Policy innovation at the interface between central steering and local autonomy: Lessons learned and ways forward

Jostein Askim, University of Oslo (jostein.askim@stv.uio.no)
Lene Holm Pedersen, KORA/AKF (lhp@akf.dk)

Introduction

This paper is about how to facilitate policy innovation in a multi-level governance (MLG) system, i.e. a system where central and local governments share responsibility for the development, implementation and improvement of policies. By policy innovation we mean innovation in substantial policies relating to e.g. health care, employment or education, not innovation in MLG as such. We conceptualize policy innovation as a three-step process involving generating policy interventions, testing and verifying the effects of interventions, and finally diffusing effective interventions.

Policy innovation is challenging in all governance settings, but some challenges are especially present in MLG settings, where innovation takes place at the interface between central steering and local autonomy. Central governments, who have primary institutional responsibility for policies, have a good overview of existing interventions but often lack the proximity to praxis necessary to generate policy interventions that are more effective than existing ones. Central governments may also have restricted authority and legitimacy to install change at the local government level. Furthermore, local governments may be able to generate innovations and to learn from each other in informal ways, thereby diffusing interventions. Such informal second-hand learning is however error prone because it involves difficult exercises of de-coding the experiences of others and re-coding them into your own policies (Rose 1993). Ineffective interventions are thereby diffused alongside effective ones. Moreover, each municipality has restricted view of their opportunity space. They tend to be aware of the doings of municipalities in their geographical proximity, but it is not necessarily so that the most effective policy interventions are practiced there.

1 An earlier version of this paper was presented at the XVI Conference of the International Research Society for Public Management (IRSPM) in Rome, Italy 11-13 April 2012
Being aware of these challenges, central and local governments find ways of collaborating in order to succeed with policy innovation. Much such collaboration is emergent and ongoing and involves e.g. institutionalized performance reporting routines and sounding out routines. Other collaboration, the type we are most interested in here, takes the shape of innovation programs where central and local governments are partners, and where local governments constitute a laboratory for experimenting with new policies.

Our first research question is whether there is a trade-off in MLG between fostering innovative practices at the local level and producing generalizable knowledge. Drawing on innovation and evaluation literatures, we outline two generic experimental approaches to knowledge production. One is denoted the de-contextualized approach; the other the contextualized approach. These translate into two ways of designing and implementing policy innovation programs in MLG settings. Strengths and weaknesses of each design are discussed.

In order to show the practical relevance of the theoretical discussion, we draw on experiences from one specific approach to programmed MLG policy innovation: Free Commune Experiments (FCE). Five such programs have taken place in the Nordic countries since 1984. Our second research question is how the above-mentioned trade-off has unfolded in the Nordic Free Commune Experiments. Among the various ways of designing collaborative MLG programs, FCE is among the most ambitious. An FCE is policy innovation program in which selected municipalities for a period of time are granted the right to undertake local policy innovation projects that necessitate exemptions (waivers) from selected national laws and regulations.

Our third research question is whether and how MLG policy innovation program designers can overcome the trade-off between fostering innovative practices at the local level and producing generalizable knowledge. Is it possible to square the circle between local freedom to innovate and field level ambitions to document and spread good practice?

The paper proceeds as follows: The theoretical section discusses the term policy innovation, presents the two above-mentioned generic approaches to knowledge production and learning, and translates them into ways of designing and implementing policy innovation programs in a multi-level government setting. The empirical section presents the five FCEs that have unfolded in the Nordic countries since the mid-1980s. In the discussion and conclusion sections we position the FCEs against our theoretical treatment of policy
innovation and program design and discuss if and how past and future FCEs can combine the strengths of both de-contextualized and contextualized knowledge production and learning.

**Concepts and theory**

*Policy innovation*

There is a substantial body of research on policy innovations (Rogers, 2003; DiMaggio and Powell, 1991). This literature contains several noteworthy empirical studies of how diffusion takes place (Dahl & Hansen, 2004; Teodoro, 2006; Mizruchi & Fein, 1999). However, diffusion-oriented scholars are critiqued for failing to take into account the moment of entrepreneurship, i.e. the very generation of new policy ideas and solutions (Mintrom, 1997). In defense of the diffusion position it has been argued that policy innovations should not be restricted to new interventions, but rather include interventions that are new to the organizational entity adopting it (Berry and Berry, 1999). In this definition innovation is not about the generation of new interventions, but rather about diffusing interventions to new organizational settings. Innovation in this definition is also not about the adoption of interventions more effective than previous ones; it is simply about the diffusion of change. Hence, there is a risk of learning the wrong lesson.

