

*Paper presented for: A performing public sector, the second transatlantic dialogue,
June 1–3, 2006, Leuven*

Explaining Organizational Learning from
Benchmarking in Networks:
Experiences from Norwegian Local Government

Jostein Askim*

Department of Political Science
University of Oslo

Knut-Andreas Christophersen
Department of Political Science
University of Oslo

Åge Johnsen

Faculty of Business, Public Administration and Social Work
Oslo University College

*Corresponding author. Email: jostein.askim@stv.uio.no

Introduction

The purpose of this article is to analyse the dynamics of organisational learning from benchmarking. More specifically, we explore variety in learning outcomes and ask why some organizations learn a lot and others less. The background for our article was the opportunity to study the largest systematic benchmarking project ever conducted in Europe. The project was a nationwide Norwegian benchmarking project for local governments, called “Networks for renewal and efficiency”. More than 300 municipalities took part in this program from 2002 to 2004, grouped in 40 benchmarking networks.

Analyzing benchmarking within relatively stable inter-municipal networks is potentially valuable to the discourses on network management (Agranoff and McGuire 2001) and organizational learning (Dekker and Hansén 2004; Hartley and Allison 2002) because it enables simultaneous estimates of the effects of network characteristics and factors internal to the organization as well as comparisons of the two. Within each group of factors we study the influence on learning of variables that from the perspective of a given benchmarking effort will be perceived as endogenous and exogenous. Network configuration (network size and heterogeneity) and management involvement are among the former; the fiscal situation and political regime of a municipality are among the latter. These issues carry lessons not only for the academic dialogue on organizational learning from benchmarking, but also for practitioners in public administration.

We call benchmarking a “practice” because we want to frame it as a limited-purpose practical theory of action. A theory of action is prescriptive, but warranted by descriptive understandings (Barzelay 2004). The theory that carries benchmarking around the world is that this practice can perform the organizational function of innovation, via vicarious organizational learning (Ammons, Coe and Lombardo 2001; Behn 2003; Hartley and Allison 2002). But several scholars have pointed to weaknesses in the descriptive understandings that underpin the perceived means-end relationship between benchmarking and organizational learning (Fedor et al. 1996; Folz 2004; Hartley and Allison 2002; Wollman and Page 2002). Some attempts have been made at mapping a causal chain between benchmarking, learning and improvement (Askim 2004; Fedor et

al. 1996), but few empirical studies have been done to identify factors which actually condition organizational learning from benchmarking activities. Hence, while generic prescriptions continue to abound, there is still little empirical knowledge within the fields of public administration and public management of where benchmarking works, and why. This is where our article aspires to contribute.

The research question we ask in this article is the following: To what extent is organizational learning from benchmarking conditioned by history-dependence, composition of benchmarking networks, internal organizational processes, and political factors? For example: Are municipalities that have prior experience with benchmarking or prior mutual networking experience able to achieve more organizational learning than those without such experience? Are larger networks more conducive for organizational learning than smaller ones? Do structural homogenous benchmarking networks work better than those that are more heterogeneous? What is the influence on learning from characteristics internal to the organization such as management participation? And finally, how do political factors such as political regime, regime stability and political competition play in?

The remainder of this article is outlined as follows. The first section reviews theory on performance management, benchmarking and organizational learning. We then present the case project in some depth. In the third section we outline an analytic framework with a set of hypotheses about what might account for variety in learning. The fourth section documents the research design, data and methods we employ. We then go on to document and discuss the results from the OLS regressions. The final section provides our conclusions, assesses limitations in our study, identifies implications for theory and practice and gives suggestions for future research.

Performance management, benchmarking and organizational learning in government

Public sector *performance management* can be defined as the design, implementation, measurement, reporting and evaluation of performance information, with the intention to improve decision-making in administrative and political processes (Johnsen 2005). Measurement of activities and outputs is as old as public administration itself (Williams

2003; Pollitt and Bouckaert 2005), but since the 1990s performance measurement has become both more extensive (more fields), more intensive (more functions) and more external in use (Bouckaert 1996). Measurement generates *performance information*, which is systematic, numerical information describing the outputs and outcomes of public programs and organizations, generated by systems and processes intended to produce such information. This includes data generated by performance monitoring systems operated by operational managers, as well as data flows from evaluations (internal or external), performance audits (internal or external) and other special exercises and reviews which have as a main purpose the discovery or analysis of performance information (Pollitt 2005). Performance information is often presented in the form of performance indicators (PIs). These may cover organizational efficiency, effectiveness or equity.

Echoing the evolution of the dialogue on NPM more generally, research on performance management and measurement has recently moved beyond broad assessments of prevalence and benefits to closer investigations of purposes and effects in different settings (Nicholson-Crotty et al. 2006). One important setting for the use of performance measures is benchmarking activities. A shorthand definition of *benchmarking* is that it is a practice that involves comparing performance and/or process information between organizations or organizational units in order to improve performance (Fedor et al. 1996:161). The term benchmarking actually covers several strategies, depending on which comparison logic that is applied. Contextualization of the current performance of the organization is at the core of benchmarking, but this can be facilitated through comparisons with benchmarks ranging from the organization's own previous performance to some normative standard or the performance of other organizations (Fedor et al. 1996; Levitt and March 1988; Mausolff 1999; Moynihan 2005; Simon 1939; Wholey 1983). Previous performance can be all-time highs or over-time averages, while normative standards are aspiration levels set by the organization itself or by someone else. The performance of other organizations can be group averages or positive outliers, so-called best cases. In local government benchmarking efforts, one can observe these comparison logics operated one by one or in combination.

The benchmarking effort studied empirically in this paper is based primarily on the third logic, where performances are compared between organizations. This benchmarking type can be further distinguished by the extent to which one moves beyond static performance comparisons to more dynamic process comparisons. The latter entails a series of steps to identify top performers in that process, analyze process components and pinpoint differences that contribute to superior performance, and to adapt key elements for use in one's own organization (Ammons, Coe and Lombardo 2001; Fedor et al. 1996). Process comparisons are most prevalent in the private sector. Public sector organizations usually pursue detailed analysis of single processes as a follow-up step, if at all. It is rarely considered an essential element of the benchmarking effort (Ammons, Coe and Lombardo 2001).

Since the turn of the century benchmarking has become practically part and parcel of contemporary public administration, especially in Northern Europe, North America and Oceania. Benchmarking is used both "vertically", i.e. as a performance monitor for principals to oversee the activities of agents (e.g. central–local government or government–private service providers) and "horizontally", when organizations voluntarily engage independently or with others in systematic search activities. Most practitioners and observers will highlight *organizational learning* as the overarching aim of such horizontal benchmarking activities (Ammons, Coe and Lombardo 2001; Behn 2003; Hartley and Allison 2002). But though organizational learning is generally recognized as a useful lens through which to study the dynamic relation between information, knowledge and organizational action and change (Dekker and Hansén 2004), the scholarly community is divided on how to conceptualize and study this phenomenon.

Many define organizational learning so as to incorporate only changes that have taken place as a result of a process. And though it may be true that most municipalities will engage in benchmarking, and indeed other performance management reforms, with a stated aim to undertake changes in routines and practices to improve performance (defined somehow), we argue that a meaningful empirical analysis of a benchmarking project benefits from a definition of organizational learning that stresses cognition as well as behavior. Organizational improvement is arguably an outcome from benchmarking

that is worth capturing, but so are attributes associated with the organizational decision making process per se. We see the outcome of a deliberation that is related to (spurred by, informed by, etc.) benchmarking experiences as a political decision. The outcome might be a decision to improve, but it might just as well be a decision to maintain status quo or even to reduce relative or absolute attributes (quality, quantity, efficiency) of a given task. Moreover, in a complex organization like a municipality it can also be the case that a decision to change has been made but has yet to be implemented due to resistance. A non-decision or no observable changes do not necessarily imply absence of learning.

