

**Benchmarking through Municipal Benchmarking Networks:  
Improvement or Leveling of Performance?**

Hans Knutsson (hans.knutsson@fek.lu.se)

Ulf Ramberg (ulf.ramberg@fek.lu.se)

Torbjörn Tagesson (torbjorn.tagesson@fek.lu.se)

Lund University, School of Economics and Management,

Department of Business Administration

P.O. Box 7080, SE-220 07 Lund, Sweden

## **Abstract**

Benchmarking has received much attention in local government. However, benchmarking within a networking environment is quite a recent phenomenon. The aim of the article is to describe municipal benchmarking networks and to analyze whether they lead to improvements in practice. The data are drawn from the National Benchmarking Project (NBP): document studies of the reports, two surveys, three focus groups, and 64 interviews in 12 municipalities. The study shows that the capacity for action-oriented use of benchmarking is weak. Poor performers often blamed their own performance on exogenous factors or methodological flaws. And some of the excellent performers actually used the results as a reason to reduce performance. Thus, the NBP actually triggered a behavior that led to the “average” rather than the “best practice” being promoted as the norm.

# **1 Introduction**

Francis and Holloway (2007) suggest that benchmarking appears to be an accepted and lasting management control tool in both private and public operations. In fact, benchmarking has received much attention in municipal operations, both in Scandinavia (Siverbo and Johansson 2007; Askim, Johnsen, and Christophersen 2007) and in other countries (Ammons and Rivenbark 2008; Bowerman, Francis, Ball, and Fry 2002; van Helden and Tillema 2005).

Benchmarking is strongly related to the idea of measuring and evaluating efficiency of services. Siverbo and Johansson (2007, 271) declare that the increased popularity of performance measurement is “one of the more lasting imprints that New Public Management has made in the public sector.” However, as Bowerman, Ball, and Francis (2001) state, benchmarking as a method could modernize local government and clarify contradictions in national and local needs; the understanding of the nature of benchmarking in the public sector, though, still is incomplete, and more studies are needed in order to find more types and purposes of benchmarking and its implications. Francis and Holloway (2007) identify a lack of critical analyses of the practice and implications of benchmarking and consider that both qualitative and quantitative studies over longer time spans are needed.

In Scandinavia, municipal benchmarking networks have gained increased attention (e.g., Askim, Johnsen, and Christophersen 2007). Saunders, Mann, and Smith (2007, 609) show that benchmarking within a networking environment is quite a recent phenomenon. In their study of one multi-organizational network, they found that benchmarking within a networking context had several positive impacts; participants’ systematic work in network meetings was important to

the results, but in order to see how practices actually were implemented, longitudinal studies were needed. In the same spirit, Ammons and Rivenbark (2008, 304) address the difference between the collection and reporting of measures and the actual use of them.

Askim, Johnsen, and Christophersen (2007) develop a framework for analyzing organizational learning outcomes in their evaluation of the Norwegian benchmarking project. They conclude that municipalities obtain organizational learning from benchmarking, but that this learning is dependent on how the networks have been formed and shaped and under what internal conditions participating municipalities have involved themselves; to what extent the level of learning also is related to actual performance is an issue that needs further research.

Since the late 1990s, the Swedish Association of Local Authorities and Regions (SALAR) has collected data in order to describe and measure quality in service. However, on the municipal level, individual municipalities have experienced a greater need and use for quality management and multi-dimensional control systems in order to describe municipal operations in new and better ways (Hellström, Jönsson, and Ramberg 2009). In the 1990s, cost measures dominated internal descriptions (Ramberg 1997), which further fueled the growing interest in systematic benchmarking of operating practices. In some cases, municipalities have organized local benchmarking initiatives (Tagesson 2002; Siverbo and Johansson 2007). In the mid 2000s, SALAR and the National Board of Health and Welfare initiated the National Benchmarking Project (NBP, National Board of Health and Welfare & SALAR 2005). In NBP, municipalities would form networks of five to ten municipalities that together would develop reports and compare performance and quality measures. Good examples were to be identified and

highlighted in order to inspire other municipalities to improve their operations. The project was run between September 2007 and March 2010 and involved 190 of Sweden's 290 municipalities.

The aim of the present study is to describe municipal benchmarking networks and to analyze whether they led to improvements in practice. The structure of the paper is as follows: section 2 presents the theoretical framework and section 3 the method and research design. In section 4, we present the empirical observations and offer concluding remarks in section 5.

## **2 Theoretical Framework**

Benchmarking should not be seen as isolated from other management control techniques in municipalities. If benchmarking is used, it needs to be integrated with both formal and informal control and become part of an articulated "control idea" in the overall management control system (e.g., Hellström and Ramberg 2009). A number of studies point out the importance of how the municipal control idea is formed and implemented and how it is adapted to changes in the municipality external environment (e.g., Brorström and Siverbo 2008; Knutsson et al. 2008). Changes made need to take into account not only actual work, but also how authority and responsibilities are organized and what objectives the organization targets (Knutsson et al. 2008). If you want to create improvement, it is important to describe what you want to accomplish. You also need to be able to monitor whether you succeed in realizing change and to adapt control in relation to actual changes made. This is expressed by Kaplan and Norton (2004, xiii) in the well-known saying, "You can't manage what you can't measure (and) you can't measure what you can't describe."

To make conscious the ambitions of an organization by articulating them verbally or in writing may seem self-evident. Yet, it is easier said than done. Goal congruence and compliance to objectives seem to be eternal themes in the fields of strategy, leadership, and management control. However, the need to articulate ambitions varies over time and with industry, organization, organizational level, and type of operation. Still, the mere fact that it is hard to articulate objectives should not let the issue be downplayed. To be able to show how you have performed is a central tenet in management control theory, as important as goal congruence and adaptation to change (e.g., Anthony and Govindarajan 2004). Information is core, for without it the consequent lack of knowledge will be a severe impediment to change. If this deficiency is paired with lack of incentives or resources, problems increase (Barney, 1996).

