KRISTIANSTAD
UNIVERSITY COLLEGE

Department of Business Studies

Working Paper Series
ISSN:1650-0636

Working Paper Series 2003:3

Do comparisons of performance measures give misleading information? The case of the Swedish water and sewage sector

TORBJÖRN TAGESSON

ABSTRACT

In recent years interest in benchmarking and comparisons of performance measures has increased among Swedish municipalities. Comparisons of performance measures are used for different purposes. They are used in benchmarking processes, with the aim not only of improving operations but also of providing standards in the context of accountability. Regardless of the purpose of the comparisons of performance measures, it is important that the compared measures are defined and applied homogeneously and consistently, i.e. they must have information correspondence in order to be comparable; otherwise, erroneous decisions can be made. However, commensurability is not enough. If the compared measures are to have any information value, the measures also need to be chosen carefully, so that, in a relevant and reliable way, they reflect the measured objects' most important qualities. This paper analyses performance measures used in three benchmarking projects in the Swedish water and sewage sector. The measures are analysed in the light of how well they reflect the measured objects' most important qualities, their relevance, commensurability, and reliability. The findings are distressing; the appropriateness of the performance measures in these three projects must be strongly questioned. In particular, the measures were too aggregated as to capture, reproduce, and describe the most important qualities of the measured objects in a relevant way. As for commensurability, it was so poor that one is entitled to ask whether the comparisons are not doing more harm than good.

Key words: accountability; benchmarking; comparisons; commensurability; performance measures; reliability

Department of Business Studies	Phone: +46 44-20 31 06
Kristianstad University College	Fax: +46 46 44 20 31 03
SE-291 88 KRISTIANSTAD	www.hkr.se/web-eko/sve/forsk/index.htm
SWEDEN	e-mail: e@hkr.se

1. Different purposes for comparison of performance measures

In a monopoly business where the price mechanism of a free market does not exist, there is a risk that efficiencies may be lower and prices higher than in a competitive business. To avoid a situation where enterprises with a monopoly position could take advantage of their monopoly power, governments and municipalities in Sweden run many operations. In a geographically limited monopoly, such as a water and sewage enterprise, it is possible to make comparisons with similar enterprises located elsewhere. By means of such comparisons—which may be termed *comparative competition* (Stahre, Adamsson and Eriksson, 2000)—one can form an idea of whether an enterprise, relative to other enterprises, is effectively run or not. However, such comparisons demand that there is some kind of measure, in the form of a parameter value, which can be compared. From a customer point of view it would probably be of interest to measure how effectively municipal operations are run (Avgiftsgruppen, 2002). Stakeholders that represent different groups of consumers and other citizens have increasingly begun to demand accountability from politicians and civil servants. Even from a management point of view there is an interest in measuring how efficiently different operations are run (Ammons, 1996; Tagesson, 2002a). Those measures can be a basis of action for improvement as well as supporting claims for accountability within the organisation. Measurement can constitute an important basis for communication between the organisation and its stakeholders (e.g. Ijiri, 1975).

Ijiri (1975), Ammons (1996), and Corvellec (1997) indicate performance measurement as an important condition for accountability. Ammons suggests that it could be useful in numerous management tasks such as planning and follow-ups. Ijiri points out that performance measurement is used within the organisation for decision-making and goal-setting, but he also indicates its use in communicating with the organisation's stakeholders.

> Implicit in accountability are the goals to be achieved. Proper accounting for activities, therefore, calls for the measurement of performance with respect to these goals. (Ijiri, 1975, p 33)

Ijiri thus indicates that measurement is not only of interest with respect to demands of accountability but is also of great importance for management control and the establishment of goals for improvement. A similar view is held by Epstein (1984): "… the link between performance and accountability is communication" (p 129). Fine and Snyder (1999) cite

performance measurement as the first step in the process to improve an operation's service and efficiency. They define performance measurement as:

> … the selection, definition and application of performance indicators, which quantify the efficiency and effectiveness of service delivery methods. (Fine and Snyder, 1999, p 24)

It could also be defined as the measurement of costs and/or quality for a certain activity or service that is carried out within an operation (e.g. Coe, 1999).

Measuring and defining performance measures may be useful when management wants to analyse how the operation is developing over time, but the measures say nothing about whether the operation is run efficiently or not. Benchmarking, therefore, may be the necessary next step to developing and improving the operation.

There is a strong link between the existence of performance measures and benchmarking. The benchmarking process would not be possible without performance measures. Nyhan and Martin (1999) also point out that a logical starting point in a benchmarking process is to begin comparing and contrasting the performance of one's own activity with the performance of other similar activities. These comparisons should help one to identify best-practice providers. By using these providers as references and models it should be possible to improve one's own activity (Spendolini, 1992; Martin and Ketter, 1996; Nyhan and Martin, 1999).

