Performance Assessment Using Key Performance Indicators (KPIs) for Water Utilities: A Primer

By Sanford V. Berg

Professor Emeritus, Economics—University of Florida
Senior Fellow, Public Utility Research Center

February 26, 2020 (Revised) ***Water Economics and Policy,*** Vol. 6, No. 2 (2020)

Abstract

Key Performance Indicators (KPIs) are widely recognized as a basis for evaluating water utility operations in developing countries and for designing both regulatory and managerial incentives that improve performance. A number of methodolgies can be used for assessing performance, with KPIs and OPIs serving as more comprehensible and potentially more comprehensive than more technical empirical benchmarking studies. Data initiatives in low and middle income countries require resources that could be used for other activities with more immediate payoffs. However, regulatory oversight requires data analysis of trends, current performance and realistic targets. Quantitative studies can provide clues regarding the extent of economies of scale, scope, and density, but policy-makers need much more detail and specificity than most scholars provide. Here, the focus is on information systems that provide accurate, reliabile, and relevant data. KPIs represent the foundation for those developing, implementing, and responding to public policy--incentivizing water utilities in developing (and developed) countries to contain costs, improve service quality, and expand water access over the long run.

1. Introduction

This paper on assessing water utility performance takes a primer/tutorial approach (with examples from case studies). Surveys of production and cost functions in the water utility sector have identified over two hundred quantitative studies (Berg and Marques, 2011; Cetrulo, et. al. 2019). Yet the use of these technical quantitative studies by regulators for actual rate cases is relatively limited in middle and low-income nations. One reason is that the data on which utilities are being evaluated are not very reliable. Another is that identifying high performers using stochastic frontier analysis (SFA) or data envelopment analysis (DEA) requires that decision-makers have confidence in the robustness of studies. In addition, it is a challenge to communicate the implications of quantitative studies to non-technical decision-makers (regulators who implement public policy and politicians who determine public policy). However, Key Performance Indicators (KPIs) provide a starting point for ranking utilities in a country. KPIs like non-revenue water (NRW), collections, coverage, and staff per 1000 customers provide an important starting point for improving performance in the water sector.

The advanced tools and new datasets available to economists have resulted in a plethora of publications. Many articles claim to have policy implications, even when the specific context of national situations are not highlighted. For example, the existence of significant scale economies does not necessarily imply that merger or consolidations will reduce unit costs. There may be cost savings from sharing information systems or utilizing engineers more effectively, but the networks are still miles apart, so the cost savings of bringing two systems together may not be substantial.

Utility performance scores (and associated rankings) are used to identify poor performers, Yet many studies do not (or cannot) control for unique circumstances faced by utilities related to topography, hydrology, customer characteristics, and other elements affecting costs. That context includes financial and political constraints (such as affordability and service quality as objectives), age of water networks and past maintenance, and the nature of regulation. Most authors acknowledge these limitations of their modeling efforts, but that does not prevent us (myself included) from drawing conclusions and making generalizations that are broader than can be adequately supported. For example, in their survey of studies, Cetrul, et. al., 2019) conclude " . . . regulatory incentives have not promoted performance improvement of water utilities in developing countries." (p. 378). They recognize that there are different types of regulatory systems, but still conclude that incentives do not impact performance! This is far too strong a conclusion, given the wide range of regulatory regimes (price caps, cost of service, hybrids, appropriate targets, K-factors, etc.) This observation suggests that scholars conducting SFA and DEA might present their conclusions with greater humility. No study is definitive; most are suggestive. The current paper argues that while simplistic approaches to performance comparisons are not ideal, they often convey enough information to motivate those with operating and oversight responsibilities to change their current strategies. In addition, careful data collection for a few indicators provides a strong foundation for more comprehensive studies that require time series and panel data (Berg, 2010).

Several years ago, the author prepared a survey of six books on water utility benchmarking activities involving performance assessment and improvement (Berg, 2013). Those volumes showed how collecting and analyzing KPIs enabled regulators and managers to do their jobs, since decision-makers manage what they measure. Quantitative evidence on trends over time, patterns across comparable utilities, and on highest performing operators is essential if realistic targets and incentives are to be established. Since it only summarized key lessons, the earlier article lacked substantive examples that might help those engaged in evaluating water utilities. With that in mind, this study focuses on the KPIs utilized by managers and by water sector regulators. The emphasis is on the potential benefits from devoting resources to data collection and analysis, and on the pitfalls associated with the misuse of data and misplaced confidence in particular information.[1] The focus is not on analytic techniques for assessing performance (such as statistical or data envelopment analyses) but on steps to be taken *prior* to more comprehensive quantitative research studies.[2] The purposes of this study are two-fold: (1) practitioners in developing countries can see how information systems are established that enable regulators to

[1] This study extends earlier work by the author on data availability (Berg and Phillips, 2017) and utilizes material developed for the Public Utility Research Center's Advanced Training Program "Benchmarking Infrastructure Operations." It reflects contributions from practitioners around the world. In addition, portions of this study draw upon material prepared by the author to answer a Frequently Asked Question on KPIs for www.regulationbodyofknowledge.org . The World Bank supported the development of that material, and that earlier study benefited from comments by Jemima T. Sy and Anna Aghababyan. Neither sponsoring organizations nor reviewers are responsible for this survey of KPIs.

[2] As of 2010, Berg and Marques (2011) had identified 190 quantitative studies using cost or production functions of water and sanitation services. A complete listing of the studies is available on-line.

monitor, evaluate, and incentivize operators; and (2) analysts can better appreciate the strengths and limitations of Overall Performance Indicators (OPIs) in establishing incentives.

Let us start with the definition for KPIs from the Glossary in the *Body of Knowledge on Infrastructure Regulation*[3]:

> "KPIs are indices of cost and production outcomes for a firm. KPIs enable decision makers to track trends and identify areas needing changes, including operational procedures, maintenance/inventory practices, and capacity investments. These indicators represent valuable information regarding whether infrastructure performance is improving or becoming worse, so tracking KPIs is essential for performance assessment and enhancement (benchmarking). KPIs can be classified into a set of categories that allow managers to establish primary responsibilities within the key departments of the regulated business (whether public or private). Since performance outcomes are inter-related and inter-dependent, creating an Overall Performance Indicator (OPI) is problematic. While specific departments might be responsible for collecting the data on one or more indices, performance will generally rely on the activities of several departments. Thus, the associated key performance outcomes refer to overall technical and financial operations, service quality, and customer experience. For example KPIs could include hours per day of service (technical), nonrevenue water or line losses (operational), cash flows and collections (financial), and customer complaints (indicating value for money, from the users' standpoint)."