Benchmarking is one of many examples of management control techniques which need clear description, both of what should be accomplished and of goal fulfilment. The need for accountability is apparent (Bowerman, Francis, Ball, and Fry 2002). Other examples are quality improvement, management by objectives, and the balanced scorecard. Usually, these management control techniques are based on some form of underlying method, describing how to realize potential.

### **The Idea of Benchmarking and Its Development**

The term *benchmark* was originally used by land surveyors (Ammons 1999). If the position and the altitude of a particular landmark can be determined, this position can be used as a reference point (benchmark) for other landmarks. If we don't know where we are, we will have trouble in knowing where to go. Comparing one's own position and situation with others in this way is a

common human trait. The newborn's behavior mimicking the mother's sounds develops, through growth, into a more reflective, knowledge-based behavior, still in reference to behavioral norms and values proliferated by people in general. Besides, an organization following its main tendency does not have to spend resources explaining its choice, that is, there are no costs of social innovation (Collin et al. 2009). Thus, the need to relate to others in order to learn what to do is not a modern management idea, but an old, universal, necessary and omnipresent, yet implicit institution. However, one obvious risk with this mimetic behavior is that the "average" rather than the "best practice" is promoted as the norm (Llewellyn and Northcott 2005).

The explicit and systematic use of benchmarking in order to control organizational behavior can be traced back to a U.S. company, Rank Xerox (Francis and Holloway 2007). Top management had been unknowing of, and hence undisturbed, by a deteriorating performance. It took Rank Xerox close to seven years to initiate its "Leadership through Quality" program in 1983 (Walker 1992)—years filled with denial, self-scrutiny, analysis, dismissal, product revisions, rationalizations, customer reviews, and more. The program mainly consisted of three parts: (1) "competitive benchmarking," (2) workforce involvement, and (3) a full-fledged compliance to customer demands (Walker, 1992).

The late 1970s and early 1980s were a time when many U.S. companies realized the need for developing strategies and management control techniques as they faced competition from Japanese companies (e.g., Kaplan 1983; Francis and Holloway 2007). Systematic and step-wise benchmarking, as developed by Rank Xerox, is a response to that competition and has shaped the view of what benchmarking is and how it is done (Anand and Kodali 2008).

Anand and Kodali (2008, 258) compare 35 different benchmarking models and note that most of them share several key themes including measurement, comparison, identification of best practices, implementation, and improvement. These ingredients are fully captured in the model suggested by Camp (1993). According to Anand and Kodali (2008), Camp's definition of benchmarking, dating back to 1989, is one of the most quoted.

Anand and Kodali (2008), as well as Francis and Holloway (2007), state that there is no single prevailing definition of benchmarking. Fong, Cheng, and Ho (1998) classify benchmarking models according to (1) *the nature of referent other*, (2) *the content of benchmarking*, and (3) *the purpose for the relationship*. To separate public sector from private sector benchmarking, Bowerman et al. (2002) discuss benchmarking in terms of “compulsory” or “voluntary” adoption. Private companies mainly use benchmarking voluntarily in order to satisfy owners and their residual claims. Accountability and responsibility, however, is according to Bowerman et al. (2002), larger and wider in public organizations, which is why benchmarking to a greater extent is compulsory in this sector (e.g., Bernstein 2001).

However, Swedish municipalities by law and tradition are relatively autonomous from the central government, and the degree of formal accountability is generally low (e.g., Cassel 2000; Tagesson 2007; Tagesson and Eriksson forthcoming). Thus, it is probably rather an institutional pressure and not compulsory or legislated demands that explains Swedish municipalities' interest in benchmarking. Municipal politicians, who act in a political market, have to consider voters and other stakeholders in their ambition to be re-elected (Downs 1957; Copley et al. 1995). Consequently, municipalities have to react to media pressure with regard to per capita cross-

comparisons between municipalities and municipal services (e.g., Ammons 1999; Tagesson 2002). Of course, the institutional pressure on individual municipalities increased even further when SALAR and the National Board of Health and Welfare initiated the NBP.

Thus, proactive benchmarking and participation in the NBP may be one way to deal with institutional pressure from the media and external stakeholders. However, participation in the NBP does not necessarily mean that the practice is integrated into the municipality's managerial and operational processes (e.g., Dillard, Rigsby, and Goodman 2004; Ammons and Rivenbark 2008). Benchmarking may be seen as “intertwined with institutional political processes and the operation of other forms of organizational and calculative practice” (Burchell et al. 1980, 13).

However, by decoupling, the municipalities can gain legitimacy by participating in the NBP while their activities vary in response to practical considerations (Meyer and Rowan 1983). Depending on situation and possible consequences, organizations and their agents may respond to institutional pressure in a variety of ways—ranging from passive conformity, compromise, and avoidance to defiance and proactive manipulation (Oliver 1991). The existence of institutional response patterns is also confirmed by van Helden and Tillema (2005) in their study of benchmarking in wastewater treatment by Dutch water boards.

The design and use of measures is expected to improve decision making. Better measures contain more information and stimulate analysis and reflection (Ammons and Rivenbark 2008). Relative measures such as productivity are considered better than absolute measures such as simple frequency. Bowerman et al. (2002) claim that benchmarking is useful for accountability

purposes and that the system in which it is used, be it quality control or balanced scorecard, is not important.