Interest in benchmarking and comparison of performance measures has increased among municipalities in recent years (e.g. Coe, 1999). Today it is common in many municipal operations to use performance measures to plan and follow up activities (Ramberg, 1997). Ammons (1999) sees two possible explanations why interest in benchmarking has recently shown a marked increase among municipalities. The first explanation could be that government officials finally conceded to media pressure and their use of cross-sectional comparisons. The other explanation could be that government officials have been influenced by successful benchmarking projects in industry. Both explanations are possible and one could therefore assume that the interest is partially driven by motives other than merely striving to improve operations. It may at first hand be the giving and demanding of "explanations, justifications and excuses" (compare Kirk and Mouritsen's (1996) definition of accountability,

p 245) that underlies this appetite for benchmarking and comparison of performance measures in the municipal sector.

To sum up: The use of performance measures assumes that comparisons are possible. The comparisons of performance measures can be used for different purposes. They can be used as part of a benchmarking process that aims to improve the operation, but also as a standard in the context of accountability.

No matter the purpose of performance measure comparison, it is important that the compared measures actually are comparable. If not, erroneous decisions can be made, both in the matter of demand for responsibility and in the matter of actions taken to improve the operation.

This paper will deal with the questions of how to obtain commensurability between performance measures from different organisations. The overarching aim of the study is to analyse the conditions for comparisons and benchmarking from performance measures used in the water and sewage sector in Sweden. This implies that I will analyse these particular performance measures in the light of how well they reflect the most important qualities of the measured objects, as well as their relevance, commensurability, and reliability.

The paper is organized as follows. Part 2 describes the theoretical framework and the basis of analysis. Part 3 describes empirical findings from three benchmarking projects. Part 4 presents the analysis. Part 5 presents concluding remarks.

2. Theoretical framework

Performance cannot be defined without settling issues of measurement. In other words, the design of the measures, such as assumptions, measuring standards, and available data, determines what the performance measures describe (Corvellec, 1997; Ramberg, 1997). The measures used to estimate performance can be discussed partly in terms of description and partly in terms of measurement. The dimension of description addresses what type of data the measures contain and how they are combined and presented (section 2.1, following), while the dimension of measurement addresses how the data are defined, collected, and registered (section 2.2).

2.1 The description of performance measures

Ammons (1996) identifies four types of performance measures useful for municipal operations:

- *Workload measures,* which indicate the amount of work or the number of incidents, etc.
- *Efficiency measures,* which reflect the relationship between work done and the resources required getting the work done
- *Effectiveness measures,* which depict the degree of goal achievement or otherwise reflect the quality of the performance
- *Productivity measures,* which combine the dimensions of efficiency and effectiveness.

The various measures describe different things. In relation to each other, the measures are what Ramberg (1997) calls 'description specific'.

Osborne and Gaebler (1994) state that both effectiveness and productivity measures are important, but when public organisations start to measure performance they often merely use efficiency measures.

Measures can be either absolute or relative (Mossberg, 1977). Absolute measures, such as a cubic metre of water, are easier to understand than relative measures, but on the other hand they do not give as much information as the relative measures (Ramberg 1997). Relative measures, usually called 'key figures', often are a condition for comparisons to be made. The denominator of the measure works as a kind of benchmark that facilitates meaningful comparisons; for example, a study of gasoline consumption for a car without relating it to driving distance would be rather meaningless.

Different measures can be more or less comprehensive (Östman, 1980), that is, composed of fewer or greater numbers of data (Östman, 1980; Ramberg, 1997). Measures with a high grade of comprehensiveness will make it possible to capture efficiency as well as effectiveness in a single measure, provided that the measure is developed and chosen with adequate care and attention.

There are certain fundamental criteria for developing and choosing a good measure. Johansson and Östman (1992) point out four basic criteria for suitable measures:

- *The effect criterion*: What kind of effect on individuals, organisations, and society has the information given by the measure? According to Johansson and Östman (1992), the information value presupposes the kind of effect it has on any outcome.
- *The measurement system criterion*: How good is the quality with respect to the ability to withstand data absorption, processing and utilization? Measures can be judged with respect to what extent, on the one hand, the data are verifiable and reliable; and on the other hand, as to what extent the measures themselves are comparable (ibid.).
- *The reproduction criterion*: How well do the measures capture the most important qualities of the measured objects? According to Johansson and Östman (1992), managers have a tendency to emphasize the measurement system criterion, while the people they manage emphasize the measures' relevance, that is, the reproduction criterion.
- *The user criterion*: How well are the measures and the information they give adaptable to the qualities and abilities of the users? Good measures should be accepted and be understandable to the users (ibid.).

The effect criterion must be regarded as the primary criterion, because the effect measures normally become the ultimately conclusive argument for selection of measures (Johansson and Östman, 1992). The other criteria, according to Johansson and Östman, must be seen as links in the judging of effects. Furthermore, it is an established fact that there is a mutual relation between the criteria of measurement system, reproduction, and user. Johansson and Östman point out that the users' apprehension of measures is unlikely to be independent of the reproduction quality, which depends on how good the quality is with respect to absorption and processing of data. At the same time, the criteria can often contradict one another. It is common, for instance, that the matters of reproduction and consistency have to be weighed against each other (ibid.). To find general measures that will entirely fulfil all users' demands and requirements will not be possible. A more reasonable starting point, therefore, is to search for a measurement system that is well adapted for continuous use with many purposes and which has ample penetrative power (ibid.).