The definition notes that KPIs go beyond cost and production outcomes to include other dimensions of performance, including customer perceptions. However, this definition does *not* acknowledge that some areas—like financial sustainability, professional capacity-building, and resource sustainability—require more comprehensive analyses of business plans, staff development programs, and long term hydrological conditions. Traditional KPIs often do not include these dimensions of performance involving long term outcomes. Thus, one limitation of KPIs is the tendency to focus on areas that reflect current operations rather than the sustainability of the utility. The more comprehensive definition (recognizing the limitations of benchmarking) provides an outline of this study: overall performance assessment for utilities, accuracy and reliability of individual KPIs, and limitations of current tracking and incentive systems. The sections identify some steps that can help address these three topics.

2. Overall Performance Assessment for Utilities

Water utility regulators in many countries utilize OPIs in ranking utilities and evaluating overall performance trends (Cabrera, Dane, Haskins, & Theuretzbacher-Fritz, 2010). A number of steps can be taken to limit aggregation problems: identify key performance objectives, use relative performance scores rather than rankings, ensure that weights reflect current priorities and recognize the importance of trends, group "comparable" utilities by basic operating conditions as well as size (to control for circumstances beyond managerial control), and present comparisons in ways that are clear and appropriate for target audiences. Examples from Albania and Peru are presented to illustrate how regulators have applied these steps. These countries were selected because the annual reports of their water sector regulators utilize KPIs in evaluating performance; other nations could have been chosen as well since the tools are applied in many developing countries.

[3] See www.regulationbodyofknowledge.org .

2.1 Identify Key Performance Objectives before focusing on Key Performance Indicators.

For a performance indicator to be "key", it must capture some dimension of performance that is important to those receiving service or to those who hope to have access in the future. Those responsible for providing oversight of infrastructure services and for delivering those services should reach a consensus regarding what is valued and what feasible. Generally, legislation will identify dimensions of performance that elected representatives seek to improve: for example, water service should be affordable, produced efficiently, and available throughout the nation. Key Performance Objectives could include availability, network outage impact reduction, robustness of operations to extreme events (including weather and conflicts), quality of service, downtime, notification of delivery problems, customer satisfaction (via surveys or number of complaints), integrity of billing and collection processes, affordability, access, efficiency, productivity, innovation, security, and safety. These goals can then be associated with a set of KPIs. Once the objectives are prioritized, and data collection has commenced, the performance evaluation process can begin: starting small, and slowly increasing the number and accuracy of KPIs. Another advantage of starting with objectives is that indicators that are "easy to measure, but relatively unimportant" will not become the focus of performance evaluation.

2.2 Use relative performance rather than rankings to create an Overall Performance Indicator (OPI).

Having a single indicator is convenient, though it can mask weak or strong performance in specific areas. Nevertheless, regulators and managers often combine KPIs to create an Overall Performance Index. Care must be taken to ensure the process is transparent and reflects priorities. In particular, aggregating a set of KPIs to create an OPI should *not* be done by adding rankings to create an overall score. Rather, analysts should utilize raw scores relative to best score (or the target score). For example, take a situation where three KPIs are used to evaluate the relative performance of three utilities depicted in Table 1. Higher scores are desired.

Table 1. Three Indicators (A, B, and C) and Three Utilities

Firm /	A	B	C
U1	**98 (1)**	**85 (1)**	**55 (3)**
U2	**96 (2)**	**83 (2)**	**60 (2)**
U3	**94 (3)**	**82 (3)**	**82 (1)**

Scores relative to 100% (with *Rankings* in Parentheses).

In the above example, U3 ranks third for indicators A and B, and comes in first for KPI C. If we add the *rankings*, U1 has five points, U2 has six points, and U3 has seven points—where the lowest score could be labelled "best". However, this conclusion would be incorrect, since the procedure treats ordinal numbers (rankings or positions) as though they were cardinal numbers that could be added together. Taking the best score for each indicator as the denominator, we can normalize the data and create a "relative performance" KPI score for each utility, as shown in Table 2:

Table 2. Three Indicators (A, B, and C) and Three Utilities

Firm /	A		B		C		OPI
U1	98/98	+	85/85	+	55/82	=	2.7
U2	96/98	+	83/85	+	60/82	=	2.7
U3	94/98	+	82/85	+	82/82	=	2.9

Individual Scores are relative to the Best Scores for each Indicator.

U3 has the best overall performance (2.9) if the three KPIs are added together (basically, given equal weight), where the other two utilities are tied with scores of 2.7. Of course, if measurement accuracy is low for the three indicators, the three utilities could be viewed as performing equally well. Note that instead of using the highest score in each KPI category, the Target KPI score could be used in the denominator. This shift will give some benefit to the utility whose performance is closest to particular targets. Such an adjustment might be appropriate if exceeding the target is very costly and the perceived benefit is small. Thus, the "best" choice for normalization depends on the purpose of the comparison. A comprehensive analysis would check the robustness of the results to different measures.

2.3 Give weights to the various indicators that reflect priorities.

In the above example, if KPI A is given a weight of .9, and B and C are each given a weight of .05, then the OPI changes and the revised rankings would also change. Clearly, determining the weights can be a challenge. In Peru, the water regulator (SUNASS) gives equal weight to each of nine KPIs, but there are different numbers of operational and financial KPIs, so weights are implicit: determined by the number of KPIs in the four categories (Corton, 2003):

1. Quality of service (three--compliance with residual chlorine rule, continuity of service, and percentage of water receiving chemical treatment),
2. Coverage of service (two--water coverage and sewerage coverage),
3. Management efficiency (three--a combination of service continuity and volume per person at a connection, percentage with meters, and the ratio of uncollected bills to total billings), and
4. Managerial finance efficiency (one--ratio of direct costs and other operating expenses to revenues)

Of course, one problem with KPIs (and with cost and production function studies) is that costs can be reduced today at the expense of greater costs in the future. Maintenance, staff training, and systematic remediation can all be deferred—improving indicators of current performance but burdening future customers. Thus, authorities setting targets and weights and providing oversight also need to monitor budgets to ensure that particular outlays are not deferred. However, avoiding micromanagement then becomes a challenge. Ther is no single, simple way to establish weights for an OPI, nor can the problem of information asymmetry be eliminated.