Simon et al. (1954) show that measures have three important functions. The first is to measure current status (*scorekeeping*). The second is to define and clarify what is of major importance to the company (*attention directing*), and the third function is to assist the organization in solving its problems (*problem solving*) through adequate resource allocation. Management control is a matter of influencing the organization and its members' actions by providing facts, priorities, and practical approaches to the work. The core of using measurements is an ambition to improve (Behn, 2003). According to Moynihan and Landuyt (2009), this ambition is closely related to organizational learning, which in the benchmarking literature in turn is strongly connected to the idea of "best practice" and process comparisons (Francis and Holloway 2007; Anand and Kodali 2008). The analysis of a measurement value should be able to signal a need for change and stimulate the organization to search for knowledge from those who perform better.

### **Benchmarking and Organizational Learning**

According to Askim et al. (2007), benchmarking rests on the assumption that it supports organizational learning and innovation. Unfortunately, scholars have created a jungle of overlapping and blurred concepts and theories around the theme of organizational learning (Huysman 2000). Visser (2007), though, suggests a clear-cut distinction between three levels of organizational learning: deuterio-learning, meta-learning and planned learning. The first two levels relate mainly to individuals and groups, whereas the third level concerns the organization.

When it comes to meta-learning, Visser puts his faith in the theories of Argyris and Schön (e.g., Argyris 2004) and their concept of “single-loop” and “double-loop” learning.” Double-loop learning occurs in the detection and correction of errors/non-conformances that require changes in the action strategies and values that govern the theory-in-use for these action strategies. Single-loop learning is the detection and correction of errors that do not require changes in the theory-in-use that governs the action. Learning through correcting errors or learning by inquiring whether you are actually doing the right thing are two fundamentally different approaches to learning, according to Argyris. However, Visser claims that both approaches are types of meta-learning, that is, learning through reflection about context. Meta-learning is characterized by periodicity and that it is a cognitive and conscious action. It may enable control and organization, improving both individual and organizational performance.

In order to institutionalize meta-learning on an organizational level, norms and values have to be established to support individual and group behavior (Visser 2007). This is so-called planned learning. It should permeate both actions by co-workers and management and supporting organizational structures. Structure could be, for example, a balanced scorecard or a benchmarking model. Continuous meta-learning—a planned learning—is, however, not easy to accomplish (Moynihan and Landuyt 2009) but could be a way to approach organizational learning in a municipality context, both in theory and in practice.

Benchmarking as an idea, but also as a method and technique, can carry many meanings and therefore may be difficult to study. Still, some common traits can be found, originating from the method developed and used by Rank Xerox (Anand and Kodali 2008). Behn (2003) underscores

the relationship between measurements, management control systems, and the drive for improvement.

To capture improvements from a municipal benchmarking network project calls for a number of different research questions to be articulated. One question concerns the quality of the goods and services provided by the municipality. Have there been any changes in the content provided? Another battery of questions addresses whether there have been any changes in structure or form, e.g., organization, work methods, or routines, used by the municipality in their provision of goods and services. There is also the question of cost savings without reducing quality of goods and services. The relation between measurement and management control is also important to take into consideration. To what extent have measures and measurement techniques from the networks helped in developing the existing management control system of the municipality? And can this development of new measures be related to institutional pressure?

### **3 Research Design and Methods**

To achieve the aim of this study as defined in section 1, it was necessary to use a multiple research design, specifically using four main data collection methods: survey, focus group, document study, and interview. As benchmarking has a strong connection to organizational learning, it was important to capture not only behavioral changes but also changes in the understanding of benchmarking itself. Hence, techniques to capture understanding outside of those directly affected by the benchmarking networks were necessary, reaching groups such as politicians and centrally- and locally-based employees.

To capture the changes over time in the various benchmarking networks, the data collection was divided into three periods: 2007–2008, 2008–2009, and 2009–2010. During the first year (2007–2008), the networks decided what municipal services to compare and how to measure them. After these decisions, each network made the first measurements and wrote an initial network report. From the data collection perspective, year one was mainly focused on studying these reports. A total of 100 reports were produced in the three-year period of the NBP. The data from these reports were recorded according to three measurement categories: *quality*, *cost*, and *efficiency*.

During the next two periods (2008–2009 and 2009–2010), two surveys were distributed. In the spring of 2009, the survey aimed at capturing *initial signs* of progress. A total of 169 municipal project leaders were sent the survey, with a response rate of 75%. The main question was about how respondents estimated impacts from the project on changes in service content, form of service, or costs. In early 2010, a second survey was sent to 944 municipal directors, finance directors, and project leaders. The purpose was to find out whether participation in the NBP had made some *lasting imprint* in the respondent's own organization and its way of planning and controlling work. The response rate was 51%.

In addition to investigating lasting organizational imprints for the third time period (2009–2010) it was also logical to consider some *changes in professional and collective values and approaches* related to benchmarking. The interviews and focus groups were therefore conducted during this final period. Twelve different municipalities were visited in the fall of 2009. The selection covered both old and young networks as well as municipalities with varied conditions.

The visitations comprised interviews with active politicians (primarily chair of the municipal board), project leaders, finance directors, and one or two incumbent managers of elderly or family care or schools. In this way, a broad view of the benchmarking work was gained. In all, 64 interviews were conducted.

The activity selected for discussion in the three focus groups was elderly care. This municipal service was the first in the NBP and the network concept had reached a wide level of diffusion early on in the project. To see the impact that comparisons and benchmarking might have, a ten-year-old network acted as one focus group; its purpose was to capture changes, understandings, and attitudes towards benchmarking within the network. A second focus group session was also arranged, this time with elderly care department managers, its purpose being to obtain the professional management perspective on benchmarking and change. The third focus group session was convened with first-level managers, in an attempt to capture explicit developments in local operations.

## **4 The National Benchmarking Project**

In this section, empirical observations are described and analyzed. They follow the structure of Camp's (1993) four phases of a benchmarking process: planning, analysis, integration, and action.