Ammons (1996) provides a more detailed list of expectations for suitable measures, which is also in line with Johansson and Östman's four basic criteria[1]:

- *Valid*. They measure what they purport to measure.
- *Reliable*. They measure accurately and exhibit little variation due to subjectivity or use by different raters.
- *Understandable*. Each measure has an unmistakably clear meaning.
- *Timely*. The measures can be compiled and distributed promptly enough to be of value to operating managers or policymakers.
- *Resistant to perverse behaviour*. The development of a performance measure raises the profile of the performance dimension being examined.
- *Comprehensive*. The most important performance dimensions are captured by a suitable set of measures.
- *Nonredundant*. By favouring unique measures over duplicative measures, …each measure contributes something distinctive.
- *Sensitive to data collection cost.*
- *Focused on controllable facts of performance*. … Good sets of performance measures emphasize outcomes or facets of performance that are controllable by policy initiatives or management action.

Harty et al. (1992) use practically identical criteria to those of Ammons (1996) when choosing measures and methods for data collection.

Foltin (1999) provides a similar enumeration:

> It is essential that the measures chosen are available, definable, consistent, constant over time, timely, understandable, comprehensive yet not overlapping and true to established goals and objectives (p 43).

If the measures are to be meaningful, the user must be able to relate them to some kind of standard (Foltin, 1999). These standards—that is, benchmarks—can, according to Foltin (1999) be:

- goals settled by professional bodies (e.g. trade associations)
- other organisations' performance

[1] The list given here is a somewhat shortened version of Ammons' list (for full details see Ammons, 1996,13–14)

- a compiled index of selected organisations
- settled goals or former performance within the own organisation

The latter consideration is an expression for how the operation develops over time, but does not say whether the operation is run in a comparatively good way or not. Accordingly, a natural development of the work with performance measures is to make use of some kind of benchmarking or comparison with other operations.

2.2 Measurement characteristics

If performance measures are to be used for comparisons with other organisations, it is not enough that the two organisations have agreed upon the description of the measurements. The characteristics of the measurements are equally critical**.** The following discussion separately addresses two type of issues: organisational considerations (2.2.1) and allocation considerations (2.2.2).

2.2.1 Organisational considerations

It is important to agree upon how to classify the operation into activities or functions (Ramberg, 1997) and on what aggregation level the measures should be done (Coe, 1999). In other words, this calls for agreement on what the measured objects are. Magnusson and Wahlgren (1995) define an object, in this sense, as a phenomenon for which economic reports and estimates are set up, a definition that is well in line with how Östman (1980) uses the concept.

The problem with organisational considerations is illustrated by Andersson's (1998) deduction when, studying a benchmarking project within the Swedish municipal educational system, he points out that different organisational solutions considerably affect the outcome of performance measures and make it difficult to do comparisons. The problem may be further illustrated with a couple of examples.

Within local police authorities the categorization of personnel and cases can be developed differently:

> In police investigations, police departments assigned cases to detectives in two ways. Some departments assigned all reported Part 1 cases to detectives while the others assigned only more serious Part 1 cases, leaving less important cases in the hands of patrol officers. Departments did not consistently keep the clearance rates of cases investigated by patrol officers. (Coe, 1999, p 112)

Here is another example, from waste management:

> Measuring the cost of garbage collection was a simple matter. Each city knew the number of tons collected. However, getting a handle on trash collection was more difficult. … The cities collected, in varying combinations and definitions: yard waste, leafy items, hazardous materials and bulk goods. Of these, only yard waste and leafy items could be compared. Bulk goods and hazardous materials were left out of the study because the nature of what the cities collected in these categories varied so widely. (ibid.)

The list of examples could be much longer.

2.2.2 Allocation considerations

Another problem when comparing performance measures between different organisations is the difference in how the organisations calculate and account for costs:

> The biggest pitfall in comparing costs among governments is failing to recognize the differences in their budgeting and accounting systems. Governments vary widely with respect to how they budget for departments, divisions within departments, line items within divisions, funds and indirect costs. (Coe, 1999, p 3)

The reported information about cost data is built upon more or less subjective assumptions (e.g. Cheng and Harris, 1998), which means that the commensurability is not as good as one at first sight might believe:

> … performance measurement systems, which are presented as objective, are grounded in subjective judgements related to constructions of indicators and data; and result in the selective and subjective integration of information. (Cheng and Harris, 1998)

If meaningful comparisons are to be made, the demand for harmonised accounting must be high. However, harmonisation is not enough; the accounting principles applied must also lead to robustness of the accounting model, and the principles must be used correctly in a true and

fair way (Tagesson, 2002a, 2002b). Different principles can lead to different allocation of costs. This applies also to the allocation of overhead costs within time as well as allocation of investment expenditure over time (Tagesson 2001). The problem may be illustrated by a few more examples.