Failing to pay attention to the newness and effectiveness of interventions makes the term innovation no more encompassing than the term change. Hence scholarly knowledge of innovation loses much of its potential for aiding in the improvement of public policy and welfare service delivery. In order to effectively aid practitioners, we believe scholarship on policy innovation should distinguish more between, one the one hand, how to encourage innovative practices at the local level, which may or may not be new and effective; and on the other hand, how to secure decontextualized, generalizable knowledge from innovative practices. Our discussion below of approaches to knowledge production and their design implications for programmed policy innovation is informed by this distinction.

*Experimental knowledge production*

We distinguish between de-contextualized and contextualized approaches to producing knowledge through experiments. Table 1 summarizes important differences between the two. We see this discussion of ideal types not as specific to policy innovation, but rather as generic to all production of knowledge through experiments. The de-contextualized approach
involves learning through randomized, controlled experiments. Participants in the study are randomly allocated to a treatment-group or a control-group. After randomization the treatment group is exposed to the intervention whereas the control group is not. Otherwise, treatment and control group participants receive the same treatment, so as not to disturb the effect of the intervention. Systematic evaluation is critical. The effect of the intervention is measured by comparing the difference between the two groups on the central output measure before and after the interventions. Developments can be compared for those who have been treated and those who have not, and placebo effects of being part of a treatment can be eliminated since all other potentially influential factors have been controlled for.

[Table 1 about here]

By contrast the contextualized approach is powered by praxis-oriented learning. This type of experiment is based on self-selection as the participants themselves choose to be part of the experiment. No participants are assigned to control groups. The motivation to participate in the experiment is often to undertake the experiment. Interventions are typically designed in the process of experimentation and the source of innovation is the closeness to praxis and the tacit knowledge in the particular organizational context. Design criteria such as internal and external validity are not central. Knowledge need not be generalizable to a larger population to be valid; knowledge is valid as the experience with the experiment rests in the organizations that have undertaken the experiment. As such the focus of evaluations of experiments is to narrate the process of undertaking the processes rather than on the documentation of effects, or, at most, very context-sensitive analyses of effects.

*Design of policy innovation programs in MLG settings*

As mentioned above, central and local governments sometimes collaborate on experimental policy innovation programs. Informed by the above discussion of types of experiments, we can in theory distinguish between two types of such programs: Contextualized programs, which will have their focus and strength in the generation of policy interventions (discovery), and de-contextualized programs, which will have their focus and strength in testing and verifying effects of interventions (see Fig. 1). We return in the discussion and conclusion sections to whether these are alternative or complementing types of policy innovation.

[Figure 1 about here]
De-contextualized programs aim to secure that rules and regulations are based on policy solutions (interventions) with documented effects. Policy innovation is seen as the transfer of effective policies and interventions from one site to the next. By contrast, contextualized programs aim to generate new policy solutions and, more broadly, foster a proactive culture of policy innovation and entrepreneurship at the local level. Policy innovation is seen as the discovery of new policies and interventions. These differences in focus and aim have important implications for program design, as outlined in Table 2. The different approaches to verification of results have implications for selection of participants. When verification rests upon a scientific study of intervention effects (the de-contextualized model), selection needs to be controlled by some coordinating agent that oversees the experiment. When verification rests upon context-sensitive description of processes and outcomes, participants can self-select to the experiment.

Table 2 contains the term “coordinating agent”. That role will often be taken by central government in a MLG setting. We nonetheless use the “level-neutral” term because, in principle, local governments can conduct local experiments of the de-contextualized type as part of more comprehensive experiments.

Diffusion of results
Transferring and diffusing interventions beyond the participating local governments is a shared ambition for both types of collaborative programs. Lessons from policy experiments can be transferred in many ways. We focus on two. One way is “vertical” in the sense that central government draws lessons from policy experiments and instructs local governments to change by means of rules and regulations. The other way is “horizontal” in the sense that interventions are voluntarily diffused among local governments. Such diffusion can involve quick imitation or rigorous analytical efforts with de-coding the experiences of others and re-coding them into your own policies (Rose 1993). Here there is no centralized instruction on whom (entities) and what (policies) to imitate.
Nordic Free Commune Experiments, 1984-2012

Method and data
This empirical part of the study covers all FCEs ever conducted. The first four are the FCEs in Sweden, Denmark, Norway and Finland in the 1980s and 1990s. The fifth is a second Danish FCE which started in 2012. Our analysis of the old FCEs is based on second-hand data from existing research reports. Our analysis of the present Danish FCE is based on official documents, qualitative analysis of FCE project applications submitted by local governments, and on participant observation at an FCE conference at the Ministry of Economy and Interior in January 2012.