### **Phase 1 – Planning**

In the planning phase, three steps are taken: (i) identify what is to be benchmarked, (ii) identify comparable organizations, and (iii) determine data collection method and collect data. These three steps are well reflected in the way NBP was organized. The project was organized through networks, where three to five process leaders from SALAR supported networks by providing competencies and structure. The networks' composition mainly depended on factors such as size, location, and individual desire to cooperate. Most networks were started within the NBP project; some, though, were established before the start of the NBP, while others originated from other cooperative constellations and links.

***Identifying what is to be benchmarked.*** In each network, a steering committee was formed comprising leading politicians, municipal directors, and each municipality's project leader. The committee decided which service or part of service to focus on. It met periodically, and in between meetings a group of project leaders and a SALAR process leader worked to define, describe, and assure quality data, as well as record and analyze the data within the network. Twice a year, gatherings of network project leaders took place where about 70 project leaders were given the opportunity to discuss their respective areas in smaller groups. According to process leaders, these gatherings provided explicit learning opportunities for project leaders, in turn leading to further quality assurance of measures used.

The objective was for every network to cover at least two services, or parts of services, during the time period. Each covered service was documented in a report, first checked and approved by the steering committee, then sent to each respective municipality board. The result of the

benchmarking work was in this way handed over to the municipality along with the responsibility for further analysis and potential action.

***Identifying comparable organizations.*** In the first year of NPB, 20 networks were initiated, comprising altogether 46 separate municipalities. The number of network participants varied between three and eleven municipalities, with a mean and median of about seven. This variation did not change over the life of the project. A second invitation to join the project was given in spring 2008, and an additional 26 municipalities joined in the fall of that year. In the second and third year, participation increased gradually, with a peak of 190 participating municipalities in the third year, organized in 28 different networks. According to the process leaders, only three municipalities decided to leave the project, generally because they were not involved with the service selected for benchmarking.

The planning phase of NBP involved many leading decision makers in the municipalities. The exact number was not determined, but a cautious estimate of 25 networks of seven municipalities, each covering three services (affecting seven people per municipality) would yield 1,225 people directly affected by the benchmarking project.

***Determining data collection method and collecting data.*** Networks clearly focused on increasing their knowledge about content (i.e., quality) of goods and services. The top five services were elderly care, social welfare, disability care, elementary school, and preschool (see table 1). These services share a lack of firm quality measures. Network project groups have worked intensely with the development of new data collection methods, be it standardized

surveys, the formal use of external reviewers, or making new and better use of existing national data.

-----  
Insert Table 1 about here.  
-----

It is apparent how similar the NBP in this respect is to the Norwegian national benchmarking project (Askim, Johnsen, and Christophersen, 2007). The NBP process leaders talk about this period as “the shop floor period”, meaning the period when the benchmarking activities are undertaken. It peaks in the annual project leader gatherings, but in the day-to-day process, the shop floor attitude is most apparent at the beginning of each new service chosen by the network.

Networks that in 2009 decided to engage in benchmarking for the top five services listed in table 1 recorded a significant number of explicit experiences using measures and methods. Moreover, some of the measures developed by the NBP networks in the service of elderly care were the first to gain status as national quality indicators. It was also decided that three other services—preschool, social welfare, and disability care—would attain the same status in the fall of 2010.

In all 100 network reports, 1,372 records were made. In the network reports from the first year, about 80% of the measures were about quality, and 10% were each related to cost and efficiency measures. This distribution was stable over consecutive years, but some additional measures were used in later reports. About 25% of these measures were of a quality index type.

Quality measures dominate the reports, many of them having been developed in services where such measures were previously lacking. About 14% of the measures were of a baseline character,

either describing structural conditions, such as households with financial support from the municipality, or describing volume, such as number of citizen suggestions or number of initiated investigations. Structural measures show initial conditions, whereas volume measures show quantitative outputs not expressed in terms of quality, cost, or efficiency. Both types are needed to facilitate learning based on understanding.

Networks focused on quality measures, while other descriptions of operations, such as process descriptions were left out. According to the study frame of reference, process descriptions are vital to completing the full benchmarking process. Although there are indications of more activity descriptions in the national quality measures, this aspect of benchmarking could be further developed.

***Findings from the planning phase.*** The NBP reflects well the three steps of the planning phase. The municipal impact is large, in terms of participating municipalities and operational services, as well as the number of participating and affected persons. There is richness in the data collection methods developed, and there has been systematic quality assurance work on the collected data. Structural capital for further benchmarking work has been built up in the planning phase. Beyond their apparent time-saving consequences, however, there is a risk of letting predetermined measures and methods reduce a healthy and critical mindset of participants. It is in the process of learning how to define and capture performance, where cause and effect is scrutinized and discussed, that the potential benefits reside.

## **Phase 2 – Analysis**

The analysis phase comprises three steps: (i) determine current performance levels, (ii) project future performance levels, and (iii) communicate benchmark findings and gain acceptance.

Numerous empirical accounts illustrate the substance of this phase; they include case interviews, process leader interviews, surveys, and reports from networks doing a second follow-up study.

The NBP approach made the analysis a natural part of the project group meetings as well as meetings in the individual municipalities once the reports were published. Descriptions were compared in the meetings as reported in the documentation.

As one report was published, the network steering committee decided on the next service to study. When the project group began this service, process leaders asked for feedback from the municipalities, particularly for actions observed in the first selected service. The twelve case studies indicate that differences among municipality performances were valuable for the people involved in the project. This has motivated participants to think through operations and their priorities. One project leader said, *“If they all performed the same, it wouldn’t be interesting. It’s the differences we can learn from.”* When the results were presented, attention was mainly directed to how one organization performed over the others. From the 64 case interviews, it was clear that participants had to be able to explain their own municipality’s relative performance to their home organization.