One condition for comparisons between two organisations to be true and fair is that apportionment between periods should be built upon similar judgements and principles within the two compared organisations. The question is of special interest if the cost for depreciation and capital charge is an essential part of the total costs:

> How should the costs of fixed assets be allocated over their service lives? (Cheng and Harris, 1998)

Malm and Yard (1985) put two requirements on the theoretical base for calculating the costs of fixed assets: cost accuracy and time accuracy. By cost accuracy they mean that the method of calculating the costs of fixt assets (including both cost of depreciation and capital charge) should give an annual cost, the total present value of which should be equivalent to the investment expenditure. As for time accuracy, Malm and Yard suggest the following definition:

1. The depreciation should occur within a reasonable estimated economic life.
2. There should be an acceptable purpose for the periodic incidence of depreciation within this economic life.

Within many municipal operations, there exist large investments with long economic life. When the depreciation plan is decided in connection with the investment activity, it is a question of an approximation afflicted with uncertainty.

When it comes to the periodic incidence of the expenses, it is possible to have different **bases** for matching. Yard (1997) identifies at least six different matching alternatives. This shows the complexity of allocating costs to different time periods. Differences in the application of principles and estimates, as well as the judgements made can accordingly lead to different patterns of allocation and consequently to differences in accounted costs in various accounting periods. A simulation of different principles, estimates and judgements, all used in Swedish water and sewage operations (Tagesson, 2001), in one and the same operation showed that

this could affect the reported costs as much as 20% within the same accounting period (Tagesson, 2002a).

Also, questions of allocation of overhead costs within time become central when it comes to comparisons of performance measures. During the last decades the overhead costs as a share of total costs have increased (e.g. Johnson and Kaplan, 1987). This fact denotes the complication of comparisons based on performance measures. There are, in fact, a number of different principles and criteria for allocation of overhead costs.

Ahmed and Scapens (1991) identify no fewer than five different criteria[2] for allocation of overhead costs. Use of different criteria within the compared organisations will lead to dissonance and misleading comparisons.

Besides such problems there are many other delimitation and consideration problems that might be handled differently in different organisations. A classical example is the delimitation between maintenance and investment.

2.2.3 Summing up

If measures are to be useful, they must be able to relate to some kind of standard, otherwise it is not possible to judge whether the performance illustrated by the measure is good or bad. This standard, or benchmark, is usually a comparison with performance measures from other organisations.

However, if the comparisons are to be meaningful, there must be correspondence between the description and measurement characteristics of the measures used in the different organisations. Since the measures work as information concepts, information correspondence is necessary. Information correspondence means that the measures are defined and applied homogeneously and consistently (e.g. Magnusson and Wahlgren, 1995), and requires unity about:

- what the measure should describe
- what the measured object is
- how employment of resources are defined

[2] The criteria were: (i) neutrality, (ii) ability to bear, (iii) cause and effect, (iv) received benefit, and (v)

This theoretical review shows that there many different pitfalls when it comes to the design of measures to be used in various comparisons. The use of measures for comparison purposes calls for unitary definition and application of the measures and the data they comprise. The measures also need to be chosen carefully if they are to relevantly and reliably reflect the measured objects' most important qualities.

In the remainder of this paper, I will report on the study of three different benchmarking projects in the Swedish water and sewage sector and on the performance measures used in those projects, examining the following hypothesis:

Hypothesis

Several of the performance measures used in comparable purposes do not in a relevant and reliable way reproduce the measured objects most important qualities, which should be compared.

3. Measures for comparisons and benchmarking – experiences from three benchmarking projects in the Swedish water and sewage sector

This part of the paper will describe the design of measures used for comparisons and benchmarking. The empirical support is based on interviews, observations, and written documents from three benchmarking projects:

Water and sewage cooperation in the south of Dalarna (WAC-SD),
Water and sewage cooperation in the south of Sweden (WAC-SS), and
A project run by the industry body Swedish Water (SW).

3.1 The measured objects

The measured objects are decided according to which classifications of activities and functions are performed. The most aggregated level is the (water and sewage) operation as a whole. A classification in a less aggregated level, often used in the water and sewage context, is one based on operation by functions. In all three benchmarking projects the following classification into functions were used:

- Production of drinking water (PDW)

reasonableness and fairness.

- Purifying of sewage water (PSW)
- Distribution of drinking water (DDW)
- Distribution of sewage water (DSW)
- Administration (Adm)

These functions were then divided into a number of different subordinated functions or activities. In WAC-SD, they had some measured objects that were broken down into different sub-systems for water supply and sewage operations, respectively. Table 1 shows the measured objects on different aggregation levels.

Table 1 Measured objects on different aggregation levels

Aggregation level	Measured objects				
Level 1	The water and sewage operation as a whole				
Level 2	PDW in total	PSW in total	DDW in total	DSW in total	Adm in total
Level 3	PDW broken down to sub systems	PSW broken down to sub systems	DDW broken down to sub systems	DSW broken down to sub systems	

A problem with the comparisons, which is pointed out by the respondents, is how the water and sewage operation as a whole should be separated from other municipal operations. This matter is not self-evident. For example**,** with respect to resources for administration, these are often shared with other municipal operations or divisions when the water and sewage operation is run within a legal form of business activity. There is also an unclear frontier between the water and sewage operation and the street operation regarding the handling of drainage water. In the same way there is an uncertainty when it comes to the handling of fire alarm posts. Whose responsibility are the fire alarm posts—the water and sewage operation's or the fire department's? Divergence on the highest aggregation level of causes gives after-effects on lower aggregation levels.