FCE program theory
Free Commune Experiment (FCE) denotes a policy programme in which a limited number of municipalities for a limited period of time are granted exemptions from selected national rules and regulations. Based on existing descriptions of the Nordic experiences it is possible to distinguish some key general characteristics of FCE as a type of policy program. First, the purpose of the program is to try out policy solutions (interventions) whose implementation would be illegal absent such exemptions.

Second, FCEs tend to have two rounds of applications. At first municipalities are invited to apply to be included in the program. Their proposals state broad intentions but do not specify rules and regulations that inhibit their policy innovation. Central government, typically a coordinating secretariat located in a ministry, assesses the proposals and selects participating municipalities based on certain criteria. Then selected municipalities apply for exemptions from specific national rules and regulations and describe interventions they want to implement. Each municipality will tend to apply for a long list of exemptions. Central government then assesses these proposals and approves or dismisses them based on certain criteria. In this round the coordinating secretariat harmonizes assessments with line ministries that are institutionally responsible for the affected rules and regulations.

Third, any particular waiver tends to be made valid for all municipalities in the FCE programme and not only for the municipality that applied for it. Experiences gained are to be evaluated and provide the basis for further development of central-local relations, including the range of devolution (Rose 1990: 213).
In short, municipalities that take part in a FCE are by no means ‘free’ of all legal constraints. Rather, the FCE program theory is to provide participants with an opportunity to experiment with new policy solutions within boundaries that are temporarily wider than usual (Rose, 1990: 234). As such FCE holds a potential for stimulating entrepreneurial spirits and practices in the public sector.

**Sweden 1984-1992**

The first Swedish Free Commune Act was passed by the Swedish parliament (*Riksdagen*) in 1984. The FCE was part of a wider public sector reform program and was a response to societal criticism about excessive bureaucratization and regulatory zeal (Rose 1990: 227; Baldersheim and Ståhlberg 1994). The FCE was administered by the Ministry of Public Administration.

Seen in relation to the above design templates for experiential multilevel governance programs, the *purpose* of the FCE and its *problem-solution understanding* were half de-contextualized and half contextualized in their logic. The element we interpret as de-contextualized was that the FCE specifically empowered participants to introduce more flexible structures for political committees (intervention i.e. pre-defined and exemptions pre-arranged). The contextualized element was an open invitation to participants to apply for exemptions from almost any rule or regulation and insert any intervention in their place.

The government initially planned to *select participants* to the FCE according to the de-contextualized design template. The intention was to sample municipalities to be representative of the whole population of municipalities in terms of size, geographical distribution and political control. However, as the selection process unfolded, criteria pertaining to the content of the proposals were prioritized over applicant attributes. Special priority was given to proposals whose projected results could be adopted widely. The resulting sample over-represented municipalities with a track record of willingness and ability to conduct change and renewal. No control group was identified. Our interpretation is that this selection process was in line with the contextualized design logic.

Twelve municipalities were admitted to the program in 1984. These were given several months to formulate and submit their applications for exemptions from national rules and regulations. 280 applications were eventually submitted. Line ministries initially stalled in their processing of applications but sped up after pressure from the prime minister. Exemptions that were granted were made valid for all participants and not just for initial
applicants. Twelve more participants were admitted in 1988, when the FCE was prolonged for the first time and a further 15 in 1990, when the FCE was prolonged again.

As for verification of results, the initial plan was to let the coordinating secretariat in the Ministry of Public Administration be responsible for evaluating the experiment. However, line ministries distrusted the secretariat’s deregulation philosophy and initially launched their own FCE evaluations. In 1986 the Secretariat and line ministries agreed to entrust the evaluation to a presumably neutral external party - the University of Gothenburg. Participating municipalities were not obliged to evaluate their own experiences under the program. In 1990, as part of the second FCE prolonging, the Government appointed its own investigator of the FCE. All in all, external observers judged the evaluation of the Swedish FCE as patchy. The design of the programme also reduced its “evaluability”. It is difficult to isolate the impact of governance interventions in a programme that spans eight years. During this period a range of contemporaneous and related events and changes would have taken place. As a result, distinguishing “signal from noise” would be neigh impossible. Our interpretation is that the verification of results followed none of the two design logics. Due to the critical importance of evaluation in the de-contextualized design model, the lack of a coherent evaluation design and practice is most detrimental to that design model.