The project leaders were generally enthusiastic about the NBP. However, among the representatives from the various selected services, the comments were more fragmented. One person representing social welfare stated, *“When I got the work in my lap it wasn’t fun, but since*

*then it really has been interesting to meet with others and discuss why things work like this or like that.*” Another social welfare representative summarized the project: *“I’ve really looked forward to the cooperation with experienced colleagues, I even came back early from my holiday to participate.”* However, other representatives were more skeptical. One social welfare manager remarked, *“We are tired ... What will all this lead to? The conditions from one municipality to another are so different that it is useless to make comparisons.”* Skeptical views of this kind were made by a number of other respondents, and not just from municipalities that were ranked low in the comparisons. One respondent from a relatively well-performing municipality drily stated that *“we have given more than we have received; we should really have been in another network with other member municipalities in order to learn ourselves.”* Several other respondents from other municipalities pointed to this situation in the networks. The projects were often decoupled from political decision making. Several respondents mentioned that findings in the NBP project had not been disseminated within the organization, that reports had only been passively passed on to politicians, who had not been interested in the findings or potential implications. Some expressions of doubt about the ability to act on findings were also made by respondents. It was also mentioned that the strong preoccupation with numbers was disappointing, whereas practical aspects of operations were expected to be discussed more than what was the case.

In the project groups, according to the process leaders, there were some lively discussions about performance levels and how performance could be measured and used for comparisons. From the case interviews, some conscious trade-offs between cost and qualities were revealed. However, the focus was on identifying activities that stood out in terms of cost or quality, rather than

identifying “best practice.” In municipalities where activities showed relatively high cost levels in the comparisons, this provided motivation for cost cutting. Similarly, a high relative quality ranking was seen as a reason to reduce quality (and hence, costs): *“It is possible to prioritize from the reports,”* as one representative from elderly care stated.

Top managers, project leaders, and leading representatives from the different services were actively engaged in the analysis work: *“We have begun comparing ourselves with others, and each time we do it, it becomes less threatening and becomes a natural part of work in the organization.”* Systematic comparisons are still mainly done at top management level and seem to have gained less attention locally. From the first (“initial signs”) survey, it is evident that the planning and analysis phases of the NBP have had some impact on operations. The most apparent observation is that the focus has not been on cost rationalization. The second observation is that some participating municipalities have changed the type of services offered and the way planning and control is done.

From the second (“lasting imprint”) survey, responses suggest that the NBP participation is expected to have some (49%) or large (33%) impact on how the municipality uses systematic comparisons. The NBP participation is expected to have some (31%) or large (52%) impact on the longevity of comparison work. Overall, NBP participation is seen as largely (72%) affecting future municipality development in a positive way.

Only three networks produced a second report for a particular service (in total, seven reports for five services), and these reports indicated how benchmarking methods were developed and

measures refined between reports. However, there was no clear trace of any attempt to analyze actual changes in performance.

*Findings from the analysis phase.* On one hand, the networks worked well in describing and analyzing the measured performance, as is clearly documented in the numerous reports. These reports also seemed to positively affect the norms and values of benchmarking. On the other hand, the municipalities' ability and actual effort in transferring experience from the reports into practical action at home varied greatly and in some cases the outcomes have been weak. Further, the consequence of the networks' focus on developing quality measures has been that the development of other measures, such as process measures, has to a great extent been neglected.

### **Phase 3 – Integration**

Phase 3 comprises two steps: (i) establish functional goals and (ii) develop action plans. One input to describe Phase 3 is data from the second survey. The purpose of the control survey was to investigate whether the NBP made any lasting impressions on how planning and control is done in the municipality. About 40% of respondents were of the opinion that the NBP had a large or very large impact on how the municipality worked with systematic comparisons. Concerning how this work would last, close to 60% anticipated the work to be permanent. Most interesting is the fact that 80% found the work of large importance to future systematic comparisons. These figures suggest that systematic comparisons will prevail and that people do understand the usefulness of them.

A leading politician stated that it takes a long time to change an existing political discussion, traditionally focused on the distribution of resources and not on the results of the distribution. Another politician described how the integration phase builds on learning: *“The NBP has clearly shown how important it is to enter things into the computer; statistics improve, so we have put much effort into routines and training, but it is obviously an issue of generations.”* A finance director pointedly commented that *“the NBP has gotten us much further in two years than our own benchmarking has done in many years.”*

*“Our board got the report, but took no actions or asked no questions. Nothing will come from it.”*

This quote from a project leader suggests that this way of working may be somewhat uncomfortable to politicians, upsetting normal routines and ways to discuss priorities. Still, improved descriptions of performance should allow improved resource allocations at the political level. At the operational level, examples of explicit use of performance measures show how it is quite manageable to integrate new measures into the day-to-day discussions in the workplace: *It concerns our everyday work ... It is so much easier to commit yourself to the questions when it is about your own work,*” as one respondent said. The process leaders also gave examples of how some municipalities plan to use the NBP measures in their budgets and follow-up routines, a definitive sign of integration of project ideas into the existing organization.

***Findings from the integration phase.*** The second survey indicates that internal planning, follow-up, and external reporting have improved as a result of the NBP. This way of integrating and institutionalizing project results into daily routines is an indication of lasting effects from the project. The case studies show that knowledge about what systematic comparisons mean and

may lead to has improved, and some participants find it reasonable and useful to work in that way. Overall, the climate for discussing and comparing performance has also improved.

Measures of quality are the most frequent measures in the NBP reports. As the majority of operational services have been in need of quality measures and descriptions, this result is imperative. Still, there is a flip side to this coin. While the quality focus has left process measures without attention, the creation of national quality indicators in preschool, disability care, social welfare, and elderly care is of merit. All in all, there are indications that some participating municipalities are starting to integrate measures and methods developed in the NBP into existing management control systems.