The possibilities of using measures, with reference to these measured objects, also are limited by the fact that situation-dependent variables differ substantially between the water and sewage operations. For example, among the operations within the WAC-SD project there is one operation, Borlänge, where the water supply system is integrated into one big system with

two waterworks and one purifying plant, while in Säter the water supply system is divided into twelve sub-systems and the sewage system into seven sub-systems.

3.2 The measures

In all three benchmarking projects studied, management worked with workload measures, efficiency measures, and effectiveness measures (e.g. Ammons, 1996). The number of measures used varied greatly among the three benchmarking projects. In none of the three projects were productivity measures used. The total number of measures for each project and their distribution according to Ammons' (1996) classification is shown in Table 2.

Table 2 Classification of used measures in the three benchmarking projects[3]

Projects	Workload measures	Efficiency measures	Effectiveness measures	Total number of measures
WAC-SD	10	25	35	70
WAC-SS	5	18	11	34
SW	20	25	10	55

3.2.1 The efficiency measures

In WAC-SD and WAC-SS, efficiency measures were defined for measured objects on aggregation level 1 and 2, i.e. the water and sewage operation as a whole and broken down on the different main functions (PDW in total, PSW in total, DDW in total, DSW in total and Adm in total). Depending on the objective, measures on all three aggregation levels could be derived in the SW program.

All efficiency measures were designed as key ratios, where cost data were related to some kind of quantity data. The cost data were expressed as total cost for the measured object including as well as excluding costs for depreciation and capital charge. The quantity data were expressed in WAC-SS as the amount of charged cubic metres of water. In WAC-SD, the quantity data for the operation as whole and the functions PDW and PSW were expressed as number of connected persons, and the amount of cubic metre of water registered at the waterworks, or alternatively at the purifying plants, depending on the measured object. For the

[3] Sometimes the delimitation between workload measures and effectiveness measures is hard to define. In doubtful cases the measures have been classified as effectiveness measures.

functions DDW and DSW, the quantity data used were number of connected persons and length of pipes, respectively. SW used the amount of charged cubic metres of water and number of connected persons for the aggregation level operation as a whole. For the measured objects PDW and PSW, produced/ processed amount of water, expressed in cubic metres, were used as quantity data. For PSW, the quantity data also were expressed as number of connected persons plus biological oxygen demand from industry. For the functions DDW and DSW, quantity data were expressed as the amount of charged cubic metres of water, number of connected persons, and length of pipes. In WAC-SD, there was also a key figure for Administration, expressed as costs for administration related to total costs. SW had similar measures but with a higher level of cost breakdown.

3.2.2 The effectiveness measures

All benchmarking projects used a large number of effectiveness measures. WAC-SS used those kinds of measures on aggregation level 1 and 2, while WAC-SD and SW used them on all aggregation levels. WAC-SS and SW used compounded measures, while WAC-SD used both absolute and compounded measures and one multiple measure. Examples of effective measures expressed as absolute measures were number of closedowns and customer complaints. Examples of compounded measures used were number of cellar flooding per 1000-connected people, number of stops per 10 kilometre pipes, etc. The multiple measure that WAC-SD used was loss (expressed in litre) per day and meter pipes. WAC-SD had graded all effectiveness measures on a four-grade scale. The value on the measure could be indicated as:

1. Very good persistence
2. Good persistence
3. Less good persistence
4. Bad persistence

3.2.3 Definition and absorption of data

In all three benchmarking projects, care was taken when defining quantity data used in the measures. The quantity data were chosen with care to be verifiable, reliable, and easy to capture. As for cost data, the definitions in both WAC-SD and WAC-SS were meagre. No discussion about the reliability or verifiability existed about the cost data used, except in one report from WAC-SD, where it was stated that the participants' economic accounting was so

diversified that it was impossible to do true and fair comparisons of measures which included cost data (Hägerman, Johansson, Johansson and Larsson, 2000). SW also established the fact that the water and sewage operation's method of accounting did vary considerably and that consequently it was hardly possible to make economical comparisons among water and sewage operations (VAV, 2001). SW then defined and proposed common accounting standards and principles, with the purpose of harmonising the accounting among Swedish water and sewage operations. Unfortunately, some of the purposed accounting principles, such as net accounting and direct write-off of reinvestments, contradict generally accepted accounting principles and do not lead to robustness in the accounting model (e.g. Tagesson, 2002a, 2002b), thus rendering true and fair comparisons more difficult rather than facilitating them (ibid.).

4. Analysis

This part of the paper analyses the design of the measures on the basis of Johansson and Östman's (1992) four basic criteria.