The diffusion of results in the Swedish FCE followed the horizontal design logic. Legal changes were made, e.g. to rules for political committee structures, but decisions did not rely on experiential learning based on evaluations of the FCE. A range of other changes to central-local relations were also indirectly inspired or informed by the FCE experiences. The government was reassured by the FCE experience that it was a good idea to increase local discretion on how to organize their political committees. Rules restricting committee flexibility were therefore permanently lifted in 1992. Second, devolution of responsibilities in the education sector in the early 1990s was informed by positive FCE experiences with granting local governments more freedom in the running of schools. However, observers have not found direct linkages between the FCE and the changes that followed. Hence we interpret the diffusion of results from the Swedish FCE as more horizontal than vertical.

Denmark 1985-1993
The Danish Free Commune Act was passed by the Danish parliament (Folketinget) in 1985. The Ministry of the Interior administered the FCE programme. The FCE was prolonged once, in 1987. The stated purposes of the Danish FCE were very general: increased self-
government, improved services and increased efficiency. In addition, de-bureaucratization was a key background motivating factor, as this theme had been on the central-local relations agenda in the preceding years. We can infer from the *definition of problems and solutions* that the purpose of the FCE was in clear accordance with the contextualized design template for policy innovation. Unlike the Swedish case, the government did not pre-specify interventions or pre-arrange exemptions. Participants were invited to apply for exemptions from almost any rule or regulation and insert any intervention in their place. This was a conscious choice by the Minister of the Interior, intended to foster local entrepreneurship and ideation.

*Selection of participants* to the Danish FCE also to some extent accorded with the contextualized design template. There was no control group and the number of participants was fairly large. 27 participants were selected in 1987 and a further 36 in 1989, when the FCE was prolonged. There was, however, a sampling strategy. The selection of participants aimed at a geographical, structural, size-related and political spread. All in all, the selection process adhered mostly to the contextualized design template. Selected waivers were made valid for all municipalities and not only for FCE participants.

As in Sweden, participants were given generous time to formulate their applications for specific exemptions and interventions to put in their place. By 1992 the participants had submitted 866 applications. Only 390 were approved. Danish line ministries, like their Swedish counterparts before them, were slow and restrictive in their processing of applications. Ninety-two applications were flat-out rejected; 210 were withdrawn by the municipalities because negotiations with line ministries were too slow and time consuming; a further 109 were neither approved nor rejected but simply got stuck in processing. Many municipalities reportedly found line ministries to lack understanding of the innovative logic and resistant towards change. Strongest criticism was aimed at line ministries with little previous direct interaction with municipalities.

As for *verification of results*, there was no comprehensive evaluation of the Danish FCE. The opposition called for a comprehensive and coordinated evaluation when the Free Commune act was passed, but the Minister of the Interior refused due to a general skepticism that the knowledge produced would contribute with something useful. In addition evaluation could be costly and produce politically unwarranted results (Albæk, 1994). Instead the Ministry of the Interior commissioned an evaluation of four specific FCE projects deemed immediately policy relevant. In addition the participating municipalities were required to
evaluate the experiences themselves; i.e. a fairly radical contextualized approach to the verification.

The relative lack of comprehensive evaluation did not prevent diffusion of results. FCE experiences, if anecdotal in nature, were used as arguments for changes in laws in parliamentary discussions. Also, line ministries followed up the FCE with general rights for municipalities to experiment within their areas of authority.

**Norway 1986-1992**

The Norwegian parliament (Stortinget) passed a Free Commune Act in 1986. The Ministry of Local Government and Regional Affairs administered the FCE programme. The FCE was prolonged once, in 1987. The FCE had two stated purposes. One was to advance more flexible and joined-up administration at the local level. The second was to advance better coordination in the way national government entities handled local government issues - i.e. improved multilevel governance. Both purposes pertain to governance, neither to e.g. the quality or efficiency of service production. As such the purposes correspond to neither of the two design templates for experiential multilevel governance programs. As in the Danish case, the absence of pre-determined definitions of problems and solutions invites the interpretation that the Norwegian FCE’s purpose was in accordance with the contextualized design template. Participants were invited to apply for exemptions from almost any rule or regulation (Baldersheim and Fimreite 1994).

**Selection of participants** to the Norwegian FCE was slightly more de-contextualized than in Sweden and Denmark. Following the issuing of an open-ended invitation to apply for inclusion in the programme, central government gave municipalities two months to submit their applications. The number of participants was pre-set to 24 but 26 were selected. The selection process was faithful to preset selection criteria and resulted in a sample that was fairly representative in terms of size and geographical spread. There was however no control group. Participants submitted 260 applications for exemptions from national rules and regulations. The applications sprawled 80 acts. Seventy projects were eventually approved.