In the case of Uganda, the national water and sanitation utility has internal management contracts with each of its managers for geographic areas. The national targets reflect the priorities identified in the operator's performance contract with its oversight committee (established at the ministerial level). The targets and weights applied to the components in the incentive plans are jointly determined by the head

office and the local manager. Both weights and targets depend on the current levels of performance, the importance of achieving objectives from the standpoint of cash flows and customer satisfaction, and opportunities for improvement. The incentive framework is dependent on KPIs, targets and weights (Mugisha, 2011).

The Water Regulatory Authority in Albania (WRA) utilizes a much more complex system where a maximum number of points is given for meeting the target for each of ten KPIs (with zero points for not reaching a minimum, WRA, 2016). Furthermore, each target has a given weight. For example, two of the targets each have a weight of .20, so reaching a collection efficiency of 82% and non-revenue water of less than 30% would give the utility 40 of the potential 100 points. Albania also has a category called "Regulatory Perception" (weight of .05) based on compliance with the law. Within this category, equal weight is given to each of the four components: possession of a valid license, operating with an approved tariff, paying regulatory fees, and having timely and complete replies to WRA information requests (p. 70).

Note that justifying explicit weights can be problematic unless stakeholders have participated in the process that created the weights. In particular, that average weights may not be the same as marginal weights when considering priorities. The relative importance of performance on specific KPIs in the future is likely to be different from fundamental weights that might be given each KPI at current levels of performance. Thus, regulators creating an OPI should identify what matters in incremental terms. If all utilities meet a well-accepted target, that particular KPI could be assigned a lower weight, allowing the OPI to reflect future priorities. For example, the inclusion of the "Regulatory Perception" KPI was somewhat important for the WRA since it only began publishing performance reports in 2012—and some utilities were still not complying with regulatory rulings. By 2015, WRA regulated 57 water (and wastewater) utilities, whose "Regulatory Perception" scores ranged from 10 to 95. The OPI ranged from 16.99 to 99.5. Presumably SUNASS (the Peruvian water regulator) omitted consideration of this factor since it did not face the same kinds of compliance issues—its authority and benchmarking system has existed for several decades.

2.4 Give some attention to performance trends when evaluating utility scores.

In the case of Albania, "Top Improvers" are also identified: utilities with scores increasing by more than 10 points (out of a maximum of 100). Of course, a very high performing utility may find it difficult to further increase its OPI since the incremental cost of further improvements is likely to be high. So one should not expect rapid improvements in scores for high ranking firms. Conversely, low performing firms may be able to improve performance through improved internal incentives and the targeting of "low hanging fruit". However, low performing firms may also have inherited very poor networks that create challenges for managers who seek to improve KPIs. In any event, some attention should be given to trends in performance. It can be argued that a low scoring utility that shows improvements in a number of performance dimensions warrants positive recognition (and reward) for taking steps towards improving overall performance. Utilities with KPIs (and an OPI) that are moving in the wrong direction need to be identified as well—if only to alert managers that those overseeing performance are concerned with developments. Of course, the OPI scores over time will not be consistent if weights change and components are added or deleted. Nevertheless, when the KPIs are defined consistently over time, unique circumstances might explain developments, but at least the burden of proof is placed on those closest to the situation. Furthermore, when documented performance improves, both the operator and the oversight agency can take some credit for positive trends (as depicted later in Section 3.3).

2.5 Recognize basic operating conditions as well as unique opportunities (such as scale).

Benchmarking scores require comparability among the units for the scores have any policy relevance. Since geography, topology, hydrology, customer density, customer incomes, scale of production, and age

of networks are beyond current managerial control, those evaluating performance need to identify comparable groups. In the case of Peru, researchers have grouped water utilities by comparable conditions (mountains, forest, and the coast, Corton, 2003). Empirical analysis found that unit operating costs were lowest in the forest due to the source of water. Without taking such conditions into account, managers could be unfairly penalized for not scoring as well as other managers. Similarly, the WRA in Albania has three size groups—as it tries to control for the economies of scale. The standards for particular KPIs differ across the groups. For example, the target indictor for "good performance in staff efficiency" (staff/1000 connections) is set at 10, 6 and 4—for utilities in the smallest, intermediate, and largest markets, respectively. Most regulators utilize some system for grouping utilities they regulate to avoid inappropriate comparisons.

2.6 Present comparisons in ways that are clear and appropriate for target audiences.

Note that the *regulatory system* goes beyond the regulatory agency and the water utility operator to include stakeholders that are in a position to support, block, or blunt reforms that would improve performance. Thus, communication with all stakeholders is central to gaining the trust and support of those supporting initiatives that could improve sector outcomes. Given the difficulties of creating an OPI (a composite indicator) that adequately captures performance, care must be taken in presenting results. In particular, excessive numerical precision is inappropriate. Arbitrarily selecting the top four firms (when the score for the fifth is close to that of the fourth) opens the regulator up to criticism. Note that a single number is easier to interpret than a set of trends in many separate indicators. Nevertheless, it is better to group firms into categories. Some regulatory commissions give grades of A, A-, B+, B, B-, etc. Others use high pass, pass, and fail, where high pass means that targets are met and pass implies that there is systematic movement towards targets. Such summary categories are much easier for citizens to understand and better capture the inherent limitations of numerical comparisons. In addition, assessing progress over time requires some consistency in the components of the OPI. The *OECD Handbook on Constructing Composite Indicators* (OECD, 2008) provides a good summary of the strengths and limitations of OPIs.

3 Accuracy and Reliability of Individual KPIs

Accurate and reliable KPIs enable decision-makers to track trends and identify areas needing changes; these would include operational procedures, maintenance/inventory practices, and capacity investments. For low and middle income nations, these indicators represent valuable information on whether infrastructure performance is improving or in decline. In the former case, those groups responsible for particular stages of production should be rewarded for meeting targets. Managers will not tend to publish data on declining performance. If politicians are not providing the government funds for infrastructure that had been promised in campaigns, they are not likely to publicize weak performance either, since their decisions are part of the problem.