Emphasizing cognition as well as behavior might help us avoid underestimating organizational learning due to e.g. resistance to implementation, but we do not necessarily overcome the problem of symbolic use of information (Feldman and March 1981). On the other hand, one could argue that actions taken due to symbolic use of information is not counterfactual to organizational learning.

This view leads us to define organizational learning as processing of information which changes an entity's range of potential behavior (Hartley and Allison 2004; Huber 1991). Fiol and Lyles epitomize this perspective when saying that learning is about "the development of insights, knowledge, and associations between past actions, the effectiveness of those actions, and future actions" (1985: 811).

Case presentation: Networks for renewal and efficiency

The benchmarking effort studied in this article is a nationwide Norwegian benchmarking programme for local governments called "Networks for renewal and efficiency" (*Kommunenettverk for fornyelse og effektivisering*). The project was initiated by the Labour government in 2000, and was carried out in co-operation with The Norwegian Association of Local and Regional Authorities (*Kommunenes Sentralforbund, KS*) and the Ministry of Local Authorities and Regional Development.

The Ministry for Labour and Public Administration, which was responsible for the modernisation program for the public sector, wanted to stimulate the use of performance indicators and to document and stimulate development work regarding efficiency, effectiveness and quality throughout local government. Although a centre-right minority coalition took over government after the national election in 2001, the benchmarking

project was implemented from 2002. Following a piloting phase with 9 municipalities in 2001, all municipalities were invited to participate, with the government financing the direct operating costs of the project. During the period 2002–2004 the project encompassed 313, i.e. 72 percent of all 435 Norwegian municipalities. From 2005 Networks for renewal and efficiency was established among the Association of Local Authorities' permanent member services. Since then municipalities have had to pay a modest fee to participate.¹

The operational project design of Networks for renewal and efficiency had three components: Performance measurement, comparisons and networks. The measurement component coincided with another Government-initiated local governmental benchmarking effort, called KOSTRA. This is a data registration and information scheme designed to help central government keep track of expenditures and activities in the local government sector. From 2002 reporting to KOSTRA was made compulsory for all municipalities in Norway. This means that there was suddenly a vast range of comparable (and increasingly reliable) financial and non-financial performance indicators available. The majority of municipalities had little previous experience in using such data for planning or reporting purposes, so it is fair to say that the stage was set for a project like Networks for renewal and efficiency. Comparisons and dialogue over KOSTRA data was therefore the first stage in the project design.

These comparison exercises took place in networks comprising 4–8 municipalities. The 143 municipalities that entered the project in 2002 were divided into 25 networks. The configuration aimed to balance a perceived need for a certain level of network homogeneity in terms of municipal size with a need for low coordination costs, i.e. small traveling distances. As the number of participating municipalities grew, so did the number of networks. More than 40 networks were active at some point during the project period. Each network chose two or more service areas to focus on. Most preferred primary education and/or elderly care, but there were also networks that chose to work with areas such as kindergartens, social welfare and children welfare.

Each network was assigned a so-called network guide from among the project staff. Guides were responsible for arranging and facilitating network meetings, e.g. by

¹ Annual fee of NOK 8,000–35,000 (USD 1,200–5,200) per municipality.

preparing presentations with comparative KOSTRA data and by chairing and spurring discussions. The municipalities were exposed to various kinds of comparisons, both between the network partners and between their own performance and national and relevant grouped averages and top scores. Each municipality sent about a handful of participants to each network meeting. These participants varied from CEOs to street-level bureaucrats depending on the more specific theme of the network meeting in question. In the meetings the participants identified and discussed similarities and differences in performance.

After meetings with various KOSTRA data comparisons, the networks shifted focus towards service quality. This required new performance measurement exercises, as KOSTRA does not contain much information about this. Hence the Networks for renewal and efficiency project developed standardized user and employee surveys for the different service areas, which were conducted in (and by) the municipalities. Other, less perception-based quality measurement techniques were also developed. Results were reported to the Association of Local Authorities, who analyzed the relationship between resource use and service quality. These analyses were then reported back to the networks to fuel new rounds of deliberations.

Next up after a few rounds of deliberations over issues related to service quality came a network meeting specifically for the municipalities' chief and deputy executive officers. The municipalities were asked to review the lessons they had drawn from the network meetings and to identify areas for improvement. Based on these presentations the network partners discussed whether or not to continue cooperating beyond the project phase. A few networks decided to split up, most only reconfigured a bit, and some got new network guides. Many networks started again from the top with measurement activities to get over-time data, and most agreed on a joint plan of action to learn from each other's successes and failures in finding courses of action and implementing changes.

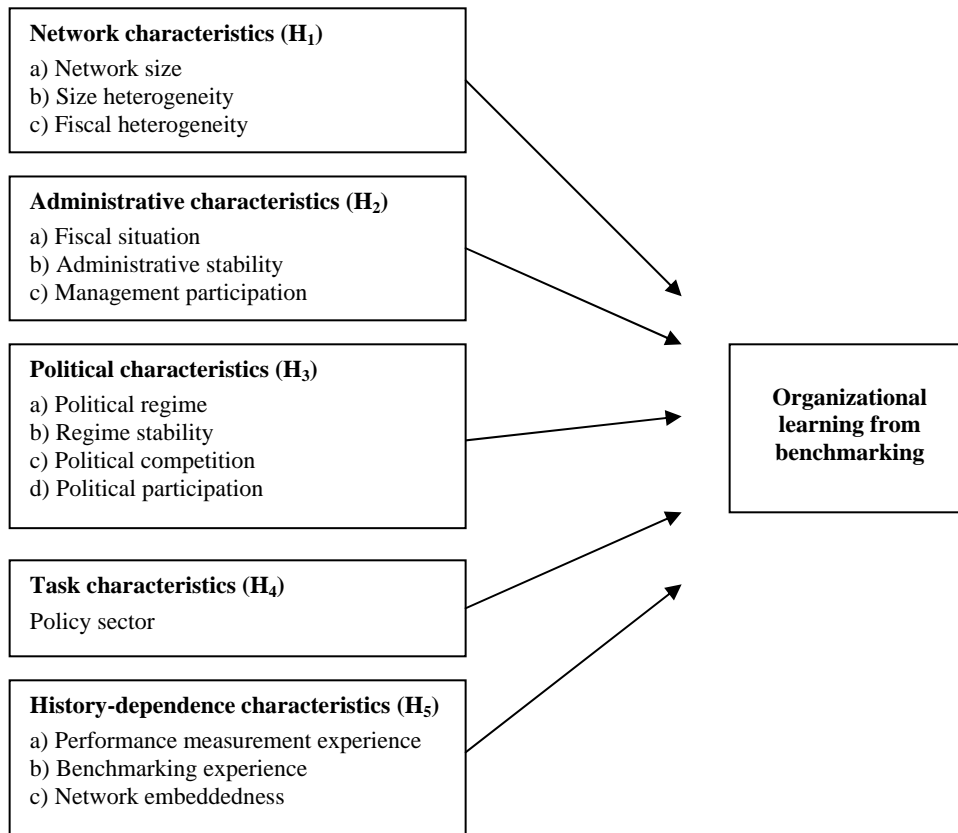
Conceptual framework and hypotheses

In this section we review works on network characteristics, organizational characteristics including political factors, and history-dependence to find potential explanations for

variety in the level of learning a municipality manages to extract from benchmarking. We formulate in total 14 hypotheses that later in the article will be tested against evidence from the Norwegian benchmarking program.