Central government’s assessment process was similar to the Swedish and Danish experiences, with the coordinating ministry brokering agreements between municipalities and reluctant line ministries. As in Denmark, line ministries with limited previous municipal experience were most reluctant to grant exemptions from their rules and regulations.
What sets the Norwegian FCE aside from the other cases is its more de-contextualized approach to verification of results. Evaluation was an integral part of the FCE. A comprehensive evaluation programme was launched from the start and entrusted to external research institutions. 30 reports and four books were produced. The evaluation covered the whole program and four projects were evaluated in more detail. A 1991 governmental white paper on the FCE drew heavily on findings from the evaluation.

Results of the FCE diffused both vertically and horizontally. Two major pieces of legislation were introduced in 1993 as a direct result of the FCE – a new local government act and a new general act on experimentation in public governance. Furthermore, experiences from the FCE projects diffused broadly among local governments. The FCE also had unanticipated long-term effects. A major administrative reform in 2006, a merger of the national apparatuses for employment, pensions and social services, had its precedent in a FCE project from the late 1980s.

**Finland 1988-1996**

The Finnish parliament passed a Free Commune Act in 1988. The Ministry of the Interior administered the FCE programme. The FCE was prolonged once, in 1992. The stated purposes of FCE were very general: increased self-government and popular participation, improved services and increased efficiency. Finland had gone through a severe economic crisis in the 1980s so empowering municipalities to make cut-backs in public expenditures was an important motivating factor. The lack of preset direction for the FCE and little central government guidance for applicants invites the interpretation that the purpose of the programme was contextualized. Participants were invited to apply for exemptions from almost any rule or regulation and insert any intervention in their place.

Fifty-six municipalities were selected to participate in the FCE. What sets the Finnish FCE aside from the other Nordic experiences is that its program boundaries were perforated. Similar to the Danish case, but to a larger extent, selected waivers were made valid not only for FCE participants but for all municipalities, thus giving the programme features more akin to a national administrative reform than a limited experiment.

Verification of results was scarce. There was no comprehensive evaluation, only reports that analyzed experiences for selected target groups and projects. FCE experiences nonetheless diffused both vertically and horizontally. Among the most tangible results were a new Local Government Act, permanent deregulations in the education sector and a follow-up
experimental policy program on citizen participation. Many line ministries, in spite of initial resistance to experimentation within their areas of responsibility, later changed legislation with inspiration from FCE projects.

*Denmark II, 2012*

The second, now ongoing, Danish FCE was born in a context of deregulation and “cutting red tape”. Local governments and their national association (KL) complained that central regulation had gone too far. Thus one purpose of the FCE is to get rid of rules and regulations which cause unnecessarily administrative burdens. Furthermore, it is a purpose of the FCE to support and enhance a culture of entrepreneurship in the local governments. Interventions are created as part of the local experiments and local governments apply for waivers when intervention descriptions have been finalized. The Danish parliament passed an act in mid-2012 that grants necessary and prioritized waivers.

21 municipalities applied to obtain status as free communes. Among these, nine were selected as participants in the FCE. By January 2012 215 applications for specific interventions had been received by the Ministry of Economy and Interior. One observation about these applications – interpreted as definitions of problems and solutions – is that many applications, especially in the employment sector, seek to reduce “red tape” (excessive bureaucratization). A second observation is that (as in previous FCEs) some but not all suggested interventions necessitate waivers. It thus seems to be a lasting phenomenon that local governments perceive of national rules and regulations as more inhibiting than they in fact are. A third observation is that the applications are more specific on problems (rules in their way) than on solutions (interventions to implement).

The low solution specificity was given three explanations at a ministerial FCE workshop. One was that local governments were given only two weeks to complete project proposals after having been admitted to the FCE program. As a result, said one workshop participant, “[we used] what we had on the shelf when the application was due”. Second, local governments, disillusioned by previous engagements with line ministries, were reluctant to invest time in detailed project proposals; they would rather wait to see line ministries’ first moves and adjust to that. Third, many participants are less geared towards changing operational practices than towards thinking abstractly about desired changes in values and ways of thinking – at least at this early stage in the program.
Getting exemptions from national rules and regulations seem not to be the only motivation for local governments to be included in the FCE program. It also matters to be perceived as a modernizing front runner among external and internal audiences. FCE status provides local political and administrative executives leverage in pushing reform agendas inside their own organizations. As expressed by one municipal FCE workshop participant: “We need to get rid of some corrosion. It is not just central regulations that cause us trouble”.