Data produced and shared in a timely, consistent, comparable manner to guide, assess, and improve infrastructure policy involves defining and collecting data, fostering cooperation, and drawing upon stakeholder input during the process. Even in small scale, community-based operations, basic records ensure that funds are being utilized according to the agreed-upon business plan. Transparency regarding performance promotes accountability as well. In addition, donor-sponsored initiatives are more likely to receive follow-on funding for project expansions if those managing the project can document volume delivered, number of customers served, service quality, collections, and other dimensions of performance.

3.1 Data Quality and evidence-based decisions

Managers, policy-makers, and sector analysts would do well to pay more attention to data quality. Unreliable or inaccurate data can lead to inefficient investments and inappropriate initiatives by operators. Reliability refers to the decision-maker's confidence regarding how the data were collected, transmitted and stored. For example, Energy and Water Utility Regulatory Authority (EWURA) for Tanzania, has adopted the International Water Association's approach to characterizing confidence in data (EWURA, 2014). EWURA designates data as (A) having high reliability when it has "sound textual records, procedures, investigation or analysis properly documented and recognized as the best method of assessment." Reliable (or B) data have "minor shortcomings, e.g. some missing documentation, reliance on unconfirmed reports, involves some use of extrapolation." Low reliability (C data) would involve "extrapolation from limited samples for which Grade A or B is available." Data that is Without Reliability (D) reflect "unconfirmed verbal reports, cursory inspections or analysis." By identifying the processes underlying data collection, EWURA is explicitly evaluating the information management system of each utility it regulates.

EWURA also explicitly addresses accuracy. The range of error is used to establish four accuracy bands, from 0-5% (band 1) to "worse than plus or minus 50%" (band 4). Thus a data point labeled A2 has a reliable collection system, but the range of error is still plus or minus 5% to 20%. By utilizing a standardized confidence indictor that builds on reliability and accuracy, EWURA is able to initiate programs that improve data quality over time—thus improving confidence in data collection systems and subsequent analyses and performance comparisons.

By highlighting these characteristics of data used to evaluate utility performance, the regulator is establishing incentives to improve data collection procedures. Management's track record on this dimension of utility activities is given explicit attention, which then places pressure on managers to strengthen processes used by different departments. Data quality is enhanced by ensuring that responsible departments provide data to a central point (thus avoiding information silos), and reviewing source documents to reduce inaccuracies and corruption.

KPIs can be classified into a set of categories that allows managers to establish responsibility within the key departments of the regulated business (whether public or private): technical (for engineers), operational (for the engineers and accountants), financial (for the finance teams), and commercial (for the teams facing the users). The associated key performance outcomes would then be linked to these categories: quality, hours/day of service (technical), non-revenue water or line losses (operational), cash flows and collections (financial), and customer complaints (indicating value for money—from the standpoint of users).

In low income, fragile, and/or conflict affected states, decision-makers tend to lack source documents containing basic data on sector characteristics and performance. Nevertheless, it is important that financial reports be produced so analysts can see the linkages among the three major financial records: income statements, balance sheets, and statements of cash flows. In addition, managers (and regulators) should have access to the following:

- Asset registries (installation date, capacity, location, maintenance records),
- Business Plans (with targets and descriptions of managerial incentives for meeting those targets)

- Operating statistics (number of workers, number of customers, reliability, service quality, etc.)

In all cases, the focus should be on improving the accuracy and reliability of the data underlying basic organizational records that provide the "raw material" for constructing KPIs. Such information also helps investors considering public-private partnerships and bond purchases to better understand conditions on (and under) the ground, reducing risks associated with "negative surprises". However, such information is often lacking in low income countries due to a number of factors:

a. Records might have been destroyed during a conflict;

b. Data may be scattered and hard to access (stored under "bad" conditions—high humidity or retained in "information silos");

c. Data collection and authentication might not be a high priority relative to just getting service to citizens;

d. Lack of staff capacity with regards to basic accounting and information systems;

e. Current management avoids transparency so that the true state of operations will not be available to those critical of current outcomes;

f. Some managers avoid transparency so that potential corruption will not be revealed.

The absence of data documenting the financial and operational trends does not mean that nothing can be done. One recommendation would be that data collection (and reporting) procedures be initiated as a *requirement* for further external funding (by development partners or governments). Basic targets can be set based on "comparable" operations in other regions or nations. Such targets should reflect the priorities of policy-makers and the capabilities of existing operators. Stakeholders need to have input into the process so benefits and costs can both be factored into the setting of realistic targets.

3.2 Procedures for Collecting KPIs, Making comparisons, and Designing Incentives

The procedures for obtaining KPIs and designing incentives begin with data sources that document company operations. The extent of process formalization depends to some extent on the size (and age) of the organization(s) under review. Although the basic steps of creating a team, working with stakeholders, selecting KPIs, authenticating information, and disseminating informative reports apply to all types of organizations, the contexts differ, so some of the tasks differ, depending on maturity and size of the operator.

In situations where there is a lack of technical and institutional capacity, several options are commonly used for collecting and storing KPIs. One option consists of training and empowering Community-Based Organizations (CBOs) in data collection techniques. For example, a water access and quality program implemented by the Asian Development Bank in Pakistan requires empowered local water users to provide biannual reports for monitoring and evaluation purposes. Another option consists on putting donor agencies in charge of KPI collection, at least during the early stages. An example of this is the Danubis Water Platform (www.Danubis.org), an online repository that includes KPIs for countries in the Danube region. The data are collected by the countries themselves, but the website, which includes KPIs, is managed by the World Bank and the International Association of Water Supply Companies in the Danube River Catchment Area. Since local utilities own the Danubis data, access is limited. A more comprehensive example of an accessible database is IBNET website (https://www.ib-net.org/ , funded by the World Bank Group) that makes data available on individual water utilities for comparison purposes.