Figure 1 illustrates the analytic framework developed in the preceding sections. Our dependent variable is the level of learning achieved by the individual municipality after having participated in the benchmarking program Networks for efficiency. We have identified 14 factors that we assume influence this learning outcome. These hypotheses are denoted H_1 – H_5 in Figure 1, and are for illustrative purposes placed in boxes for network, administrative, political, task, and history dependence characteristics.

Figure 1: Conceptual framework: What affects organizational learning from benchmarking activities?



Network characteristics

Inter-organizational learning is recognized as important in facilitating improvement in public services. A wide array of contemporary improvement efforts are premised upon the belief that governmental organizations can and will learn from each other. Examples are pilot initiatives, demonstration projects, beacon organizations, and, of course, benchmarking (Hartley and Allison 2002). Inter-organizational benchmarking can take place in more or less structured settings – from ad hoc bilateral (or even unilateral) comparisons to highly structured benchmarking sessions within groups of organizations. If these groups are stable over time we can call them networks. There is some research on what makes benchmarking partners choose each other (Fedor et al. 1996), but fairly little is known about the effects on learning outcomes of different network configurations. Our first and rather crude hypothesis is that network configuration matters for learning outcomes (**H₁**).

One factor of interest is the size of the network. By size we mean the number of municipalities that participate in the network that structures the benchmarking activities. On the one hand, it is imaginable that increasing network size has a negative effect on learning, due to increased coordination costs. On the other hand, a larger network broadens the avenue of experiences the municipality is exposed to through the benchmarking activities, and this might broaden the scope for learning. On balance, we assume that increasing network size increases the individual municipality's learning outcome from the benchmarking activities (**H_{1A}**).

Networks also vary in terms of partner heterogeneity. While some benchmarking networks are made up of quite different partners, others are very homogenous. The units within the organizational field of local governments, i.e. the municipalities, are probably more similar than companies within a subset of the private sector, and especially so within a given country, due to a shared legal and fiscal framework. Inter-municipal differences of interest to the issue at hand are nonetheless hardly exhausted. One candidate is municipal size. The 435 Norwegian municipalities vary in size from less than 300 to more than 500,000 inhabitants. A second source of network heterogeneity is fiscal situation. Due to among other things variety in social benefits demand and revenues from energy production, Norwegian municipalities vary from practically bankrupt to very

wealthy. Several authors argue that partner heterogeneity has a positive influence on inter-organizational learning. Some emphasize the value of learning from contexts outside the organization's usual frame of reference (Camp 1989, 1995) and others that dissimilar referents may be preferable because they may help blunt the affective consequence of comparisons (Fedor et al. 1996). We therefore hypothesize that increased size heterogeneity within the network has a positive impact on organizational learning (**H_{1B}**). For the same reasons, we expect to find a positive effect also from heterogeneity concerning the fiscal situations of the municipalities in the network (**H_{1C}**).

Municipality-level characteristics

Network configuration is an endogenous factor, seen from the perspective of the people that design a given benchmarking structure. At the level of the individual municipality, there are fewer factors of this sort that are potentially influential for the learning outcome of benchmarking experiences. Exceptions are those associated with setting up what Moynihan (2005) calls internal learning forums. We will cover two such factors here, but also a range of more exogenous ones, some covering administrative and others political municipal characteristics.

Administrative characteristics

Our basic assumption is that characteristics associated with the municipality's administrative situation and leadership influence the level of learning it draws from benchmarking activities (**H₂**). The first municipal-level factor we want to highlight is the fiscal situation of the polity. While some assume that a weak fiscal situation may serve as motivation to change (Hood 1995), others hold that organizational learning requires interpretation effort and i.e. dedication and time, which might difficult to mobilize among organizations with low slack. A sound fiscal situation can also mean that the organization is less on the defensive and has a lower guard. We hence expect a strong fiscal situation to be associated with a high level of learning from benchmarking activities at the level of the individual municipality (**H_{2A}**).

Municipalities vary considerably in their level of turnover among executives. Some change Chief executive officers (CEOs) and other top administrators every other year or so while others keep them for decades. A high level of executive stability may on the one

hand have a negative impact on inter-organizational learning, including learning from benchmarking, as a very stable municipality might be excessively self-referencing. Moreover, a new CEO might want to latch on to lessons from benchmarking efforts in order to break with the past and make changes that may help establish his/her position. On the other hand, benchmarking, like all change initiatives, requires resources and persistence to have an impact on the organization. We can imagine administrative stability to increase the ability to bring ideas and initiatives from benchmarking onto the local agenda and to implement any resulting policies aimed at change. On balance we therefore expect to find a positive effect of stability, i.e. that municipalities that changed their CEO during the benchmarking project period experienced less organizational learning than municipalities with stability at the top (**H_{2B}**).

We furthermore assume that municipal commitment and inclination to organizational learning increase with the degree of participation by top level management. We expect municipalities where the CEO took active part in follow-up activities to experience much learning from the benchmarking efforts (**H_{2C}**). Patterns of participation relate to what Moynihan (2005) calls an organizational learning forum. Learning forums are routine events that are established with the intention to consider and discuss performance information. Based on studies of US performance management reforms, Moynihan points to a common assumption that “performance information will automatically become a factor in existing decision processes” (Moynihan 2005:211) and claims that this is what currently guides designers of performance management systems. He goes on to argue that this assumption is counterproductive to organizational learning and that performance management designers need to start taking learning forums internal to the organization as seriously as they do the routines that simply collect and disseminate data.

Political characteristics

Some might argue that politics increases the likelihood of experiential learning because politicians need to learn from successes and failures to secure reelection. Dekker and Hansén (2004) argue that politics may be incorporated into organizational learning and may actually contribute positively to learning processes within the public sector. Others

will hold that one thing that separates politicians from administrative leaders is that the former have less regard for the experiential record of the polity; they only focus on the future. Both positions hold the view that we cannot fully understand the dynamics of organizational learning in the public sector without considering its broader political environment. This is a view that is shared by this paper. We expect political features associated with the individual municipality to affect the level of learning it extracts from participating in the benchmarking program (**H₃**). More specifically, we expect to find effects on learning from the “color” and stability of the political regime in place, the level of political competition, and from politicians’ direct involvement in the benchmarking effort. We will outline each of these sub-hypotheses in turn.

First, one might assume that benchmarking, as a practice associated with NPM, is championed by parties on the right hand side of the political spectrum. If this is true, we can expect municipalities that are ruled by such parties to make more use of this practice than municipalities ruled by centre or left-side parties. An earlier study of municipal performance measurement in Norway has however showed a positive relationship between left-wing party regime and applications of performance measurement (Johnsen 1999), and we recall that the benchmarking programme in question was in fact initiated by a Labor government. Performance measurement and benchmarking can be seen as the kind of “soft” NPM policies that some interpret as defensive strategies to avoid more radical NPM policies heralded by the political right, such as contracting out (Grønlie 1998). We hence expect to find a negative relationship between right-side political regimes and levels of organizational learning from benchmarking (**H_{3A}**).

The benchmarking project studied in this article straddled two electoral periods, before and after the 2003 municipal elections, and hence potentially two municipal councils with different parties in control. We assume that regime stability may make it easier to extract learning from benchmarking because a stable regime have vested interests in past performances and hence an incentive to learn from these. A new regime on the other hand, will want to start with clean sheets and have much less interest in learning from performances associated with the previous regime. Hence, we expect regime stability to increase the individual municipality’s learning outcome from the benchmarking activities (**H_{3B}**).