#### **Phase 4 – Action**

Phase 4 comprises two important comprehensive steps in the benchmarking process: (i) implement specific actions and monitor progress and (ii) recalibrate benchmarks. It is difficult to observe any explicit changes in service content from the network reports. Only a few networks managed to produce a second report for a service, and in those follow-up reports changes were only implied. The information on action emanates mainly from the process leader interviews, case studies, and focus group interviews.

Process leaders mentioned actual changes of various kinds. One example is the balance between home care and institutional care within elderly care. This situation concerns both changes in supply and changes in form. In some cases, home care has been developed whereas forms of service in between home and institutional care have been removed; in other cases, intermediate

forms have been developed. Process leaders claimed that the quality index of elderly care has had a large impact on how many municipalities allocate resources, for example, in terms of choices of food, frequency of cleaning, and time spent outdoors. Similar actions were observed in disability care.

Measuring and benchmarking the continuity of staff meeting with elders and how personnel divide their time between direct and indirect tasks are other examples of how quality has been approached. The quality of information on web pages, printed information, as well its accessibility have all improved dramatically due to the measurements conducted as part of the NBP. This is supported in the case interviews.

When it comes to changes in the form of municipal provision of goods and services, the process leaders related these changes to the methods used to describe quality. Various index measures and so-called typical cases have supported the changes in the way goods and services are provided, for example, how businesses and entrepreneurs are approached and how accessibility for disabled persons is implemented. Handling time for administrative matters, such as decisions about social welfare or building permissions are another examples where execution of work has been improved by the NBP. In the case interviews, examples stand out of an improved balance of the forms of elderly care, improved staffing, and work scheduling. Policies for welfare grants and a transfer of social welfare from closed to open forms of care are other examples.

Cost rationalizations have been scarce, and only some examples were recorded. Cost comparisons in one case caused one municipality to raise rents and fees for residents, meanwhile

focusing more closely on actual service needs. This rendered a dramatic decrease of the net cost of an individual for the municipality. In another case, coordinated food distribution for school and elderly care reduced costs as well. The most apparent example of cost rationalization, though, was how technology transfer between two different municipalities made it possible to record the work done by external suppliers; thereby, remuneration was directly connected to actual performance and not based on standard estimates.

The focus groups showed how measurements actually played a very modest role in the search for improvement. It was, first and foremost, the experiences from their own operations that guided improvement efforts. The extent to which measures were used hinged mainly upon the respective measure's ability to credibly represent actual performance. This was strongly supported by case study observations.

These examples were still the exception; the process leaders confirmed that cost reductions were few in number. Two reasons emerge in the empirical material: cost reductions could not be observed since a second measurement had not been made, but the most important factor was a steadfast distrust of existing cost data.

For the politicians, the results of NBP to a greater extent were about showing or confirming the quality of existing service, rather than looking for potential improvement. One politician said that NBP was all about *“getting a confirmation that we have good quality and low cost in our social welfare services.”* Considering the results from the network benchmark, another politician said, *“We can now lower our ambitions in a more informed way.”*

*Findings from the action phase.* First, changes in content were observed within the benchmarking networks. Quality changes were related to the methods developed in the NBP for describing quality. Elderly care, disability care, and social welfare stand out in this respect. In these services, performance is to a large extent a matter of quality, but it had not been possible to describe and measure it before. This may be why these services were early starters in the NBP, which in turn gave the networks time to develop adequate measures and methods to be captured in this study. Instances of well-performing municipalities reducing their ambitions indicate the reduction of disparities in performance rather than their continuous improvement. One exception is the website information where participants generally improved. It is worth noting that this kind of change has little direct implications for the employees.

Second, changes in form of goods and service provision can be related to methods and measurements developed in the NBP. Improved staff scheduling, improved staff continuity among the elderly, and social activities for the elderly are examples of how actual work changed in participating municipalities. New methods have been developed, where the typical case method and technology-driven reporting are two clear examples.

Third, to identify any cost rationalization during the three-year period is not so easy. The changes that have been made concerning content and form of service, even though not explicitly cost-oriented, may well prove to be driving costs in the long run. For example, a decrease in cleaning frequency will likely free resources, just as the improved offering of a different variety of meals will claim additional resources. The most apparent cost rationalization was the

technology transfer aimed at recording the work done by external suppliers. Beyond the obvious reason that quality measurements were the main objective of the project, the problems in estimating fixed and variable costs as well as their variation over time, paired with a lack of trust of established cost data, are the main reasons why cost rationalizations were not implemented.

To sum up, the NBP has produced some changes in behavior, both in terms of content and form. No major cost rationalizations have been identified, but some reductions are implied. These are to be found in the networks that started early, which suggests that time matters. However, what has been found is a clear contribution of revealing important differences in ambition across the participating municipalities. These differences have presented clear potential for improvement.

## **5 Conclusion and Discussion**

This section presents our conclusions on learning, organizational support, and the issue of leveling vs. best practice.

### **Learning**

Knowledge and understanding about benchmarking has been established. The NBP has helped diffuse knowledge about benchmarking in 190 municipalities. There are also many examples of how participants have made practical use of that knowledge. This outcome suggests that participants have gained an understanding about benchmarking and the relevance of systematic comparisons over time. The conclusion is further emphasized by the fact that NBP measures and methods have been transformed into a system of national quality indicators.