4.1 The measures in relation to the effect criterion

A fundamental condition for performance measures to be considered as meaningful is that the information given from the measures should have some kind of effect on individuals' and organisations' actions and performance. It is also reasonable to assume that the information given by the measures can have different effects, depending on what kind of user is receiving the information. In the three benchmarking projects cited I did not perceive any articulated wish about effects of the use of the measures. In the three benchmarking projects, other than generally expressed terms like: "… possibilities … for … comparisons among water and sewage operations and thereby increase the basis for exchanging experiences and getting stimulus for improvement and productivity manoeuvres" (translation from a WAC-SS document; VASAM, 2001, p 2). The identification of actual effects from the use of performance measures would require another kind of study than this, so I chose not to do further analysis of the measures in relation to this criterion.

4.2 The measures in relation to the user criterion

There is some uncertainty as to who are the contemplated users of the measures. There are many potential stakeholders, such as government officials, politicians, or subscribers, who could be interested in performance measures regarding water and sewage operations. Since

those different stakeholders have different prerequisites and qualifications, it is doubtful whether they all need access to the same types or numbers of measures. It would be desirable to have a more evident analysis of how measures could be adapted for different users on basis of their need of information and their potential use of that information. However, the kind of measures of interest for different groups of stakeholders was not discussed within these three benchmarking projects. While the question may not be of primary interest for this study, it is still of general interest since there is a mutual connection between user criterion, reproduction criterion, and measurement criterion (e.g. Johansson and Östman, 1992).

4.3 The measures in relation to the measurement criterion

In their view of the measurement criterion Johansson and Östman (1992) were concerned with the measures' quality with respect to the ability to absorb data, and undergo data processing and utilization. Measures are judged with respect to the possibility of verifying the data used in the measures, and as to whether they are reliable and comparable (ibid.).

As for the quantity data used in the effectiveness measures and in the denominator in the efficiency measures, significant effort was made to define those data. The careful definition of the quantity data was aimed at ensuring commensurability. The data were chosen with respect to their use for verification and recording purposes. The cost data that constituted the numerator in the efficiency measures were used without any adjustment although there was an understanding that those data were misleading. There were some discussions about allocation of overhead costs among different measured objects and functions, but with respect to the allocation of investment expenditure over time there were no serious discussions about the effect of application of different principles or how commensurability could be obtained.

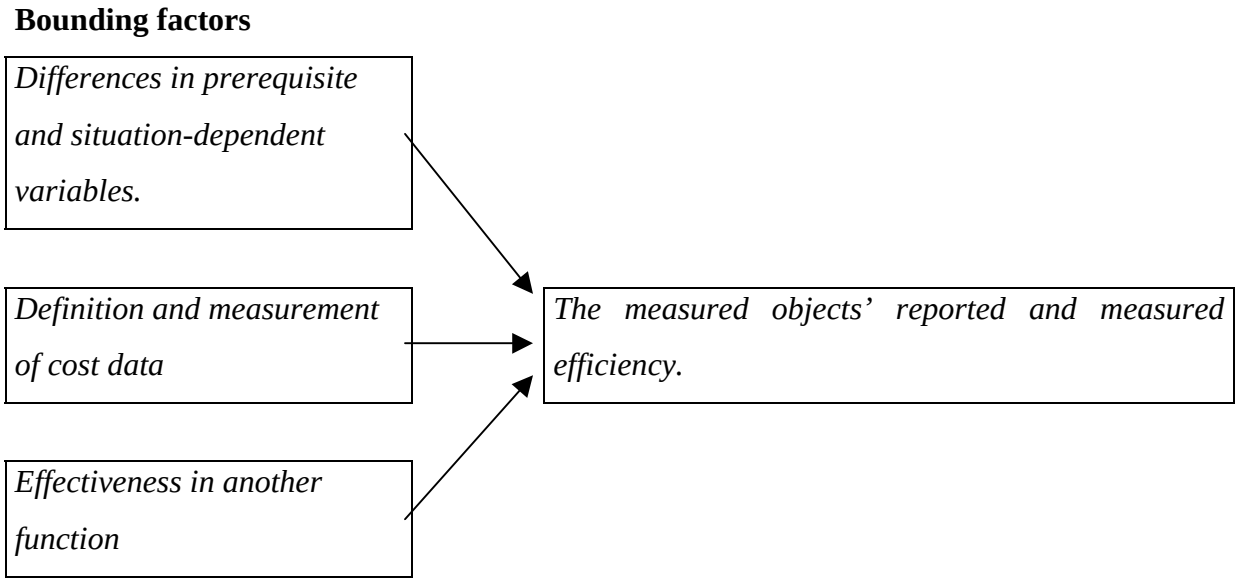
This insufficiency of reliability and commensurability of cost data indicated that the commensurability of efficiency measures was limited or even impossible.

Even when the data are strictly defined, there could be reasons to question the commensurability. Among the organisations that took part in the benchmarking project some very different conditions were in evidence. It is natural to presume that situation-dependent variables constitute restrictions for action-dependent variables. It seems that the organisations chose benchmarking partners and their standards from a geographical point of view rather

than from situation-dependent variables. In many cases the measures also were too aggregated and stereotyped. Surprisingly enough, there were no discussions in the benchmarking projects about the fact that the effectiveness in one function could affect the efficiency in another function.

Figure 1 illustrates some of the factors that bound the commensurability of efficiency measures.

