indenrigsministeriet] (2012) *Program for workshop om udvikling og tværgående evaluering af frikommuneforsøg* [Program for workshop on development and cross-cutting evaluation of free commune experiments], København: Økonomi- og Indenrigsministeriet
- Ministry of Economic Affairs and the Interior, Denmark [Økonomi- og indenrigsministeriet] (2014), *Frikommuneforsøget* [The Free Commune Experiment] web-page, <http://oim.dk/arbejdsomraader/kommunal-og-regionaloekonomi/frikommuneforsog-og-udfordringsret/frikommuneforsog.aspx> (uploaded 22.12.2014)
- Ministry of Local Government, Norway [Kommunaldepartementet] (1986a) *Forsøksvirksomhet i kommunene og fylkeskommunene – frikommuner* [Pilot schemes in local and regional governments – free communes]. Rundskriv [circular letter] h-14/86. Oslo: Kommunal- og arbeidsdepartementet
- Ministry of Local Government, Norway [Kommunaldepartementet] (1986b) *Frikommunesøknader – samlet fremstilling og innholdsfortegnelse* [Free commune experiment applications – full overview and table of contents]. Oslo: Kommunal- og arbeidsdepartementet
- Ministry of Local Government, Norway [Kommunaldepartementet] (1991) St.meld. nr. 38 (1990-91) *Frikommuneforsøket* [White Paper on the free commune experiment]. Oslo: Kommunal- og arbeidsdepartementet
- Ministry of Local Government, Norway [Kommunaldepartementet] (2005) NOU 2005:6. *Samspill og tillit. Om staten og lokaldemokratiet (Lokaldemokratikommisjonen)* [Norwegian official report 2005:6. Collaboration and trust. On the state and local democracy [The commission on local democracy]. Oslo: Kommunal- og regionaldepartementet
- Ministry of the Interior, Finland [Inriksdepartementet] (1988) *Regeringens proposition til Riksdagen med förslag til lag om försök med frikommuner* [The Government's proposition to the Riksdag for a law on free commune experiments]. Helsingfors: Inriksdepartementet/Statens tryckericentral.
- Ministry of the Interior, Denmark [Indenrigsministeriet] (1988) *Oversigt over samtlige forslag til frikommuneprojekter pr. 1/12 1988* [Overview of all proposals for free commune projects as of 1/12 1988]. København: Indenrigsministeriet.