A further political characteristic that varies between municipalities is the level of conflict in the municipal council. Some polities resemble cozy clubs of old acquaintances while others are virtual war zones. As benchmarking experiences require interpretation among organizational participants in order to have an impact on preferences and on organizational behavior, we can assume the level of political conflict to influence levels of learning from benchmarking. More conflict makes it less likely that interpretations of past experiences are shared within the polity. The best attainable operationalization of conflict is the level of political competition, which is considered high when political blocks are equal in size. We expect a high level of political competition in the municipal council to be associated with a low level of organizational learning from benchmarking experiences (**H_{3C}**).

A final political characteristic that can be assumed to impact on learning is the level of active involvement by politicians in drawing lessons from benchmarking experiences (cf. Dekker and Hansén 2004). This expectation is equivalent to the one formulated above about top administrator involvement, and it can be viewed as another potentially important feature of municipal learning forums. We expect to find that municipalities where politicians took active part in follow-up activities experience more learning from the benchmarking efforts (**H_{3D}**).

Task characteristics

Several contributors to the dialogue on the link between information and decisions have argued that there is reason to expect systematic differences between policy sectors due to sector specific differences in the tasks that are to be accomplished (Askim 2004; ter Bogt 2003; Johanssen 1995; Greve 2003; van Helden and Johnsen 2002; Macintosh 1985, 1994; Meyer and Rowan 1977; Moynihan 2005; Wilson 2000). The proposed implication for this article is that municipalities' learning from benchmarking efforts will vary between policy sectors. Most scholars develop typologies of tasks and argue that each type can be expected to be associated with a certain level or type of utilization of performance information. Most municipal task are however hard to characterize according to any typology. To take the case of elderly care, the accomplishment of this task comprises component activities that are both hard and soft (rendering Johanssen hard

to apply in our case); both highly complex and less complex (ditto Greve); and both observable and less observable (ditto Wilson and Macintosh).

An expectation of less sophistication but with presumably higher face value is that the impact of benchmarking will decrease with increased resistance from among the “street-level bureaucrats”. Our survey covers benchmarking experiences from the sectors elderly care and primary education. Given prior experience with extensive resistance towards performance measurement from teachers and teacher unions in Norway, we expect the level of learning drawn from benchmarking experiences to be lower in primary education than in elderly care (**H₄**).

History dependence characteristics

Some hold that the ability to learn from others in a network is a matter of mutual dependency in the present (Agranoff and McGuire 2001). Others argue that these are history dependent processes. Cohen and Levinthal (1990) argue that the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends is critical to its innovative capabilities. This capability was labelled a firm’s absorptive capacity. The development of absorptive capacity may depend on individual as well as organizational factors, and it is history (or path) dependent in the sense that it is largely a function of knowledge built through prior related experiences.

We assume that the dynamics of external information utilization is similar in public sector organizations, including municipalities. Coupled with historical municipality-level data, the Norwegian benchmarking program provides an opportunity to study the impact of experiential history on organizational learning. Our hypothesis is that a municipality’s ability to recognize the value of information from benchmarking, and to assimilate and apply it, is influenced by knowledge built through prior related experiences (**H₅**).

First, some municipalities have a long history of using performance information for planning or reporting purposes, while others were fairly new to this at the time of the initiation of the benchmarking program. We expect to find that municipalities that have much experience with utilizing performance information have better prospects of learning from benchmarking efforts (**H_{5A}**). A second expectation relates more specifically to previous experience with the practice of benchmarking. We hypothesize that

municipalities with more benchmarking experience have built what we can call a method specific learning capacity (cf. Fedor et al. 1996), and that this will increase the municipality's learning outcome from present benchmarking activities (**H_{5B}**).

Finally we assume that inter-municipal relations too can have a history, and that this can influence present dynamics in a network. Some might argue that new acquaintances are most promising because they expose the organizations to new ideas, but others may argue that embedded ties perform better. We follow the latter assumption, and expect previous structured interactions with the same municipalities that participate in the network in the benchmarking effort at hand increase the municipality's learning outcome from benchmarking activities (**H_{5C}**). We can call this a partner specific learning capacity hypothesis.

Research design, data and methods

Our material on the Networks for renewal and efficiency program provides a promising opportunity to research such network effects as hypothesised above, as the program encompassed more than 300 municipalities in more than 40 networks. We conducted a survey among the 114 municipalities of these municipalities that participated in networks for primary education and the 142 that participated in networks for care for elderly and disabled, in total 256 municipalities. Distinct electronic questionnaires (Questback) were developed for each policy sector, and the municipalities were instructed to answer for only one. The questionnaires were sent by email to the CEO and one senior manager within the selected policy sector in December 2005. In the cases where we received more than one reply from a municipality, we chose to use the reply from the CEO. There were two rounds of following up non-response, and by mid January 2006 we had received replies from 279 respondents from 203 municipalities. The municipal response rate was 79.3 percent (203 of 256 municipalities).

The dependent variable in our study is organizational learning, and in accordance with the definition discussed above, our operationalization acknowledges that not all learning will materialize itself in organizational change. We measure the dependent variable by six indices, each with a potential range from 0 to 12. Three of the indices correspond to a distinction between three functions that information and experiences from

benchmarking can serve in a municipality. These are agenda-setting, decision support and solution. The fact that an issue is put on the municipal agenda as a result of benchmarking or that benchmarking experiences are used as documentation in an ongoing decision process concerning a topic does not necessarily mean that the organization will transform or even adjust its policies or actions. Change *might* prove to be the outcome of a decision process, but it needs not be. Either of these two applications nonetheless satisfies our definition of organizational learning. Solution denotes situations where benchmarking experiences have had decisive influence on the content of a new policy or a new routine in a municipality.

The remaining three indices cover the municipalities' use of information and experiences from benchmarking in their treatment of strategic, structural and operational issues, respectively. The survey contained 12 questions on usage of experiences from the Network for renewal and efficiency. The survey questions were dummy variables where non-use is coded "zero" and use is coded "one". The first four questions were related to *strategic issues*, defined as overarching questions related to municipal service production and priorities between programs. These questions were related to changes in budget allocations between policy sectors, changes in budget allocations between service units within a policy sector, changes in budget allocations between user groups within a policy sector, and ways of co-operation between policy sectors. The next four questions were related to *structural issues*, defined as questions related to organization and routines for planning and reporting on municipal and/or policy sector level. These questions were related to organizational issues, co-operation between service units within the same policy sector, quality of systematic management information, and usage of systematic management information in planning and reporting documents. The final four questions were related to operational or *practical issues*, defined as organization and structuring of matters concerning users and/or employees more directly. These questions were related to working conditions for the employees, the interface between employees and users, the interface between the CEO and service unit managers, and the flow of information about the services to users and their relatives. Each of the 12 questions was also mapped according to the three categories of the first dimensions (agenda-setting, decision support, and solution).

The 12 questions are obviously not exhaustive relative to the universe of issues that may be affected by benchmarking experiences, but the sample is developed based on consultations with key actors in the municipal field such as representatives from the Association of Local Authorities and municipal managers.

Data sources and operationalizations of the independent variables are documented in Table 1.

As the data structure represents two levels, municipalities and networks, multilevel analysis is in principle relevant. The dependent variable can be expected to have a municipality level component as well as a network level component. The number of municipalities within the networks is however too low to satisfy the requirements for multilevel analysis (Hox 2002). The data are hence analyzed using multiple (OLS) regression analysis.