Swedish municipalities understand the point of systematic comparisons. Many of them have a positive attitude to benchmarking and the indications are that benchmarking is here to stay. What does this mean for the Swedish municipal sector? Incentives and capacity sometimes are lacking in municipalities, which is why any immediate benchmarking revolution is not likely to happen. Rather, it is more adequate to think of an evolving movement. Comprehensive approaches are pending that could go beyond comparing individual performance measurements, process descriptions, and explicit action plans. We have observed indications of meta-learning and in some cases an evolving understanding of the need for institutionalization of benchmarking practices—that is, planned learning (Visser 2007). Some 75% of participating municipalities have committed themselves to continued activity within the SALAR benchmarking network structure. This corresponds to almost 50% of the 290 Swedish municipalities.

A few instances of process indexation have been observed, and the ways measures are developed are gradually being re-examined. This takes time. Concurrently, internal capacity has to be improved. Obviously, here resides a great potential for a more elaborate and efficient benchmarking model, but the time it takes is probably also time needed; in order to accept and implement new ways of working, knowledge, and understanding likely need first-hand experience.

### **Organizational Support**

Differences in benchmarking attitudes have an effect on implementation. Experiences and results from measurements have been used with varying zeal and intent in different municipalities. Reports seem to have been put to more use where politicians have shown greater interest in

them. Politicians also seem to be more interested in showing good performance rather than taking measurements as a starting point for development and continuous improvement of operations. Thus, support from top management is crucial to the constructive use of benchmarking information (e.g., Askim, Johnsen, and Christophersen 2007). A critical aspect, though, is the necessary goal congruence between political, managerial, and operational levels. Without congruence, the willingness to alter behavior is reduced. In most of the cases, participation in the benchmarking project has not been widely known or acknowledged, while the main activities have been a matter of data gathering.

There are also indications that smaller municipalities are more involved than larger ones. Other signs point toward inferior performance being the most important driving force for change. However, systematic benchmarking may create a norm for performance, thus leveling municipal performance rather than improving it.

### **Leveling, Not Best Practice**

Variation in action and performance makes it possible to classify municipalities in a matrix (see figure 1). Action denotes the effort made by municipalities to change behavior. Performance refers to how municipality performance is shown in NBP performance measurements.

-----

Insert Figure 1 about here.

-----

In the first category, low action and low performance, municipalities have refrained from benchmarking information. It is rooted in inferior relative performance. Denial takes the form of

questioning the measurements and they attribute low performance to exogenous factors or methodological flaws. The second category involves municipalities which act on the performance measurements on the basis of identified potential for improvement. The signifying attitude of these municipalities is effort in order to improve. These are municipalities that are closest to embracing the ideas behind benchmarking, even though they not always aim to achieve best practice. The third category comprises the municipalities which perform well and are content. They use the information from their benchmarking experiences for confirmation and recognition of their own excellence. These municipalities show an attitude of complacency. The fourth and final category comprises municipalities with the characteristics of high performance and high action. However, their action aims to adapt to prevailing norms. Hence, the action by municipalities in category two and four leads to a leveling of municipal performance rather than improvement (e.g., Llewellyn and Northcott 2005).

The fact that the NBP was initiated by SALAR and the National Board of Health and Welfare may have created institutional pressure and a sense of political control (Burchell et al. 1980), thus leading to a defensive benchmarking strategy with the aim for the municipalities to prove to be “good enough” (or not the worst) rather than striving to be the best (Bowerman et al. 2002).

## **References**

- Ammons, David N. 1999. A Proper Mentality for Benchmarking. *Public Administration Review* 59(2): 105–109.
- Ammons, David N., and William C. Rivenbark. 2008. Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina

- Benchmarking Project. *Public Administration Review* 68(2): 304–318.
- Anand, Gurumurthy and Rambabu Kodali. 2008. Benchmarking the Benchmarking Models. *Benchmarking: An International Journal* 15(3): 257–291.
- Anthony, Robert N. and Vijay Govindarajan. 2004. *Management Control Systems*. 11th ed. Chicago: Irwin.
- Argyris, Chris. (2004). *Reasons and Rationalizations: The Limits to Organizational Knowledge*. Oxford: Oxford University Press.
- Askim, Jostein., Åge Johnsen, and Knut-Andreas Christophersen. 2007. Factors behind Organizational Learning from Benchmarking: Experiences from Norwegian Municipal Benchmarking Networks. *Journal of Public Administration Research and Theory* 18(2): 297–320.
- Barney, Jay B. 1996. *Bringing Managers Back In: A Resource-Based Analysis of the Role of Managers in Creating and Sustaining Competitive Advantages for Firms*. Texas: A&M University.
- Behn, Robert D. 2003. Why Measure Performance? Different Purposes Require Different Measures. *Administration Review* 63(5): 586–606.
- Bernstein, David J. 2001. Local Government Measurement Use to Focus on Performance and Results. *Evaluation and Program Planning* 24: 91–101.
- Bowerman, Mary, Graham Francis, Amanda Ball, and Jackie Fry. 2002. The Evolution of Benchmarking in UK Local Authorities. *Benchmarking: An International Journal* 9(5): 429–449.
- Brorström, Björn and Sven Siverbo. 2008. *Perspektiv på framgångsrika kommuner –utmaningar i teori och praktik*. Göteborg: Kommunforskning i Västsverige, KFi-rapport nr 94.