Figure 1. Factors that bound commensurability of efficiency measures



To supplement Figure 1, an example where the measured object is the function Production of Drinking Water (PDW), will be given. The accessibility to subsoil water varies among different municipalities. This factor affects the production cost for drinking water. Another important factor that affects this cost is the capacity of the waterworks. An underestimated size of the waterworks will lead to high extra costs when additional drinking water must be produced. Over-dimensioning of the waterworks does lead to additional costs that on a short-term basis cannot be affected. There can be a large difference between theoretical efficiency—i.e. the efficiency when full capacity is utilized—and capacity that can be affected—i.e. efficiency when dimensioning and other external assumption have been considered. The application of different principles, judgements and estimates in the accounting, i.e. the measurement of cost data, affects the numerator in the efficiency measures and by that the

reported efficiency. It is not only the subscribers' consumption of water that affects the need and quantity of production of drinking water. If the effectiveness is low in the Distribution of Drinking Water (DDW) function inasmuch as the amount of loss is large, then the need for production of drinking water increases, which in turn affects the efficiency measure for the measured object PDW, which uses recorded amount of water in the waterworks as quantity data in the measures denominator.

4.4 The measures in relation to the reproduction criterion

The reproduction criterion concerns how well the measure captures the measured objects' most important qualities (Johansson and Östman, 1992). The higher the aggregation level, the harder it is to capture these qualities. The most important quality is not unambiguous. The empirical support showed that amount of water and number of connected people were often used as data in the denominator both in efficiency and effectiveness measures. These are not the only possible benchmarks to use when measuring efficiency and effectiveness. The number of subscribers can vary considerably even if the numbers of connected people or water consumption are the same. There could be differences in the number of subscribers such as companies and institutions to which no connected people can be assigned through population records. The same problem occurs for subscribers who use their houses as vacation homes. Further, a few subscribers or connected people can make use of a large amount of the produced drinking water. The cost of distribution of water and sewage is not primarily affected by the number of connected people or the amount of produced water. Rather it is the scattering or localization of subscribers that is crucial for the cost level for this kind of function.

The foregoing discussion has shown that to aggregate measures in this case would cause a deterioration in relevance and reproduction. The problem becomes bigger when the measured object is assignable to a higher aggregation level. Measures that are too complicated, on a lower aggregation level, can however make the measures difficult to understand, and that would not be desirable either.

5. Concluding remarks

The overarching aim of this study was to analyse the conditions for comparisons and benchmarking from performance measures used in the water and sewage sector in Sweden.

The empirical support of the study showed that with the measures and measurement methods used, it is very doubtful whether meaningful comparisons and benchmarking is possible to do on the basis of performance measures. The fact that the criteria for choosing benchmarking partners are too simple does not improve the potential for successful and meaningful benchmarking.

Many of the measures used in the benchmarking projects and communicated to external stakeholders were too aggregated **to** capture in a relevant way, reproduce, or describe the most important qualities of the measured objects. The so-called reproduction criterion of Johansson and Östman was accordingly in many cases badly fulfilled. The higher the aggregation level on the measured object, the worse became the measure's consistency. Aggregated measures were used to facilitate understanding for different users. The reproduction criterion and the user criterion are consequently in conflict with each other, at the same time as there is a mutual connection between the two criteria (ibid.). If the user criterion is to be fulfilled, the measures must be accepted by, and understandable to, the users (ibid). Too many and too complicated measures decrease comprehensibility, while too aggregated measures impede reproduction and likely acceptance.

The hypothesis that several of the performance measures used in comparable purposes do not in a relevant and reliable way reproduce the measured objects' most important qualities, which should be compared, could accordingly not be rejected.

The usefulness of the performance measures in the three benchmarking projects in this study must accordingly be strongly questioned. The positive effects of the benchmarking projects probably had the character of experience exchange on a more general level. The possibility of using the measures as standards in the context of accountability also becomes very limited, since differences in outcome may be explained by many factors other than the agents' management of the resources.

References:

Ahmed, M. N. and Scapens, R. W. (1991). "Cost Allocation Theory and Practice – The Continuing Debate." In Ashton, D., Hopper, T. and Scapens, R. W. (eds), *Issues in Management Accounting*. Hemel Hempstead: Prentice Hall.

Ammons, D. N. (1996). *Municipal Benchmarks – Assessing Local Performance and Establishing Community Standards.* Thousand Oaks: Sage.

Ammons, D. N. (1999). "A Proper Mentality for Benchmarking", *Public Administration Review,* Vol 59, p 105-109.

Andersson, P. A. (1998). *Benchmarking I kommuner.(Benchmarking in Municipalities)* Lund: KEFU.

Avgiftsgruppen (2002). *Fastigheten Nils Holgersons underbara resa genom Sverige - En avgiftsstudie för år 1999. (Realtor Nils Holgerson's wonderful journey through Sweden – A study of local charges for the year 2000)* Stockholm: Sveriges Fastighetsägare. http://www.nilsholgersson.nu/Avgiftrapport/2002/AvgiftsrapportNH2002.htm, accessed 2003-02-07.