Table 1: Independent variables, data sources and operationalizations

Variable	Source	Scale	Operationalization
Municipal size	Statistics Norway	Interval	Municipal inhabitants (1,000)
Network size	The Association of Local Authorities	Interval	Number of municipalities participating in the networks
Size heterogeneity	The Association of Local Authorities and Statistics Norway	Ratio	Sum of distance (absolute number) between number of inhabitants of the municipalities in the network and mean number of inhabitants in the network.
Fiscal heterogeneity	The Association of Local Authorities and Statistics Norway (KOSTRA)	Ratio	Sum of distance (absolute number) between fiscal situation (see below) of the municipalities in the network and mean fiscal situation in the network (1,000 NOK)
Fiscal situation	Statistics Norway (KOSTRA)	Ratio	Net operating result per capita, average 2001–2004 (1,000 NOK)
Administrative stability	Survey	Nominal	Change of Chief Administrative Officer (CEO) 2002–2005. Dummy variable: 0=no, 1=yes. Inversed.
Management participation	Survey	Nominal	Active participation of CEO in the utilization of municipal experiences from participation in the Networks for renewal and efficiency. Dummy variable: 0=no, 1=yes
Political regime	Norwegian Social Science Data Service	Nominal	Municipality with a mayor that represents the right side of the political spectrum, i.e. the Conservative party (<i>Høyre</i>) or the Progressive party (<i>Fremskrittspartiet</i>) = 1, all other parties = 0
Regime stability	Norwegian Social Science Data Service	Nominal	Change in major's party 2002–2005. Dummy variable: 0=no, 1=yes
Political competition	Statistics Norway	Interval	Party concentration in the municipal council. Herfindahl index ^a . Inversed.
Political participation	Survey	Nominal	Active participation of politicians in the utilization of municipal experiences from participation in the Networks for renewal and efficiency. Dummy variable: 0=no, 1=yes
Policy sector	Survey	Nominal	Dummy variable: 0=primary school, 1=health care and care for elderly and disabled
Performance measurement experience	The Ministry of Local Authorities and Regional Development Municipal organization database	Nominal	Municipal use of performance measurement in 2000. Dummy variable: 0=no, 1=yes
Benchmarking experience	The Ministry of Local Authorities and Regional Development Municipal organization database	Nominal	Municipal use of benchmarking in 2000. Dummy variable: 0=no, 1=yes
Network embeddedness	Telemark Research	Interval	Summary of partner (municipality) overlaps (in n) between the present network and up to ten structured inter-municipal efforts (<i>interkommunale samarbeid</i>) the municipality has participated in and judges as successful. Theoretical maximum value is equivalent to the maximum number of municipalities in a network

Notes:

^a The Herfindahl-index measuring party concentration is defined as:

$$H = \sum_{j=1}^P S_j^2,$$

where S_j is the share of representatives from party j in the municipal council and P is the number of parties. The Herfindahl-index takes the value 1 when a single party holds all seats in the council. The minimum value is $1/P$ when the seats are equally divided among the P parties.

Results and discussion

Before we present and discuss the regressions, we document a univariate and a bivariate analysis using correlations.

Table 2: Descriptive statistics

	N	Min	Max	Mean	Std. dev.	Skewness	Kurtosis
Strategy	193	0.00	12.00	3.50	2.53	0.69	0.37
Structure	198	0.00	12.00	3.57	2.70	0.62	-0.07
Practice	189	0.00	10.00	2.66	2.51	0.84	-0.03
Agenda-setting	202	0.00	12.00	3.91	3.38	0.53	-0.75
Decision support	202	0.00	12.00	3.95	3.44	0.63	-0.54
Solution	202	0.00	12.00	1.51	2.38	2.05	4.77
Network size	178	2.00	11.00	6.14	1.62	0.97	1.59
Fiscal heterogeneity	169	0.03	21.10	2.91	2.74	3.67	21.43
Size heterogeneity	169	508.00	350173.71	28383.04	67879.97	4.47	18.76
Fiscal situation	171	-18.72	1058.13	6.67	80.90	13.07	170.81
Administrative stability	198	0.00	1.00	0.62	0.49	-0.50	-1.76
Management participation	202	0.00	1.00	0.82	0.38	-1.69	0.88
Political regime	201	0.00	1.00	0.20	0.40	1.52	0.31
Political stability	202	0.00	1.00	0.50	0.50	0.02	-2.02
Political competition	202	0.14	0.52	0.77	0.07	-1.77	3.59
Political participation	202	0.00	1.00	0.47	0.50	0.12	-2.01
Policy sector	202	0.00	1.00	0.57	0.50	-0.30	-1.93
Perf. measurem. experience	171	0.00	1.00	0.35	0.48	0.66	-1.59
Benchmarking experience	171	0.00	1.00	0.54	0.50	-0.15	-2.00
Network embeddedness	142	0.00	6.23	1.43	1.51	1.22	0.69

Table 2 documents descriptive statistics for all the variables. Organizational learning varied between the six indices. The mean value is lowest for the index denoted “solution”, which measures whether benchmarking experiences have had decisive influence on the content of a new policy or a new routine in a municipality. This reinforces our belief that studies of benchmarking-induced learning need to include cognition as well as behavior/change, so as not to underestimate the magnitude of organizational learning. The three network characteristics variables varied much between the municipalities. In the Networks for renewal and efficiency program the participating municipalities were initially organized in groups of 4–8, and these networks were fairly stable over the course of the project period. Our data shows that though most were within this range, network sizes in fact ranged from 2 to 11. 38 percent of the municipalities

experienced changes in the CEOs, but managerial participation in the networks was high. CEOs participated in 82 percent of the municipalities. 20 percent of the municipalities had a major from the Conservative Party or the Progress Party, indicating a right-wing political regime, and 50 percent of the municipalities had experienced a change in political regime. Reflecting the Norwegian multi-party system, the party concentration in our material is low, indicating a high level of party competition. Note, however, that local politics in Norway is even more consensual based than national politics. The active participation of politicians of 47 percent was somewhat lower than the managerial participation. A majority of 57 percent of the municipalities participated in the networks for health care and care for elderly and disabled with the remaining 47 percent in networks for primary schools. Only 1/3 of the municipalities reported experience with performance measurement but more than half reported experience with benchmarking. The extent of embedded network experiences was relatively low.

Table 3 documents the correlations between all the variables.

The fact that the data for the independent variables stem from different sources (cf. Table 1) proved problematic for the regression analysis using the full conceptual model as specified in Figure 1, as some variables, and in particular the variables regarding history-dependence, had many missing values. Table 4 documents a regression analysis without the history-dependence variables (**H₅**). A regression with the full model, but with a reduced number of cases, is documented in Table 5.