- Mizruchi, M. and Fein, L. (1999) 'The social construction of organizational knowledge: a study of the uses of coercive, mimetic, and normative isomorphism'. *Administrative Science Quarterly*, 44 (4). 653–683.
- Norwegian Parliament [Stortinget] (1988) Innst. S. nr 161 (1988-89). *Innstilling fra kommunal- og miljøvernkomiteen om frikommuneforsøket*. [Recommendation from the parliamentary committee on local governments and environmental affairs concerning the free commune experiment].
- Norwegian Parliament [Stortinget] (1991) Innst. S. nr. 18 (1991-92). *Innstilling fra kommunal- og miljøvernkomiteen om stortingsmelding om frikommuneforsøket*. [Recommendation from the parliamentary committee on local governments and environmental affairs concerning White Paper on the free commune experiment]
- OECD (2015) *OECD innovation strategy 2015: An agenda for policy action*. Paris: OECD Publishing
- Olsen, J. P. (1996) 'Norway: Slow learner – or another triumph of the tortoise?' in *Lessons from experience. Experimental learning in administrative reforms in eight democracies*, J. P. Olsen and B. G. Peters (eds), pp. 180–213. Oslo: Scandinavian University Press
- Olsen, O. E. and P. Lindøe (2004) 'Trailing research based evaluation; phases and roles'. *Evaluation and Program Planning*, 27(4): 371–380.
- Patton, M.Q. (2011) *Developmental evaluation. Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage
- Perrin, B. (2002) 'How to – and How Not to – Evaluate Innovation'. *Evaluation*, 8(1): 13–28.
- Petterson, O. (2016) 'Rational politics. Commissions of inquiry and the referral system in Sweden' in *The Oxford Handbook of Swedish Politics*, J. Pierre (ed.), pp. 650–662. Oxford: Oxford University Press.
- Rambøll Consulting (2014) *Midtvejsevaluering af frikommuneforsøg* [Mid-way evaluation of the free commune experiment]. Report to Local Government Denmark. Aarhus: Rambøll Consulting.
- Rambøll Consulting (2016) *Hovedrapport: Slutevaluering af frikommuneordningen* [Final report: Evaluation of the free commune experiment]. Report to Local Government Denmark. Aarhus: Rambøll Consulting.
- Rogers, E. (2003) *Diffusion of innovations*. 5th ed. New York: Free Press.
- Rose, L. (1989) *Frikommuneforsøk i Norden. En bibliografi* [Free-commune experiments in the Nordic countries. A bibliography]. Research note 1989: 133. Oslo: Norwegian Institute of Urban and Regional Research.
- Rose, L. (1990) 'Nordic free-commune experiments: increased local autonomy or continued central control' in *Challenges to Local Government*, D. King and J. Pierre (eds), pp. 212–41. London, Sage
- Rose, L. (1994) 'The management of organizational change in local government: methods and results' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 189–204. Aldershot: Dartmouth.
- Rose, R. (1991) 'What is lesson-drawing?'. *Journal of Public Policy*, 11(1): 3–30.
- Røvik, K. (2007) *Trender og translasjoner—Ideer som former det 21. århundrets organisasjon* [Trends and translations—ideas that shape the 21st century organization]. Oslo: Universitetsforlaget.
- Sanderson, I. (2002) 'Evaluation, policy learning and evidence-based policy making'. *Public Administration*, 80(1): 1–22.

- Schou, B. (1988) *Tre år med frikommuneforsøg—succes eller fiasko?—fornyelse eller stagnation?* [Three years of free commune experiments—success or fiasco?—renewal or stagnation]. Research report 1/88. Roskilde: Roskilde University Centre .
- Strömberg, L. (1986) *Plan for utvärdering av frikommunförsöket* [Plan for evaluating the free commune experiment]. Gothenburg: University of Gothenburg
- Ståhlberg, K. (1994) 'The Finnish case: a reform rather than an experiment' in *Towards the self-regulating municipality: free communes and administrative modernization in Scandinavia*, H. Baldersheim and K. Stahlberg (eds), pp. 85–108. Aldershot: Dartmouth.
- Swedish Government [Regeringen] (1991) Regeringens proposition 1991/92: 13. *Frikommunförsöket* [Parliamentary White Paper: The free commune experiment]
- Sørensen, E. and Torfing, J. (2011) 'Enhancing collaborative innovation in the public sector'. *Administration & Society*, 43(8): 842–868.
- Teodoro, M. (2009) 'Bureaucratic job mobility and the diffusion of innovations'. *American Journal of Political Science*, 53(1): 175–189.
- Vedung, E. (1997) *Public policy and program evaluation*. New Brunswick, NJ: Transaction Publishers.
- Walker, R. (2006) 'Innovation type and diffusion: an empirical analysis of local government'. *Public Administration*, 84(2): 311–335.
- Weiss, C. H. (1998) 'Have we learned anything new about the use of evaluation?'. *American Journal of Evaluation* 19(1): 21–33.
- Word, J. C., Stream, C. and Lukasiak, K. (2011) 'What Cannot Be Counted: Ethics, Innovation, and Evaluation in the Delivery of Public Services'. *The Innovation Journal: The Public Sector Innovation Journal*, 16(2).

ⁱ Hanne Foss Hansen uses a similar term, 'systemic evaluation governance', to denote something else – evaluation in the form of recurrent comparisons of a number of organisational entities or an entire organisational field (for example, hospitals) to facilitate competition and learning. She contrasts this with the evaluation of programmes and the evaluation of single organisational entities (Foss Hansen 2012: 51).