3.3 Initiating and Revitalizing Data Programs

The steps operators and regulators taking going forward depends on the current situation. Some nationas are still very weak in the area of data collection and analysis. Scholars conducting research on regions or countries should understand just where a data system is along the spectrum of very weak to very strong. The accuracy, reliability, and comprehensiveness of existing data constraints what analysts can realistically achieve. Often data transparency is lacking, either because performance trends are problematic politically or because resources are not being devoted to information systems. In many cases, service is delivered through dispersed and low-capacity operators serving communities, often informally. In such sitations, developing the KPIs used for incentives is challenging. Those monitoring smaller, dispersed operators could take steps for improving data program. Similarly, data initiatives for urban systems can be undertaken to provide foundational data required for assessing and improving operator performance.

- *Create a Benchmarking Team:* where the emphasis is on collaboration between those providing oversight and the local group actually responsible for delivering service. Initially, the team should focus on very basic processes: better to have five accurate indicators than twenty inaccurate indicators;
- *Focus on Establishing Trust:* operations in both small and large areas depend heavily on community support: identifying reasonable targets and designing incentives to achieve them requires substantial local input (so working with stakeholders is a crucial step in the process);
- *Select a Small Set of KPIs*: when initiating a program, decision-makers will tend to find that the available records are quite limited, so capacity building must accompany the selection of KPIs;
- *Gather Raw Data for Yardstick Comparisons*: this stage involves the selection of the appropriate comparison group of companies and indicators;
- *Apply Data Verification Procedures:* after data collection processes have been instituted, techniques need to be developed for ensuring that procedures lead to improvements in measurements over time;
- *Perform Data Analyses:* this crucial stage involves the application of methodologies appropriate for answering key policy questions. For low income states, just establishing baselines and identifying trends generally represents an improvement over past procedures (in terms of professionalism, transparency and continuity). The analyses need not be sophisticated to identify trends;
- *Develop Policy Implications:* explore potential determinants of inefficiencies to lay foundations for future initiatives—capacity building for record-keeping, communication, and community participation are likely to be most important;
- *Engage in Information dissemination*: reporting the results of performance comparisons helps engage stakeholders in the process—for simplicity these results can be presented in simple stop-light Figures with green, yellow and red representing different degrees of progress;
- *Determine the types of incentives most likely to be both acceptable and effective*: there needs to be clear procedures for gaining local support in the planning and implementation of incentives and corrective actions (the best “carrot” for meeting goals is the provision of future funds to expand and improve the current set of services);
- *Promote Stakeholder Awareness*: all those affected by local projects deserve to be included in discussions about arrangements, accountability, targets, cost-recovery, and other issues (limiting inefficiency and promoting fairness are a key by-products of serious data collection initiatives.);

- *Prepare Regular Reports for Local Beneficiaries and for the External Organizations Supporting the Initiative:* data collection and analysis is an on-going process for monitoring, communicating, and evaluating performance outcomes. Avoid actions that will be perceived as bureaucratic micro-management. Those with oversight responsibilities need to be viewed as "partners" in the development process, not as adversaries.

Two cases are presented below to illustrate the types of improvements possible when KPIs serve as the basis for setting incentives. Note that the success stories also reflect engineering excellence, a team approach, significant dependence on internal incentives, and leadership. The example from the Philippines is a privately-owned utility regulated by local agencies. The example of a state-owned utility is Uganda's National Water and Sewerage Corporation;

Manila Water: The success of Manila Water in setting targets and devising internal incentives for achieving them is remarkable. In the case of Non-Revenue Water, district metering, leak repairs, new pipes and other initiatives have contributed to substantial success in improving performance. However, the key step involved data collection and analysis, which led to the selection of strategies appropriate for the local situations. As Table 3 shows, Manila Water took on different geographic areas over time, each with different network histories, operating conditions and opportunities. The reductions in NRW by 2018 support the point that this regulated, privately-owned water utility has improved performance associated with this indicator. Similar data (and stories) could be reported on coverage, staffing, and collections.

Table 3. Manila Water: Non-Revenue Water (end-of-period)

	% NRW at the start of operation	2016	2017	2018
Manila Concession	63.0% (1997)	10.8%	11.6%	11.4%
Clark Water	15.0% (2011)	4.6%	5.5%	6.7%
BMDC	50.0% (2017)	-	50.0%	34.3%
Obando Water	52.9% (2018)	-	-	47.7%
Calasiao Water	12.1% (2018)	-	-	5.5%
Laguna Water	48.0% (2009)	21.4%	19.5%	16.7%
Boracay Water	29.0% (2010)	12.6%	22.5%	13.8%
Cebu Water	5.0% (2015)	5.0%	3.0%	2.0%
Estate Water	47.0%(2016)	47.0%	34.0%	38.0%

Source: Manila Water*: Resilient in the Pursuit of Sustainable Water and Sanitation*, Annual Report 2018.

National Water and Sewerage Corporation: The high performance of NWSC is described in several other parts of this study—in the identification of appropriate targets and the design of internal incentives. Here, a number of data points over nearly two decades show the trends that a policy-maker would be interested in. KPIs like service coverage, staff per 1000 connections, collection efficiency, and NRW all show improvement—even as the utility took on nearly 200 smaller towns in recent years (Table 4). Great emphasis has been placed on data accuracy and reliability since incentives are based on reported data.

Table 4. Uganda's National Water and Sewerage Corporation: Key Statistics

Performance Indicator	1998	2011	2017
Number of NWSC towns	11	23	218
Service Coverage (%)	48	75	79
Population served (Million people)	na	2.5	6.3
Pipe Network Length (Km)	2,000	6,500	12,500
Total Connections	50,826	272,406	530,000
Metering efficiency (%)	73	99.8	99.9
Staff per 1000 Connections	36	6	6
Collection Efficiency (%)	60	95	98
Non-Revenue Water (%)	60	33	31
Annual Turnover (Million USD)	20	48	92
Profit (Before. Dep) (Million USD)	4.0 (-)	12 (+)	20 (+)
Asset Value (Million USD)	150	250	414

Source: National Water and Sewerage Corporation Reports

3.4 Five Additional Examples of Applying Benchmarking

The above steps have been taken in various degrees by a number of countries. The cases of Albania, Tanzania, and Peru presented earlier represent stable situations, where regulators have been able to develop data frameworks that provide relatively accurate and reliable information on a wide range of performance outcomes. Similarly, Manila Water and NWSC are both data-driven utilities, where the KPIs serve as foundations for setting realistic targets and establishing incentives for reaching those targets. Institutional arrangements that promote transparency make citizens aware of how well their service providers are complying with rules and whether performance trends are in line with comparable utilities. In this section, cases of water (in Cambodia, Zambia, Uganda, Kosovo, and Jamaica) illustrate how data collection has been central to setting targets and designing incentives. However, most other nations utilize some form of (at least) rudimentary benchmarking in their regulatory systems. When the comparisons are publicized and utilized in developing targets and incentives, they can have a positive impact on the sector.