Table 3: Descriptive statistics and correlations (N=90)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1 Strategy																				
2 Structure	.71**																			
3 Practice	.49**	.53**																		
4 Agenda-setting	.66**	.59**	.43**																	
5 Decision support	.65**	.63**	.72**	.17**																
6 Solution	.56**	.73**	.50**	.21**	.45**															
7 Network size	-.05	.00	-.05	.00	-.01	-.08														
8 Size heterogeneity	.03	.20*	.15	-.09	.20*	.26**	.13													
8 Fiscal heterogeneity	.28**	.09	.01	.21**	.07	.03	.30**	.06												
9 Fiscal situation	.03	-.15	.06	.06	.01	-.19*	.05	-.52**	.01											
10 Administrative stability	.05	.20*	.16	-.01	.21**	.15	.03	.14	.01	-.09										
11 Management participation	.20*	.27**	.21*	.21**	.17	.20*	-.02	.12	.08	-.04	.07									
12 Political regime	.02	-.02	.02	-.06	.13	-.08	-.05	.02	.22**	.05	-.24**	.19								
13 Political stability	.01	-.07	.01	-.04	-.02	.03	-.01	-.14	.05	.02	-.07	.00	-.04							
14 Political competition	.02	.04	-.06	-.00	-.06	.10	-.00	-.12	.16	-.01	.02	-.06	.15	-.11						
15 Political participation	.16	.23**	.19*	.17	.26**	.02	.06	.08	.05	-.02	.00	.41**	.17	-.01	-.05					
16 Policy sector	.16	-.02	.10	.10	.08	.00	-.35**	-.25**	.16	.12	.01	-.16	-.05	.10	.12	-.06				
17 Perf. measurem. experience	.09	.06	.00	.09	.01	.03	.11	-.18*	.24**	.06	-.03	.06	.10	.15	.09	.08	.04			
18 Benchmarking experience	.18*	.21**	.00	.11	.13	.08	.06	.11	.20*	-.10	.02	.12	.07	.15	.02	.13	.05	.33		
19 Network embeddedness	-.05	-.11	.00	-.05	.05	-.17	.28**	.11	-.04	.02	.09	-.06	.04	.01	.08	.10	-.10	-.06	.06	

Notes:

*Significant at level $p < .10$, **Significant at level $p < .05$ (two-tailed p-values)

Table 4: Regression of independent variables on organizational learning from benchmarking among Norwegian municipalities

Independent variables	Dependent variables					
	<i>Strategy</i>	<i>Structure</i>	<i>Practice</i>	<i>Agenda-setting</i>	<i>Decision support</i>	<i>Solution</i>
Constant	2.12	2.63	-0.42	1.58	0.40	2.37
<i>Network characteristics</i>						
Network size	-0.12	-0.13	0.07	-0.07	0.10	-0.19
Size heterogeneity	0.01**	0.00	-0.01	0.00	0.00	0.00
Fiscal heterogeneity	0.05	0.16**	0.22**	-0.13	0.37**	0.21**
<i>Administrative characteristics</i>						
Fiscal situation	0.09	-0.02	0.19**	0.03	0.23*	-0.03
Administrative stability	0.37	1.25**	0.50	0.54	0.54	0.65*
Management participation	0.98*	1.61**	1.31**	1.71**	0.64	1.28**
<i>Political characteristics</i>						
Political regime (right=1)	-0.56	-0.94*	-0.41	-1.06*	0.34	-0.90**
Political stability	-0.20	-0.13	-0.30	-0.13	-0.04	0.31
Political competition	-2.91	0.54	-0.73	-4.11	0.57	5.31*
Political participation	0.62	1.09**	0.69*	0.94*	1.59**	0.12
<i>Task characteristics</i>						
Sector (elderly care=1)	0.36	0.03	1.41**	0.73	1.08**	0.17
<i>Model fit^a</i>						
R-square	0.13	0.21**	0.23**	0.12	0.20**	0.17**
R-square, adjusted	0.05	0.13	0.16	0.05	0.13	0.10
N	130	135	128	137	137	137

Notes:

* Significant at level $p < .10$, ** significant at level $p < .05$ (one-tailed p-values).

^a Tolerance values range from .719 to .746 between the six models (dependent variables), indicating an acceptably low level of multicollinearity for the explanatory framework.

Network characteristics

Two of the three sub-hypotheses regarding network characteristics were supported by the data, strengthening our assumption that network configuration matters for learning outcomes (**H₁**). In **H_{1A}** we assumed that increasing network size would increase the individual municipality's learning outcome from the benchmarking activities. This does not seem to be the case in the empirical setting studied. Effects are weak and show unstable signs across the six models. This non-finding might suggest that in cases of larger networks, the positive effect of a broadened learning scope is cancelled out by the negative effect of increased coordination costs. It might however also imply that variance

in network size needs to be larger to have substantial impact on learning outcomes, or that network size simply does not matter.

Our second expectation concerning the influence of network configuration, **H_{1B}**, was that increased heterogeneity within the network regarding municipal size would have a positive impact on organizational learning. The results show that this is the case in five of six models, and significantly so for the municipalities' treatment of strategic issues. The effects may appear small, but recall that this independent variable has a very broad range in values.

Finally, in **H_{1C}**, we expected the individual municipality's learning outcome from benchmarking to increase with the fiscal heterogeneity within the benchmarking network. This hypothesis is confirmed by the data. We find significant effects of this type of heterogeneity in four of the six models. It can be, as suggested above, that it is more valuable to learn from contexts that are outside one's usual frame of reference, and/or that dissimilar referents may be preferable because they may help blunt the affective consequence of comparisons. This would imply that fiscal differences are more important than differences in size between the municipalities, as we did not find this effect for size heterogeneity.

Administrative characteristics

The results show that characteristics associated with the municipality's administrative situation and leadership, aspects close to so-called learning forums (Moynihan 2005), do influence the level of learning it draws from benchmarking activities (**H₂**). All three hypotheses regarding this were supported by the data, though to different extents. We do find that a strong fiscal situation is associated with a high level of learning from benchmarking (**H_{2A}**). This relationship is positive in four of the six models, significantly so in two. Our interpretation is that organizational learning requires interpretation effort and i.e. the kind of dedication and time it is difficult to mobilize in organizations with low slack. We also find a positive effect of stability at the top. Municipalities that changed their CEO during the benchmarking project period experienced less organizational learning than municipalities that did not (**H_{2B}**). This pattern is consistent across the six models, statistically significantly in two. Finally, we find that

municipalities where the CEO took active part in follow-up activities experienced more learning from the benchmarking efforts (H_{2C}). This hypothesis was significantly corroborated in five of the models.

Political characteristics

Three out of four hypotheses concerning politics were supported by the data, reinforcing our belief that internal political features do affect the level of learning the municipality manages to draw from benchmarking (H_3). First, we do find the expected negative relationship between right-side political regimes and levels of organizational learning from benchmarking (H_{3A}). This effect is stable in terms of signs in five of the six models, and significant in three. Our interpretation is that right-sided regimes may prefer more radical solutions like vouchers and contracting out over benchmarking, while centre and left-sided regimes use benchmarking actively, to implement solutions within the traditional scope of public services, and as a defensive strategy to ease the pressure for the radical NPM reforms favored by the right.

Secondly, we find that regime instability seems to decrease the individual municipality's learning outcome from benchmarking (H_{3B}). The relationship is not significant but the effect has the expected negative sign in five of the six models. It might be that stable regimes have vested interests in past performances and therefore have stronger incentives to learn from these than new regimes, which prefer to start with clean sheets. Though not tested here, we can speculate that an incumbent regime will have more vested interests in successes than in failures and that regime stability therefore is particularly associated with high levels of experiential learning when benchmarking casts a positive light on past performances. Part of the reason can be that this might improve the likelihood for re-election (Downs 1957).

Political competition in the municipal council does not seem to have a strong or stable effect on organizational learning from benchmarking experiences (H_{3C}). One interpretation of this is that the dynamics of learning from benchmarking simply do not activate disagreement along party lines. It is also possible that the parallel to parliamentarianism is stretched or flawed in the empirical setting studied. The fact is that a few big cities apart, municipalities in Norway do not have parliamentarian models; they

are organized with a proportionally elected municipal council (*kommunestyre*) and a proportional executive committee (*formannskap*) consisting of about 1/3 of the municipal council members. And mayors, who are normally appointed from the largest party, do not hold much formal power. This means that the division between position and opposition is more blurred in local government than in national politics. Decisions are consensual more often than not.