The FCE application form asked applicant local governments to describe how they envisioned evaluating interventions they wanted to implement. This does not mean, however, that the program’s thinking about verification of results follows a decontextualized design template. Ambitions for comprehensive nationwide evaluation seem limited; so do local governments’ evaluation ambitions. A reading of FCE applications in the employment policy area reveals that their evaluation designs are underdeveloped and lack e.g. considerations about starting-point measurements and control groups. It seems that the FCE’s evaluation mode will be documentation of experimental practices.

At this early stage in the program it is of course not possible to say much about whether and how interventions will be diffused. Based on previous FCEs we can however extrapolate that unless central government initiates a coordinated study of effects, diffusion will primarily follow the horizontal model of diffusion.

Discussion

The potential for mutual integration

The two templates on experimental designs both have strengths and weaknesses - and indeed a combination of the two may have certain advantages. Discovery and testing of interventions: The contextualized approach has its comparative strength in the discovery stage of policy innovation; the de-contextualized approach in the testing and verification of the effects of interventions. How can both strengths be combined? What can we learn from the Nordic FCEs and innovation literature? One integration strategy is sequential. In step one an open-ended contextualized innovation program generates policy interventions; in step two a de-contextualized program – with carefully selected participants and narrowly defined scopes for problems and solutions – test and verify these interventions. The Finnish FCE case illustrates such sequential integration of program design approaches. There, topically open-ended FCE was followed by a topically narrowed experimental policy program on citizen participation.
Another way is simultaneous integration of program design approaches. One policy innovation program can have two tiers, one tier which is contextualized, i.e. indiscriminately open-ended for problems and solutions, and another tier which is more de-contextualized, i.e. careful in the selection of participants, narrow in scope and with heavy emphasis on systematic evaluation. The Swedish FCE case to some extent illustrates such simultaneous integration, in that it combined open-ended experimenting and experimenting with one specific intervention, namely changes in political committee structures. The Norwegian FCE also showed evidence of such simultaneous integration, in that systematic evaluation was emphasized, both overall and in relation to certain prioritized interventions and types of local governments.

**Diffusion of results:** We defined the third and last stage in policy innovation as the diffusion of effective interventions. “Vertical” policy transfer is in principle viable for results of both contextualized and de-contextualized policy programs. The de-contextualized nature of the results may however be a weakness of such lesson-drawing. Solutions that work well “on average” may still be ill-fitted or hard to implement in many local governments. “Horizontal” policy transfer is primarily linked to lessons from contextualized policy programs. According to sociological institutionalism, imitation is primarily fuelled by repairing perceived shortages in legitimacy (Meyer and Rowan 1977). Local governments will be willing to learn from context-sensitive, vivid depictions of stories of local governments that are seen as successful and modern. A weakness of this learning mode is however the risk of learning the wrong lessons. Knowledge on the effects of contextualized policy programs often suffers from low internal and external validity. Hence, contextual programs may lead to diffusion of policy ideas which are in fact ineffective. This line of inquiry has sadly not been pursued by research on FCEs.

However, there is no reason to believe that knowledge produced in line with a decontextualized model cannot be diffused by imitation. Evidence of effects may support the codification of interventions as modern and rational, and documented effects may work by ‘force of the better argument’ (Elster, 1983). Thus, interventions which have documented effects may have a higher viability as they may provide arguments which can cut across pressure from professionals and political opponents.

But how we design programs whose results diffuse both vertically and horizontally? We believe the answer lies in selection of participants and evaluation designs. Program designers
can, as evidenced e.g. by the first Danish FCE, include “field leader” attributes (demonstrated and acknowledged ability to “modernize” etc.) as one criteria among others when selecting participants to policy innovation programs. That way chances increase that you have participants whose experiences will be visible and legitimate, hence fostering horizontal diffusion. Furthermore, program designers can – to greater extent than evidenced by the Nordic FCEs – do more to pre-define and scope down problems and solutions program participants take on board in their local experiments. We presume that local government stakeholders see the value in, and accept, more steering of experiments than was the case in the FCEs. There is certainly a potential in setting local government free to formulate new interventions – something line ministries struggle to do, due e.g. to information asymmetries. However, there may be reason to strive for a higher level of coordination when the effects of the interventions are to be evaluated. Program designers should also consider allowing fewer interventions per participant than the Nordic FCEs did. That way it will be more viable to systematically evaluate and verify the effects of the interventions, because the amount of “noise” has been reduced. Finally, program designers can foster horizontal alongside vertical diffusion by combining contextualized and de-contextualized evaluation models. Once you have designed a relatively “evaluable” program, there is good reason to commission evaluations that not only “coldly” test for causal effects of interventions, but also include thick descriptions of “best cases” and loopholes to avoid.