Case of Cambodia: KPIs for rural Water are identified in an ADB Report that includes key lessons from recent initiatives and areas for improved coordination and capacity-building in rural water in Cambodia: http://adb.org/sites/default/files/projdocs/2009/38560-CAM-RRP.pdf Accounting, auditing, and reporting procedures are given some attention. In the case of the Phnom Penh Water Supply Authority (PPWSA), there was a major turnaround from 1993-2006. For example non-revenue water was reduced from 72% in 1996 to 6% in 2010; staff per 1,000 connections was reduced from 22 to 3 during that period. Here, the Guiding Coalition was spearheaded by the company General Director, Ek Sonn Chan. Incentives include high staff salaries and bonuses for high performance (http://www.ppwsa.com.kh/ and http://www.ppwsa.com.kh/en/index.php?page=sharingthereformprocess). Once high levels of performance were achieved, managerial pressure to maintain and improve on the standards has continued. This suggests that while backsliding is possible, changes in an organization's culture can have long-lasting impacts. Data-driven decision-making coupled with strong internal incentives promote strong performance.

Case of Zambia: The publication of data on individual utility performance relative to KPIs of other utilities has served as a major element in Zambia's regulatory system. Zambia's water regulator, National Water Supply and Sanitation Council (NWASCO), makes its annual sector reports available on the web on a timely basis. NWASCO collects KPIs on a wide range of KPIs and has developed weights to be used for an overall performance score, with particular attention given to data quality. Clear Diagrams capture relative performance: For example, in addition to presenting the "numbers", stop lights and arrows are utilized to indicate levels and trends over time of individual KPIs.[4] CEO awards are presented by NWASCO—giving visibility to leaders who have achieved strong results. In addition, the sector regulator will be collaborating with a newly established institution, the Water Resource Management Authority (WARMA) to develop greater sustainablility in the nation's water resources. Future price and availability of water to residences, industry and commercial enterprises depends on understanding hydrological patterns and future demands. Inter-agency collaboration and this expansion of performance indicators illustrate a broader approach to regulation that promotes long term approaches to water sector performance assessment.

Case of Uganda: Rural and urban utilities face different challenges that warrant different frameworks. Mugisha and Berg (2008) outline the steps taken by the government-owned National Water and Sewerage Corporation (NWSC) to improve performance for twenty-three operators (under a performance contract approach monitored by a team from several ministries rather than a formal sector regulator). For urban areas, NWSC utilized KPIs for setting targets and rewards for the managers of local water systems. NWSC headquarters served a coordinating, capacity-building, and incentivizing role to meet targets set in a series of performance contracts. The success of the resulting KPI-driven internal incentive system led to a dramatic expansion of number of towns under NWSC (Figure 2 above). The institutional framework for the added rural water supply operators draws upon data from over 200 operators, with differential reporting requirements, depending on the size of the utility. The framework facilitates harmonization of national objectives with donor support.[5] The rural indicators provide information on access, functionality, value for money, access/use (sanitation), quality, quantity, equity, access/use (hygiene), management, gender, and water resources management compliance.

Case of Kosovo: The Water and Wastewater Regulatory Office of Kosovo (WWR0 http://www.wwro-ks.org/English/index.html) has under 20 employees. Established in 2004, it has worked with the operators' association and other stakeholders to utilize KPIs in setting tariffs and enforcing compliance with service standards. Examination of its annual reports shows the emphasis on KPIs in making comparisons across the seven Regional Water Companies.[6] WWRO is able to evaluate performance and set targets for each utility based on data on water quality (% passing bacteriological tests and Physical-Chemical tests), water pressure, continuity of supply (24x7), pipe bursts, non-revenue water, service coverage, % meters, complaints, volume and value of water sales, unit costs (operations plus capital maintenance), investments, and other indicators. The DANUBIS water platform facilitates networking,

[4] See: http://www.nwasco.org.zm/jdownloads/Publications/Urban%20and%20Peri-Urban%20WSS%20Sector%20Reports/sector_report_2014.pdf p. 4.

[5] Institutional responsibilities and ten "Golden Indicators" are outlined in http://www.slideshare.net/ircuser/2-ssozi-uganda . Also, see https://www.nwsc.co.ug/index.php/about-us/ourprofile .

[6] http://www.wwro-ks.org/English/Publications/AnnualPerfor/RAPORTI_I_PERFORMANCES_2014_ENG.pdf

training, and data-sharing in the Danube region.[7]

Case of Jamaica: In 2008, the Office of Utilities Regulation (OUR) established a K-Factor Fund to expand NWC's capital investment program, to enhance efficiency improvements, especially to reduce Non-Revenue Water (NRW). Most of the cash flows from the price mark-up were to be applied to NRW reduction since the levels at the time were above 60%: bringing this number down to the interim target (50%) was the highest priority for both the regulator and the operator, the National Water Commission (NWC). An external review in 2017 (HYSTRA, 2017) noted that the national target was not reached, and identified a number of deficiencies in the associated programs, including measurement and monitoring issues. That report also described the contract (a Co-Management Agreement)between NWC and Miya (an Israeli company) to reduce NRW in the Kingston area that was relatively successful. That program included incentives for reaching NRW targets for the specific area under remediation. However, weaker performance in other regions kept the national number above the national NRW target. This program illustrates how a focused initiative based on a single KPI can direct managerial attention to more intensive initiatives. It also underscores the importance of having accurate and reliable disaggregaed data to establish baselines and outcomes. Going forward, OUR will be introducing additional incentives for NWC to reach the regulatory objective.