Cheng, R. H. and Harris, J. (1998). "The New Public Management: Implicit Acceptance of the Regimen and Subjectivity of Accounting", conference paper, *Accounting for the New Public Management*, European Institute for Advanced Studies in Management (EIASM),Venice, Italy, September 17 – 19.

Coe, C. (1999). "Local Government Benchmarking: Lessons from Two Major Multigovernment Efforts", *Public Administration Review,* Vol 59, p 110 – 123.

Corvellec, H. (1997). *Stories of Achievements – Narrrative Features of Organizational Performance.* New Brunswick, NJ: Transaction Publisher.

Epstein, P. D. (1984). *Using performance measurement in local government: A guide to improving decisions, performance and accountability.* New York: National Civic League Press.

Fine, T. and Snyder, L. (1999). "What is the difference between performance measurement and benchmarking?", *Public Management (US),* Vol 81, p 24 – 24.

Foltin, C. (1999). "State and local government performance", *Government Accountants Journal,* Vol 48, p 40 – 46.

Hägerman, T., Johansson, P., Johansson, Å. and Larsson, J. (2002). *VA-plan 2050 VA-samverkan I södra Dalarna (Water and sewage plan 2050 water and sewage cooperation in the south of Dalarna).* Stockholm: Svenska vatten- och avloppsverksföreningen.

Harty, H. P., Blair, L. H., Fisk, D. M., Greiner, J. M., Hall Jr., J. R. and Schaenman, P. S. (1992). *How effective are your community services? – Procedures for measuring their quality.* The Urban Institute and ICMA.

Ijiri, Y. (1975). *Theory of Accounting Measurement.* Sarasota, Florida: American Accounting Association.

Johansson, S-E. and Östman, L. (1992). *Lönsamhetskrav – redovisningsmått – styring. (Requirement for profitability - accounting measures – management).* Lund: Studentlitteratur.

Johnson, H. T. and Kaplan, R. S. (1987). *Relevance Lost: The Rise and Fall of Management Accounting.* Boston: Harvard Business School Press.

Kirk, K. and Mouritsen, J. (1996). "Space of accountability. Systems of accountability in a multinational firm", in Munro, R. and Mouritsen, J. (eds) *Accountability, Power, Ethos and the Technologies of Managing.* London: International Thomson Business Press.

Magnusson, Å. and Wahlgren, H. (1995). *Utveckling av ekonomistyrsystem enligt RP*

(Development of economic control systems according to RP). Stockholm: Industrilitteratur.

Malm, A. T. and Yard, S. (1985). *Kapitalkostnader I kommunal verksamhet (Depreciation and capital charge in municipal operations.)* Lund: Studentlitteratur.

Martin, L. L. and Ketter, P. M. (1996). *Measuring the performance of human service programs.* Thousand Oaks, California: Sage.

Mossberg, T. (1977). *Utveckling av nyckeltal (Development of key figures).* Stockholm: EFI.

Nyhan, R. C. and Martin, L. L. (1999). "Comparative Performance Measurement", *Public Productivity and Management Review*, Vol 22, p 348 – 364.

Osborne, D.and Gaebler, T. (1994). *Sluta ro – börja styra (Reinventing Government).* Stockholm: Tidens Förlag.

Ramberg, U. (1997). *Utformning och användning av kommunala verksamhetsmått (Design and use of performance measures in municipal operations)* Lund: Lund University Press.

Spendolini, M. J. (1992). *The benchmarking book.* New York: AMACON.

Stahre, P., Adamsson, J. and Eriksson, Ö. (2000). *Nyckeltal inom VA – En introduktion (Key ratios usable in the water and sewage sector – An introduction)* Stockholm: Svenska vatten- och avloppsverksföreningen

Tagesson, T. (2001). "Redovisning och beräkning av kapitalkostnader i kommunala va-verksamheter" (Calculation and accounting of costs for depreciation and lost interest income according to capital tied up in assets, within municipal water and sewage enterprises). *Kommunal ekonomi och politik (Municipal economy and politics),* 5:45 – 84.

Tagesson, T. (2002a). *Kostnadsredovisning som underlag för benchmarking och prissättning (Cost accounting as a basis for benchmarking and pricing).* Lund: Lund University Press.

Tagesson, T. (2002b). "Effective benchmarking calls for harmonization and use of accounting principles leading to a robust accounting model", *Working Paper Series 2002:3. Department of Business Studies*, Kristianstad: Kristianstad University College.

VASAM (2001). *Nyckeltal I VA-verksamheter, Utveckling 1996 – 1999 – Jämförelser inom VASAM-gruppen (Key ratios within water and sewage operations, Development 1996 – 1999, Comparisons within the WAC-SS group.* (Unpublished manuscript).

VAV (2001). *Statistik inom VA-branchen – VAV:s framtida statistik (Statistics within the water and sewage business – The future statistics of Swedish Water).* Stockholm: Svenska vatten- och avloppsverksföreningen (Swedish Water).

Yard, S. (1997). *Beräkning av kapitalkostnader (Calculation of depreciation and capital charge).* Lund: Lund University Press.

Östman, L. (1980). *Beteendeinriktad redovisning (Behavioural accounting).* Stockholm: EFI.