Summing up, we argue even if the contextualized model is the default for experimental designs in a local government setting, a synthesis would hold a potential if the discovery, testing and diffusion faces could be integrated in the FCEs. In this case, entrepreneurship could be facilitated in the discovery phase, while a systematic test of the effects, would secure the production of knowledge and reduce the risk of learning the wrong lesson in the diffusion phase (see figure 1).

Obstacles to integration in a MLG setting
We can draw on the Nordic FCEs to find out what the obstacles to such a potential integration could be. Indeed, there may be elements in MLG which hinders the development of an integrated approach to learning and innovation. First, divisions of responsibilities may be one obstacle. On one hand central governments have responsibility for policies at the local level, but responsibility for evaluations have nonetheless tended to be decentralized. Central governments could of course set out rules for how evaluations of FCEs should be made,
perhaps as a precondition for getting status as a FCE. However, restrictive rules on this would be against the spirit of FCEs. Furthermore, the division of responsibilities within central government also may create obstacles. Some ministries may have an interest in letting the thousand blossoms blossom, whereas others are less interested in experiments within their jurisdiction, especially if experiments challenge regulations that are well-established and/or have documented positive effects. Research on FCEs shows systematic variation between line ministries and coordinating ministries (Schou 1988). Division of responsibilities at the local level may also create obstacles. Even if there is a commitment to carrying out de-contextualized-type FCEs at strategic level of the municipality, there may be a lack of interest in “controlled” experiments and dedicating resources to evaluations at the operational level.

Second, structural capacities for implementation may be an obstacle to integration. De-contextual-type evaluations may provide knowledge at the local level. However, it often requires capacities which cut across the single local government to carry out decontextualized experiments. Randomized selection of participants and the creation of an adequate control group require coordination. Even if randomized experiments at the individual level can be carried out with in a single municipality, it is for practical matters often necessary to include more municipalities in order to reach an adequate amount of participants.

Third, there may be obstacles to more de-contextualized policy innovation among political actors and interest organizations. Associations of local governments may prefer a more contextualized model where local governments are free to imitate interventions they judge to be most relevant to their municipality – regardless of whether or not they have validly documented positive effects.

Fourth, there may be obstacles at the ideational level; that is in the understanding among specialists and expert communities of what makes for valid, true knowledge. If we compare the development over time, from the old FCEs to the FCE that started in Denmark in 2012, there are examples of stability as well as change. In the 1980s there was little coordinated evaluation of the FCE in Denmark. The Minister of Interior at the time did not believe an evaluation would produce politically relevant knowledge (Albæk, 1994). Since then the so-called evidence movement has gained momentum. Its most ardent advocates have been nicknamed ‘randomistas’ due to their belief in the superiority of decontextualized experiments, also in policy analysis (Brogaard, 2011, Bhatti et al., 2006). They have complained that evaluations of public sector reforms are unscientific and argued that we incur costs when we
implement unwarranted policies (Kongsted and Konnerup 2011, Hede 2011). So-called randomistas are found among political actors involved in the present Danish FCE, notably among specialists working in the Ministry of Employment. They are well aware of documented positive effects of regulations from which the municipalities now apply for exceptions, such as the documentation of timeliness in job-interviews with unemployed. However, these actors are not central in the decision-making on the evaluation design for the FCE. So even though there has been some ideational change on the ideal model of policy innovation, stability in divisions of responsibilities and political interests in FCE cause great stability in the design of FCE experiments and evaluations.

**Conclusion**

We have described, discussed and illustrated two ways of thinking about policy innovation in multi-level governance settings: the contextualized and the de-contextualized experimental approach. Each has its pros and cons. The contextualized experiment has its relative strength in the discovery phase and its relative weakness in the verification of effects of policy interventions. This verification fault does not hinder horizontal diffusion of the intervention among local governments, though; new policies spread even without documented effects. There is therefore a danger of learning the wrong lessons. The contextualized approach to diffusion of policy innovations may be effective but it is “risky”. The decontextualized experiment has its relative strengths in verification and generalization and “secure” policy diffusion and its relative weakness in the discovery phase; it does not afford much attention to how new interventions are generated. Thus there is a potential trade-off between fostering innovative practices at the local level and producing generalizable knowledge.