The strongest finding regarding politics is the positive effect on learning outcomes from the active participation by politicians in follow-up activities (**H_{3D}**). The effect of participation is positive in all and significant in four of the six models. This is another confirmation of the importance of internal learning forums.

Task characteristics

H₄ predicted municipalities' learning from benchmarking to vary systematically between policy sectors. We expected elderly care to achieve more organizational learning than primary schools. This hypothesis was corroborated in all six models and significant in the models concerning practical issues and decision support. A thorough discussion of this finding requires relating it to the task typologies discussed above and ideally re-testing in more policy sectors than the two scrutinized here. This should be pursued in further research. For now we can suggest that the impact of benchmarking is stronger in elderly care than in education because resistance towards performance measurement has been stronger among teachers and teacher unions than among their peers in the elderly care sector. A second explanation, equally "close" to the empirical setting studied, is that the care sector in Norway has been exposed to wide-ranging performance measurement for a longer period of time.

History-dependence

Table 5 documents regressions of the full explanatory framework illustrated in Figure 1. This analysis was separated from Table 4 because of the low number of available cases encompassing the history-dependence variables. We have specified summary statistics and not regression coefficients in the table, as that is all we need to make the point, which is that history doesn't seem to matter. Adding history-dependence to the regressions does not increase the explanatory power of our framework. It is tempting to conclude that

Agranoff and McGuire (2001) are right: The ability to learn from others in a network is a matter of dynamics in the present and not of absorptive capacity built through past experiences.

Table 5: Regressions of independent variables, including history dependence, on organizational learning from benchmarking. Model statistics only.

Model fit	Dependent variables					
	Strategy	Structure	Practice	Agenda-setting	Decision support	Solution
<i>Networks characteristics (model 1)</i>						
R-square	0.10**	0.04	0.03	0.06	0.05	0.08*
R-square adjusted	0.07	0.02	0.00	0.03	0.02	0.05
R-square change	0.10**	0.05	0.03	0.06	0.05	0.08*
<i>Municipality-level characteristics (model 2)</i>						
R-square	0.17	0.20**	0.14	0.14	0.17*	0.18*
R-square adjusted	0.06	0.11	0.03	0.04	0.07	0.09
R-square change	0.07	0.16**	0.11	0.08	0.12*	0.10
<i>Task characteristics (model 3)</i>						
R-square	0.17	0.20**	0.16	0.15	0.19*	0.19*
R-square adjusted	0.06	0.10	0.04	0.04	0.08	0.09
R-square change	0.00	0.00	0.02	0.01	0.02	0.01
<i>History-dependence (model 4)</i>						
R-square	0.18	0.23*	0.17	0.15	0.19	0.21*
R-square adjusted	0.04	0.10	0.01	0.01	0.06	0.08
R-square change	0.01	0.02	0.01	0.01	0.00	0.03
N	94	96	91	98	98	98

Notes:

*Significant at level $p < .10$, **Significant at level $p < .05$

Model 1 Network size, Fiscal heterogeneity, Size heterogeneity

Model 2 Model 1 + Fiscal situation, Administrative stability, Management participation, Political regime, Political stability, Political competition, Political participation

Model 3 Model 2 + Policy sector

Model 4 Model 3 + Performance measurement experience, Benchmarking experience, Network embeddedness

The low number of cases included in this analysis does however give reason for caution, as do some of the bivariate correlations documented in Table 3. A municipality's previous experience with performance management correlates consistently positively with all six indices for organizational learning, though not significantly. Previous experience with inter-municipal benchmarking does correlate significantly with learning, more specifically with the indices that measure municipal usage of benchmarking for treatment of strategic and structural issues respectively. This gives reason to urge more research to be done on the possible history-dependency of organizational learning from

benchmarking. Further research might also test interesting hypotheses put forward in the literature, like that of Fedor et al. (1996), who argue that more experience an organization has with benchmarking reduces the impact of network configuration on learning. The low number of cases did not allow us to conduct meaningful tests of such interaction effects in this research.

A full test of the partner specific learning capacity hypothesis must also be postponed for future research. What we can read out of the bivariate correlations documented in Table 3 is that network embeddedness has a consistently negative correlation with all six indices for organizational learning, though not significantly. This indicates that new acquaintances are the most promising benchmarking partners, and we can speculate that the reason is that they expose the organizations to novel ideas and ways of working.

Conclusions

The practice of benchmarking is carried around the world on the back of a promise that it supports innovation in organizations, via vicarious organizational learning. But while generic prescriptions continue to abound, the descriptive understandings that underpin this perceived means–end relationship remain weak due to a lack of empirical knowledge of where benchmarking works, and why. This article builds on previous attempts to map and understand the chain between benchmarking, learning and improvement and to identify factors that condition organizational learning from benchmarking. As a first step, we have developed a framework for analyzing organizational learning outcomes from municipal benchmarking exercises. The framework draws on existing research and incorporates explanatory factors at different levels (network and municipality), with different time perspectives (past and present) and with different potential practical implications (endogenous and exogenous).

As a second step we have tested the framework with survey data covering experiences from a nationwide benchmarking project for local governments in Norway. Model fits were relatively modest, which probably at least in part is due to a fairly low number of available cases. But regressions identified several interesting effects. One is that network configuration matters for learning outcomes. The article demonstrates that

municipalities seem to draw more learning from the benchmarking exercises when networks are heterogeneous, i.e. when benchmarking partners are dissimilar. One way to see this is to draw on a term from Moynihan (2005) and say that we have identified an important factor relating to the inter-municipal learning forum that is set up as part of the infrastructure for the benchmarking exercises. We have also found effects of several factors relating to learning forums internal to the individual municipality. Administrative slack, administrative and political regime stability and managerial and political involvement in follow-up activities are all factors that impact positively on the level of learning drawn from benchmarking.

The article also demonstrates a negative relationship between right-side political regimes and levels of organizational learning from benchmarking. Our interpretation is that right-sided regimes may prefer more radical solutions like vouchers and contracting out over benchmarking, while centre and left-sided regimes use benchmarking actively, to implement solutions within the traditional scope of public services, and as a defensive strategy to ease the pressure for the radical NPM reforms favored by the right.

The article's findings have ramifications for practitioners in public administration. An upbeat practical implication is that there are a number of factors that are endogenous for designers of benchmarking efforts, like network configuration and patterns of participation, that influence outcomes. Related to this is a finding that the ability to learn from others in a network appears to be a matter of dynamics in the present, and less so of capacity built through past experiences.

The article aspires to contribute to the scholarly dialogue, but we have also put forward suggestions for future research. The number of variables was large relative to the available cases in our study. We therefore call for future research using the developed framework with more extensive data. Special care should be taken to investigate the effects of history-dependence variables, which was only done to a limited extent here..

A final contribution of the article is an insistence on conceptualizing and operationalizing organizational learning so that one can capture effects that are both cognitive and behavioral in nature. We have found three functions that information and experiences from benchmarking can serve in a municipality. These are agenda-setting, decision support and solution. Solution denotes situations where benchmarking

experiences have had decisive influence on the content of a new policy or a new routine in a municipality. The fact that an issue is put on the municipal agenda as a result of benchmarking or that benchmarking experiences are used as documentation in an ongoing decision process concerning a topic does not necessarily mean that the organization will transform or even adjust its policies or actions. Change *might* prove to be the outcome of a decision process, but it needs not be. Either of these two applications nonetheless satisfies our definition of organizational learning, and we argue that restricting the conceptual and operational definitions of learning to changes will underestimate the effects in public administration and politics from performance management reforms such as benchmarking efforts.