4 Recommendations for the Use of Indicators

Nobel prize-winning Economic Psychologist Daniel Kahneman (2011) describes System 1 and System 2 as alternative "modes of thinking". The first is rapid, intuitive, and relies on emotions. The second features a more deliberative and slow process that emphasizes logic. He notes that the human tendency is to substitute questions that are easy to answer for more difficult, complex questions. Benchmarking can be a tool for recognizing the need to operate from a System 2 framework when evaluating performing and designing incentives, however, it can also limit a person's perspective. Kahneman labels a common source of bias as WYSIATI ("What you see is all there is.") A corollary is "You manage what you measure." So when decision-makers form a story around the data that have been assembled, they can make mistakes. The key point here is " . . . that the less information you have, the easier you can 'construct a coherent, believable story'—even if the information you have is not true or not complete. A mind influenced by WYSIATI 'will achieve high confidence much too easily by ignoring what it does not know'."[8] The dangers of System 1 thinking in the context of performance assessment are clear. Several decades ago, Meadows (1998) identified seven limitations (or pitfalls) in the use of indicators. This section provides recommendations that can help analysts be more confident and comprehensive in developing policy implications from studies. The suggestions apply to analyses of KPI trends (the focus of this paper) and to more technical studies using DEA, SFA, or other benchmarking tools.

i. Avoid over-aggregation: if too many items are lumped together, their combined message may be indecipherable. In addition, the weights given various components might be inconsistent with the views of key stakeholders. The earlier section on creating and applying an OPI identified a number of issues with such an aggregation process. For example, a single index would not reveal regional disparities in particular KPIs. Similarly, combining data on water tanker performance (in terms of quality and

[7] http://www.danubis.org/ .

[8] The extended quote is from Scott Hempling's "Regulatory Reactivity Risks Mental Errors," April 2019.

affordability) with data from a water network service provider would not really tell us much about overall sector performance. That also implies that disaggregation (by region or urban area) can reveal issues in need of attention.

ii. Avoid measuring only those elements that are easily measurable: specialists can make the mistake of focusing on what is easily quantified rather than what is important. (e.g. "tons" of hazardous chemicals, rather than toxicity). It is better to have a *rough estimate* of an important dimensions of performance than a *precise calculation* of some metric that has minimal implications for producers or consumers. Given this point, data collection needs to focus on characteristics that are salient to customers or have clear implications for the long term sustainability of the operator. Issues around safety/health standards are particularly important for water and sanitation, but measurement can be problematic. Guidance on KPIs (including definitions) and links to a number of resources information are available at the website for the International Benchmarking network for Water and Sanitation Utilities < https://www.ib-net.org/ >. The International Water Association's *Performance Indicators for Water Supply Services* (Second Edition) is another resource for this important topic (Alegre, et. al., 2013).

iii. Avoid oversimplification: analysts need to avoid depending on a false model--does the indicator really tell us a fundamental relationship? If the indicator construction process lacks transparency and/or sound statistical or conceptual principles, stakeholders will not be convinced that the indicator is relevant for decision-making. For example, the IWA water balance definitions provide a precise way to calculate non-revenue water (NRW= System input volume minus billed authorized consumption). Nevertheless, the specific performance indicator depends on network density, where less than 20 connections/km of pipe should use liters per km Pipe Mains per day rather than Liters per Service Connection per day. Furthermore, once average pressure is incorporated into a more comprehensive Infrastructure Leakage Index (ILI), calculation of this indicator becomes even more difficult--especially in developing countries. Thus, inappropriate design of incentives that focus on specific KPIs can harm overall performance (in terms of efficiency and service quality) when excessive resources are devoted to achieving a specific indicator. For example, focusing on reducing number of staff per 1000 customers could lead to a number of inefficient outcomes. Managers might substitute other inputs for staff or outsource activities. In particular, the reduction in some activities (like maintenance) would improve a particular indicator today but greatly weaken performance in the future. *Such behavior underscores the importance of having a balanced set of indicators (and associated threshold requirements) rather than a single KPI or OPI.*

iv. Devise ways to improve data systems: if an index carries bad news, someone may be tempted to alter it, delay its release, redefine the variable, defund data collection, or suppress the information! Thus, the accuracy and reliability of data is central to a sound information system. For developing countries, KPIs allow for fair comparisons which make it easier to avoid situations where those in power give preferential treatment to their ethnic groups, tribes, or political parties. Of course, that requires that transparency be given priority in the process. Of course, legacy systems (and information silos) are not automatically dismantled when new approaches to decision-making become possible, but leadership and teamwork can support organizational change.

v. Focus on the value of outcomes as reflected in direct experience: indicators give "precise" numbers—but are these related to what people actually experience (e.g. Quality)? Technicians can sometimes miss the most important dimensions of performance experienced by customers. For example, it might be impossible to restore a high level of service immediately after a storm. Priority might be given to service being restored to a large portion of the population having access, even though the level of service is not

ideal. Focusing on 24 hour service in such circumstances could be quite inappropriate since customers may be able to store water. A related issue arises when a very large number of indicators are collected. When the Association of Water Regulators of the Americas (ADERASA) began a collaborative data collection project in the late 1990s, those managing the project started with a goal of over 130 different variables. Most water regulators in the region already had data on basic KPIs, but the jump to over 100 items (used to create more than 30 KPIs) was probably excessive. It would have been better to focus on quantifying the accuracy and reliability of data on a set of core indicators—as noted earlier. On the plus side (in this case) is the role played by a regional regulatory network in promoting comparisons within and across countries.

vi. Avoid over-confidence: indicators may lead people to think they know what is happening, when indicators may be flawed and inaccurate. In developing countries, operators are often severely under-resourced and those serving as regulators (a Ministry or a semi-autonomous sector regulator) may lack the incentive to gather and publicize performance since the indicators could indicate weak performance (and poor oversight). Furthermore, "believing is seeing": people give differential weight to facts that are consistent vs. inconsistent with preconceptions. Basically, results that differ from expectations tend to be rejected, giving excessive weight to our beliefs. This tendency can lead to simplistic (but incorrect) policy conclusions.

vii. Devise comprehensive approaches to evaluating situations: indicators are not the real system. They may miss many of the subtleties and opportunities facing the current system. For example, identifying the distributional consequences or geographic disparities associated with current service arrangements requires some disaggregation of data. In addition, KPIs from current operations may be inconsistent with the sustainability of hydrological resources, staff development, or business plans. Incentives that only focus on the easy-to-obtain KPIs are likely to lead to inefficient remedial actions. Once a serious initiative to develop a system of KPIs has been undertaken, decision-makers should focus on ten to fifteen data elements. These KPIs would be based on inputs and outputs: some financial information (revenues and other sources of funds, operating expenditures (with some breakdowns—such as total wages and salaries, investments), customer base (number and types of customers and average bills by customer type, percent with meters/usage measurement devices, collections), operations (number of staff), service quantity produced, technical and commercial losses (to gauge line losses and theft), and quantity delivered. Such data could be used to create an OPI that is informative.