This trade-off is evident in past and present Nordic Free Commune Experiments (FCEs). On one hand FCEs give room for experiments that unfold close to praxis. This carries strong potential for policy entrepreneurship and for generating new and broadly legitimate policy ideas and interventions. However, the analysis of Nordic FCEs shows variation in how experiments have been evaluated and results verified. The FCEs in Norway, in particular, but also in Sweden, employed coordinated evaluation programs. Evaluations in Denmark and Finland were more patchy and fragmented. In Denmark today the default model is that the municipalities in each application state how they are going to evaluate their own experiments.
Knowledge production in FCEs takes place in the interface between central steering and local autonomy. Our overall assessment is that that the FCEs, despite their science-leaning rhetoric (re. free commune experiments), have been better as generators of new policy ideas than as test beds for the effects of interventions. Too often evaluation designs have fragmented into process evaluations in individual municipalities. Ideas and interventions generated in the experiments have nonetheless spread widely among Nordic local governments, not only through voluntary horizontal diffusion but also via the multi-level government channel; central government agencies have taken inspiration and lessons from local experiments too. This is a risky innovation strategy. Limited knowledge has been produced about whether and how positive effects of new policies can be replicated across local governments.

However, there is an untapped potential for integrating the discovery, testing and diffusion phases of policy innovation. Strength and weaknesses of the contextualized and de-contextualized approaches to experimentation can be combined through strategies discussed here. Though its successes and failures Nordic FCEs have taught us important lessons about how to square the circle between local freedom to innovate and field level ambitions to document and spread good practice.
References


### Tables

**Table 1: Approaches to producing knowledge through experiments**

<table>
<thead>
<tr>
<th></th>
<th>De-contextualized knowledge production</th>
<th>Contextualized knowledge production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledge objective</strong></td>
<td>Better knowledge of the effects of interventions</td>
<td>Better knowledge of the development of interventions and their applications</td>
</tr>
<tr>
<td><strong>Process of discovery</strong></td>
<td>Randomized, controlled experiment, <em>ex post</em> to the development of intervention</td>
<td>Heuristic. Experience based problem solving, learning and discovery</td>
</tr>
<tr>
<td><strong>Relation to generation of interventions</strong></td>
<td>Intervention is designed <em>ex ante</em> to the experiment. The source of the innovation is not important.</td>
<td>Intervention is designed in the experiment. The source of the innovation is the participants’ praxis and experience.</td>
</tr>
<tr>
<td><strong>Criteria for valid knowledge</strong></td>
<td>Solid knowledge of effects and causality</td>
<td>True and inspiring depiction of process, outcome and context</td>
</tr>
<tr>
<td><strong>Functional requirements</strong></td>
<td>Control over the experiment. Participants selected in a randomized way; distinction between effect and control group. Measurement validity, internal validity (causality versus correlation), external validity and replicability</td>
<td>Nurture the innovative and entrepreneurial spirits of participants Participants self-select/volunteer based on interest in and ability to conduct the experiment.</td>
</tr>
</tbody>
</table>

**Table 2: Design templates for policy innovation programs in MLG settings**

<table>
<thead>
<tr>
<th></th>
<th>De-contextualized program design</th>
<th>Contextualized program design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition of exemptions (problem)</strong></td>
<td>Coordinating agent selects regulations from which exemptions can be granted. Coordinating agent agrees exemptions with line ministries prior to local level’s involvement</td>
<td>Local level selects regulations from which exemptions are necessary to develop policy solutions. Exemptions agreed case-by-case in dialogue between applicants and line ministries, brokered by coordinating agent.</td>
</tr>
<tr>
<td><strong>Definition of intervention (solution)</strong></td>
<td>Interventions (policy solutions) are defined <em>ex ante</em> by coordinating agent and applied at the local level</td>
<td>Interventions are created as part of local experiments. Some but not all necessitate exemptions from regulations</td>
</tr>
<tr>
<td><strong>Selection of participants</strong></td>
<td>Selection based on criteria that facilitate generalization of results (median or critical case, etc.). Limited number of participants is necessary to facilitate tight monitoring of experiments.</td>
<td>Self-selection facilitates inclusion of enthusiastic and able participants. Such field leaders enable <em>ex post</em> diffusion of innovations. Large number of participants unproblematic.</td>
</tr>
<tr>
<td><strong>Verification of results</strong></td>
<td>Evaluation as a scientific study of the causal effects of interventions. De-contextualization to distinguish “signal from noise”.</td>
<td>Evaluation as context-sensitive documentation and description of the experimental process and outcome.</td>
</tr>
</tbody>
</table>
Figures

![Diagram of innovation program focus]

Figure 1: Comparative focus of innovation programs