- Burchell, Stuart, Colin Clubb, Anthony Hopwood, and John Hughes. 1980. The Roles of Accounting in Organizations and Society. *Accounting, Organizations and Society* 5(1): 5–27.
- Cassel, Filip. 2000. *Behovet av kommunal externrevision*. Stockholm: SNS.
- Camp, Robert C. 1993. A Bible for Benchmarking, by Xerox. *Financial Executive*, July/August, 23–27.
- Collin, Sven-Olof, Torbjörn Tagesson, Anette Andersson, Joosefin Cato, and Karin Hansson. 2009. Explaining the Choice of Accounting Standards in Municipal Corporations. *Critical Perspectives on Accounting* 20(2): 141–174
- Copley Paul A., Jennifer J. Gaver, and Kenneth M. Gaver. 1995. Simultaneous Estimation of the Supply and Demand of Differentiated Audits: Evidence from the Municipal Audit Market. *Journal of Accounting Research* 33(1): 137–155.
- Dillard, Jesse F., John T. Rigsby, and Carrie Goodman. 2004. The Making and Remaking of Organization Context: Duality and the Institutionalization Process. *Accounting, Auditing & Accountability Journal* 17(4):506–542.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper Row.
- Fernandez, S. and Rainey, H.G. 2006. Managing Successful Organizational Change in the Public Sector. *Public Administration Review* 66(2), 168–176.
- Fong, Sik Wha, Eddie W.L. Cheng, and Danny C.K. Ho. 1998. Benchmarking: A General Reading for Management Practitioners. *Management Decision*, 36(6): 407–418.
- Francis, Graham and Jacky Holloway. 2007. What Have We Learned? Themes from the Literature on Best Practice Benchmarking. *International Journal of Management Reviews* 9(3): 171–189.

- Hellström, Mikael and Ulf Ramberg. 2009. Styra meningsfullt med mål. In *Ett smörgåsbord med ekonomistyrning och redovisning – en vänbok till Olof Arwidi*, edited by Per-Magnus Andersson, Peter Jönsson, Gert Paulsson, and Stefan Yard, Institute of Economic Research. Lund: Lund Business Press.
- Hellström, Mikael, Peter Jönsson, and Ulf Ramberg. 2009. Diffusion of Administrative Techniques: The Case of Balanced Scorecard in Swedish Municipalities. Paper presented at MPSA, Chicago, April.
- Hellström, Mikael, Hans Knutsson, and Ulf Ramberg. 2009. *Strategier för kommunal hushållning*. Stockholm: Rådet för främjande av kommunala analyser.
- Huysman, Marleen. 2000. Rethinking Organizational Learning: Analyzing Learning Processes of Information System Designers. *Accounting, Management and Information Technologies* 10(2):81–99.
- Kaplan, Robert S. (1983). Measuring Manufacturing Performance: A New Challenge for Managerial Accounting Research. *Accounting Review* 58(4): 686–705.
- Kaplan, Robert S. and David P. Norton. 2004. *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Boston: Harvard Business School.
- Llewellyn, Sue and Deryl Northcott. 2005. The Average Hospital. *Accounting, Organizations and Society*, 30(6):555–583.
- Meyer, John W. and Brian Rowan. 1983. Institutionalized Organizations: Formal Structure as Myth and Ceremony, *American Journal of Sociology* 83(2): 340–363.
- Moynihan, Donald P. and Noel Landuyt. 2009. How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives. *Public Administration Review* 69(6): 1097–1105.

- Oliver, Christine. 1991 Strategic Response to Institutional Processes. *Academy of Management Review* 16(1): 145–179.
- Ramberg, Ulf. 1997. *Utformning och användning av verksamhetsmått*. Lund: Lund University Press.
- Simon, Herbert. A., H. Guetzkow, G. Kozmetsky, and G. Tyndall, G. 1954. *Centralization and Decentralization in Organizing the Controller's Department*. New York: The Controllershship Foundation.
- Siverbo, Sven and Tobias Johansson. 2006. Relative Performance Evaluation in Swedish Local Government. *Financial Accountability & Management* 22(3):271–290.
- Socialstyrelsen and Sveriges kommuner och landsting (2005). *Nationell strategi för "öppna jämförelser" av hälso- och sjukvårdens effektivitet och kvalitet – Förslag till aktiviteter och ställningstaganden*. Stockholm: Socialstyrelsen.
- Tagesson, Torbjörn. 2002. *Kostnadsredovisning som underlag för benchmarking*. Lund: Lund University Press.
- Tagesson, Torbjörn. 2007. Does Legislation or Form of Association Influence the Harmonization of Accounting? A Study of the Swedish Water and Sewage Sector. *Utilities Policy* 15(4): 248–260.
- Tagesson, Torbjörn and Ola Eriksson. forthcoming. What Do Auditors Do? Obviously They Do Not Scrutinize the Accounting and Reporting. *Financial Accountability and Management*.
- van Helden, G. Jan and Sandra Tillema. 2005. In Search of a Benchmarking Theory for the Public Sector. *Financial Accountability & Management* 21(3): 337–361.
- Vedung, Evert. 1998. *Utvärdering i politik and förvaltning*. Lund: Studentlitteratur.

Visser, Max. 2007. Deutero-learning in Organizations: A Review and a Reformulation. *Academy of Management Review* 32(2): 659–667.

Walker, Rob. 1992. Rank Xerox – Management Revolution. *Long Range Planning* 25(1): 9–21.

**Table 1.** Number of networks, number of municipalities and number of services that performed double measurements, all reports produced between September 2007 and March 2010.

<b>Service/Part of service</b>	<b>Number of networks</b>	<b>Number of municipalities</b>	<b>Double measurements</b>
Employment issues	3	22	
Preschool	8	56	1
Roads and lighting	4	30	
Elementary school	11	78	1
Secondary school	2	12	
Disability care	12	87	1
Social welfare	14	100	2
Information	2	18	
Food	1	6	
Culture and sport	4	29	
Business	2	17	
Planning and building	5	30	
Elderly care	23	148	2
Others	2	13	

<b>Performance Action</b>	<b>Low</b>	<b>High</b>
<b>Low</b>	<i>1. Denial</i>	<i>3. Complacency</i>
<b>High</b>	<i>2. Improvers</i>	<i>4. Leveling (reduction in level of ambition)</i>

**Figure 1.** Understanding municipal attitudes towards benchmarking.