To these seven recommendations should be added an observation: decision-makers might misuse indicators. Managers might select internal targets that are easy to achieve, thus receiving bonuses unjustified by significant performance improvements. Alternatively, using a pass-fail approach, regulators (or consumer advocate intervenors) might focus on a few indicators that do not meet regulatory targets—arguing that the operator should be penalized, even when the indicators are relatively unimportant or inaccurate (Lynch, et. al., 1994). Nevertheless, benchmarking initiatives using KPIs and performance scores are essential if progress is to be made in the water sector.

5 Concluding Observations

Evaluating water utility performance using KPIs and OPIs represents an important step in identifying strong and weak performers, establishing reasonable targets, and creating internal (managerial) and external (regulatory) incentives for improving utility operations. Of course, a benchmarking initiative needs to be embedded in a *regulatory system* that goes beyond the regulatory agency and the water

utility operator to include stakeholders (including customers, Ministries,and citizens without quality service). Domestic politics and tribalism can limit the effectiveness of regulatory institutions. Stakeholders need to have a shared vision, even if they have different preferred strategies for meeting objectives. The point is that those responsible for improving water sector performance take steps towards a comprehensive benchmarking program so others can build upon solid foundations.

A review of past benchmarking studies (Berg and Marques, 2011) suggests that the following principles of sound governance are central to the collection, selection, and publication of KPIs:

- *Clarity* in the regulatory objectives that are directly linked to incentives and targets,
- *Transparency* that make information available to all stakeholders,
- *Participation* (engaging operators and other stakeholders in the prioritization of data elements),
- *Regulatory powers* enabling the commission to collect information (governance matters, Bromley and Anderson, 2018)
- *Capacity building* in the skills required for evaluating operating and financial performance,
- *Agency funding* for the recruitment, training, and retention of competent professional staff,
- *Autonomy* from political pressures that could jeopardize the credibility of regulatory analyses, including avoidance of regulatory capture by powerful consumer groups or the operator

These principles underscore the fundamental need to have evidence-based decisions regarding realistic targets, reasonable incentives, and public acceptability of rules that are promulgated (based on accurate data). Ultimately, however, *individuals* move the situation forward. Policies are not self-implementing. Those exercising leadership both "stir and steer"; individuals make sure that issues get raised and addressed and they guide the group once the appropriate strategies have been identified. If there is a window of opportunity (the political, economic, and social conditions are ready for change), leadership can help organizations develop initiatives (programs) that improve infrastructure performance. While the pitfalls of inappropriate use of KPIs are real, the potential for informing decision-makers and the public in general clearly outweigh the limitations of data collection and analysis as currently practiced by those developing, implementing, and responding to public policies in the water sector.

REFERENCES

Alegre, Helena, et. al. (2013) *Performance Indicators for Water Supply Services—Second Edition*, International Water Association.

Berg, S. (2010). *Water Utility Benchmarking: Measurement, Methodologies and Performance Incentives*, International Water Association, London, xii–170.

____ (2013). "Advances in Benchmarking to Improve Water Utility Operations: a Review of Six IWA Books," *Water Policy*, 15, 325-333.

____ and R. Marques (2011). "Quantitative Studies of Water and Sanitation Utilities: A Benchmarking Literature Survey," *Water Policy* 13, 591-606.

_____ and M. Phillips (2017). "Data Availability as a Key Tool for Regulating Government-Owned Water Utilities," *Utilities Policy,* December, 30-37.

Bromley, Daniel W. and Glen Anderson (2018), "Does Water Governance Matter?" *Water Economics and Policy*. Vol. 4, No. 3.

Cetrulo, Tiago Baliero, Rui Cunha Marques, and Tadeu Fabrico Malheiros ((2019). "An analytical review of the efficiency of water and sanitation utilities in developing countries," *Water Research*, 372-380.

Cabrera, E. Jr., Dane, P., Haskins, S. & Theuretzbacher-Fritz, H. (2010), *Benchmarking Water Services: Guiding Water Utilities to Excellence*, International Water Association, London, xix–164.

Corton, Maria Luisa (2003), "Benchmarking in the Latin American Water Sector: the Case of Peru," *Utilities Policy*, 11, 3. September, pp. 133-142.

EWURA (2014), *Water Utilities Performance Review Report 2013/2014*. December. http://www.ewura.go.tz/wp-content/uploads/2015/04/EWURA-WATER-REGIONAL-REPORT-2014-FINAL.pdf

HYTSA Estudios y Proyectos S.A. (2017) Management and Operational Audit of the K-Factor Programme: Final Report to OUR, December.

IB-Net: International Benchmarking Network for Water and Sanitation Utilities, http://www.ib-net.org/.

Kahneman, Daniel (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux, 1-499.

Lynch, John G. Jr., T. Buzas, and S. Berg (1994). "Regulatory Measurement and Evaluation of Telephone Service Quality," *Management Science*, 40, 2, February, pp. 169-194.

Meadows, Donella H. (1998), *Indicators and Information Systems for Sustainable Development*. http://donellameadows.org/wp-content/userfiles/IndicatorsInformation.pdf

Mugisha, S. and S. Berg (2008), "State-Owned Enterprises: NWSC's Turnaround in Uganda," with Silver Mugisha, *African Development Review*, Vol. 20, No. 2, 305-334.

Mugisha, S. (2011). *Utility Benchmarking and Regulation in Developing Countries: Practical Application of Performance Monitoring and Incentives*. IWA Publishing.

OECD, (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide* http://www.oecd.org/sdd/42495745.pdf

WRA (Water Regulatory Authority of Albania), September 2016 *Report on the Performance on the Water Supply and Sewerage Companies, 2015,* http://www.erru.al/doc/Raporti_Performances_2015_eng.pdf