Acknowledgements

Thanks to The Norwegian Association of Local and Regional Authorities for financing and supporting the survey among local governments in Norway. Thanks to Anders Nordstoga for assistance in data collection on networks. We are grateful for the comments to an earlier draft to this article from the participants at the annual meeting of the Network for Organization Theory in Norway, Bergen, Norway, 17 November 2005; at a Department of Political Science seminar at University of Oslo, 21 November 2005; and at the Nordic Municipal Researchers Conference, Åbo, Finland, 25–27 November 2005.

References

- Agranoff, Robert and Michael McGuire. 2001. The big questions of public network management research. *Journal of Public Administration Research and Theory* 11 (3): 295–326.
- Ammons, D.N., C. Coe, and M. Lombardo. 2001. Performance-comparison projects in local government: Participants' perspectives. *Public Administration Review* 61 (1): 100–115.
- Askim, Jostein. 2004. Benchmarking in local government service delivery: Window-dressing or a potent driver for improvement? Evidence from Norway. In *Leistungsmessung und -vergleich in Politik und Verwaltung*, eds. Sabine Kuhlmann, Jörg Bogumil and Hellmut Wollmann. Wiesbaden: VS Verlag
- Barzelay, Michael. 2004. Narratives, arguments, and institutional processualism: Learning about implementing presidential priorities from Brazil in Action. Paper delivered at Conference on smart practices towards innovation in public management, University of British Columbia, Canada, June 15–17, 2004.
- Behn, Robert D. 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* 63 (5): 588–606.
- Bogt, Henk J. ter. 2003. Performance evaluation styles in governmental organizations: How do professional managers facilitate politicians' work? *Management Accounting Research* 14 (4): 311–332.

- Bouckaert, Geert .1996. Measurement of public sector performance: Some European perspectives. In *Organizational performance and management in the public sector*, eds. Arie Halachmi and Geert Bouchaert, 223–37. London: Quorum Books.
- Camp, Robert C. 1989. *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, WI: ASQC Quality Press.
- Camp, Robert C. 1995. *Business process benchmarking*. Milwaukee, WI: ASQC Quality Press.
- Cohen, Wesley M. and Daniel A. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35 (1): 128–152.
- Dekker, Sander and Dan Hansén. 2004. Learning under pressure: The effects of politicization on organizational learning in public bureaucracies. *Journal of Public Administration Research & Theory* 14 (2): 211–230.
- Downs, Anthony. 1957. *An economic theory of democracy*. New York, N.Y.: Harper Collins.
- Fedor, Donald B., Charles K. Parsons and Christina E. Shalley. 1996. Organizational comparison processes: Investigating the adoption and impact of benchmarking-related activities. *Journal of Quality Management* 1 (2): 161–192.
- Feldman, Martha S. and James G. March. 1981. Information in organizations as signal and symbol. *Administrative Science Quarterly*, 26: 171–86. In *Decisions and organizations*, ed. J.G. March, 1988, 409–428. Cambridge, MA: Basil Blackwell.
- Fiol, C. Marlene and Marjorie A. Lyles. 1985. Organizational learning. *Academy of Management Review* 19 (4): 803–813.
- Folz, David F. 2004. Service quality and benchmarking the performance of municipal services. *Public Administration Review* 64 (2): 209–220.
- Greve, Henrich R. 2003. *Organizational learning from performance feedback: A behavioral perspective on innovation and change*. Cambridge, UK: Cambridge University Press.
- Grønlie, Tore. 1998. Drømmen om en konkurransetilpasset stat. Ytre fristilling som styringspolitisk redskap, 1945–1995. In *Ein stat? Fristillingas fire ansikt*, eds. Tore Grønlie and Per Selle, 66–106. Oslo: Samlaget.

- Hartley, Jean and Maria Allison. 2002. Good, better, best? Inter-organizational learning in a network of local authorities. *Public Management Review* 4 (1): 101–118.
- Helden, G. Jan van and Åge Johnsen. 2002. A comparative analysis of the development of performance-based management systems in Dutch and Norwegian local government. *International Public Management Journal* 5 (1): 75–95.
- Hood, Christopher. 1995. The new public management in the 1980s: Variations on a theme. *Accounting, Organizations and Society* 20 (2/3): 93–109.
- Hox, Joop. 2002. *Multilevel analysis: Techniques and applications*. London: Lawrence Erlbaum Associates.
- Huber, George P. 1991. Organizational learning: The contributing processes and the literatures. *Organization Science* 2 (1): 88–115.
- Johansson, Staffan. 1995. *Verksamhetsbedömning i mjuka organisationer. Om kommunala ledares verksamhetsinformation och försök att finna mått på effektivitet inom social service*. [Performance Assessment in Soft Organisations]. Stockholm: CEFOS
- Johnsen, Åge. 1999. Performance measurement in local government: Organisational control in political institutions. Doctoral thesis. Bergen: Norwegian School of Economics and Business Administration.
- Johnsen, Åge. 2005. What does 25 years of experience tell us about the state of performance measurement in public management and policy? *Public Money and Management* 25 (1): 9–17.
- Levitt, Barbara and James G. March. 1988. Organizational learning. *Annual Review of Sociology* 14: 319–340.
- Macintosh, Norman B. 1985. *The social software of accounting and information systems*. London: John Wiley and Sons.
- Macintosh, Norman B. 1994. *Management accounting and control systems*. New York: Chichester.
- March, James G. and Herbert A. Simon. 1993 [1958]. *Organizations*. 2nd edition. Cambridge, Massachusetts: Blackwell Publishers.

- Mausolff, Christopher. 1999. *Organizational learning in response to performance monitoring: A study of four employment programs*. PhD Dissertation. University of Pittsburgh.
- Meyer, John W. and Rowan, Brian. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* 83 (2): 340–363. In Walter W. Powell and Paul J. DiMaggio (eds.) (1991) *The New Institutionalism in Organizational Analysis*, pp. 41–62, Chicago: University of Chicago Press.
- Moynihan, Donald P. 2005. Goal-based learning and the future of performance management. *Public Administration Review* 65 (2): 203–216.
- Nicholson-Crotty, Sean, Nick A. Theobald and Jill Nicholson-Crotty. 2006. Disparate measures: Public managers and performance-measurement strategies. *Public Administration Review* 66 (1): 101–113.
- Pollitt, Christopher and Geert Bouckaert. 2004. *Public management reform: A comparative analysis* (2nd edition). Oxford: Oxford University Press
- Pollitt, Christopher. 2005. Performance information for democracy – the missing link? Paper presented at EGPA Annual Conference in Bern, 31.8.–3.9. 2005. Forthcoming in *Evaluation – The International Journal of Theory, Research and Practice*.
- Simon, Herbert A. 1939. The administrator in search of statistics. *Public Management* 21: 106–109.
- Wholey, Joseph S. 1983. *Evaluation and effective public management*. Boston, MA: Little, Brown, and Co.
- Williams, Daniel W. 2003. Measuring government in the early 20th Century. *Public Administration Review* 63 (6): 643-659.
- Wilson, James Q. 2000. *Bureaucracy: What government agencies do and why they do it*. 2nd edition. New York: Basic Books.
- Wolman, Harold and Ed Page. 2002. Policy transfer among local governments: An information-theory approach. *Governance* 15 (4): 477–501.